C

3-Cycle

▶ Canonical Network Motifs

^{13}C Fluxomics

▶ ^{13}C Metabolic Flux Analysis

^{13}C Metabolic Flux Analysis

Meghna Rajvanshi and Kareenhalli V. Venkatesh
Department of Chemical Engineering, Indian Institute of Technology Bombay, Powai, Mumbai, Maharashtra, India

Synonyms

^{13}C Fluxomics

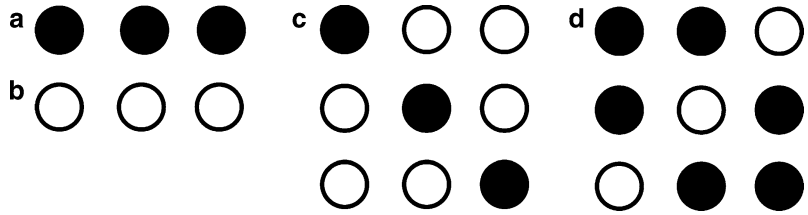
Definition

^{13}C MFA is an experimental approach to compute intracellular flux pattern. It involves tracer studies where substrates are labeled with stable isotopes, such as ^{13}C, and the labeling pattern of metabolites is subsequently measured. Detailed flux distributions can be obtained by a combination of tracer experiments (intracellular labeling information), measured extracellular fluxes and stoichiometric balancing. Nuclear Magnetic Resonance Spectroscopy (NMR) and, more recently, ▶ Mass Spectrometry (MS) have emerged as interesting tools for measuring the level of ^{13}C enrichment in metabolites (Wiechert 2001). It provides valuable insights into the cellular metabolism and is one of the methodologies used in ▶ Metabolic Pathway Analysis.

Characteristics

^{13}C metabolic flux analysis is a powerful technique for modeling biological systems. ^{13}C metabolic flux analysis is free of all the short comings of the stoichiometric ▶ Metabolic Flux Analysis (MFA), though it needs more information in addition to extracellular flux data to have the complete flux distribution. Stoichiometric MFA and ▶ Flux Balance Analysis (FBA) are restricted in their ability to fully quantify the network because they cannot be applied in cases involving the following: (1) metabolic pathways where the energy metabolites, such as ATP, NADH, NADPH, are not balanced; (2) parallel reactions where none of the competing branches are directly coupled to measurable variable or metabolite; (3) metabolic pathways containing cyclic reaction chains not linked to a measurable flux; and (4) metabolic cycles having bidirectional reactions (Wiechert et al. 1997).

W. Dubitzky et al. (eds.), *Encyclopedia of Systems Biology*, DOI 10.1007/978-1-4419-9863-7,

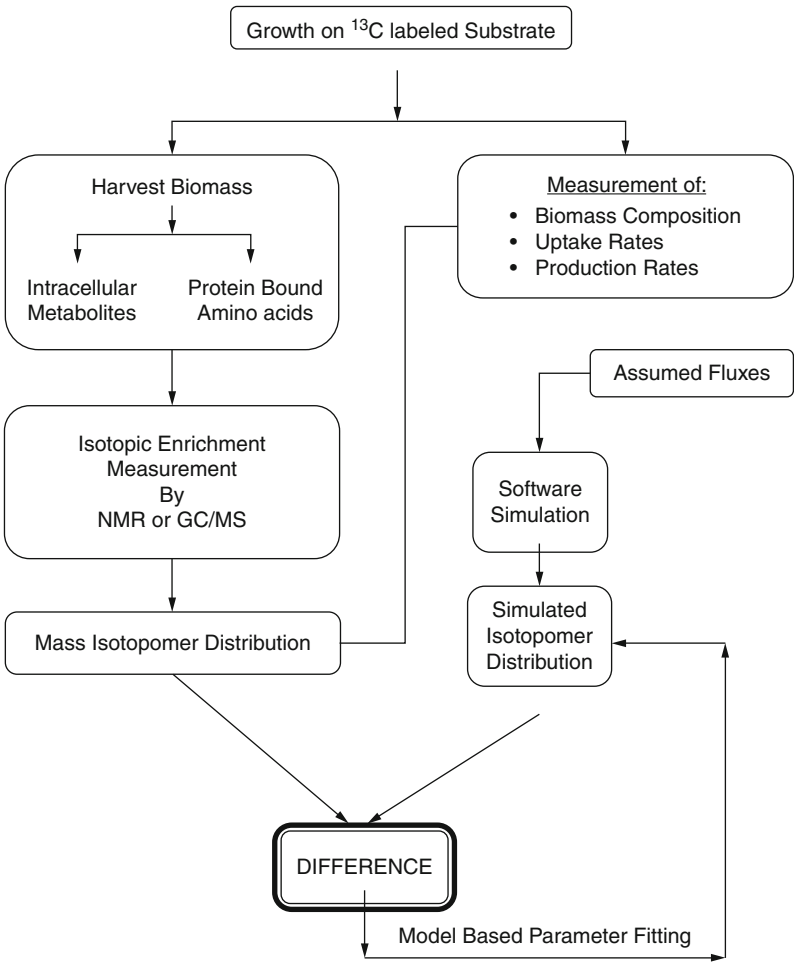


^{13}C Metabolic Flux Analysis, Fig. 1 Different labeling states of a molecule containing 3 C atoms. Labeled C is shown as *black circles*. There are eight positional isotopomers and four mass isotopomers (specified as (**a**–**d**))

Certain assumptions and rules need to be applied while performing ^{13}C metabolic flux analysis as listed below (Marx et al. 1996):

1. It is based on an assumption of a pseudo steady state.
2. It assumes that there is no preference of enzymes for labeled and unlabeled substrates (i.e., no isotope effect).
3. Bioreactor size should be small to minimize the cost of the experiment as labeled substrates are quite costly but it is also necessary that the volume reduction should not be to the extent where stable stationary condition cannot be achieved.
4. It is necessary that all the intracellular metabolites are in steady state both in terms of metabolic fluxes and isotopomer distribution. Typically it takes 2–4 bioreactor residence times to reach the isotopic equilibrium state. Time required for intracellular intermediates to reach isotopic steady state is comparatively less than the time taken by the macromolecules of biomass (proteins, RNA, DNA, cell wall) to reach isotopic steady state. The term isotopomer implies "one of the different labeling states in which a particular metabolite can be found." There can be 2^n different labeling states of a molecule. For example, for a metabolite with 3 C atoms, there can be eight different positional isotopomers while only four different mass isotopomers (Fig. 1). Knowledge of complete isotopomer distribution yields information regarding the enrichment pattern of different intracellular metabolites. This fractional enrichment information reflects how atoms in the reacting species interact with each other to form product molecules. Thus, fractional enrichment data along with C atom transition for all the species involved in a set of reactions defining the intracellular metabolism will yield intracellular fluxes.

The basic procedure followed for ^{13}C metabolic flux analysis (Wiechert 2001; Krömer et al. 2009) is to measure the labeling patterns of metabolites. Carbon labeling experiment consists of growing the organism of interest on ^{13}C substrate such as glucose labeled at a specific location. This labeled substrate is then utilized or consumed across the various pathways finally leading to enrichment of the metabolites intracellular pool. Cells are harvested and are hydrolyzed to get metabolite precursors like amino acids. The pattern of enrichment in these metabolites is identified or estimated by mass spectrometry (MS) or NMR. The output from MS or NMR consists of a spectral pattern. The pattern and the level of enrichment is a fingerprint for a particular kind of pathway involved. The level of enrichment of certain metabolite is a reflection of the permutation and combinations a labeled molecule has undergone. MS is becoming the method of choice for ^{13}C analysis due to its comparatively simpler data interpretation ability even for multimolecular species. NMR on the other hand does not require complex sample processing as MS but its spectral data becomes extremely convoluted for molecules with more than three atoms. A few deconvolution algorithms are available but are mathematically involved and this restricts its usage to only few users with time-developed skill and experience. Thus, this intracellular labeling information, in combination with extracellular fluxes, is used to compute the intracellular fluxes. Several algorithms are available to decipher the meaningful information from such spectra. OpenFlux (Quek et al. 2009), FiatFlux (Zamboni et al. 2005), ^{13}C MFA, etc., are some of the examples of the software available. The enrichment information in terms of mass isotopomer distribution acts as an input to the algorithm. Algorithm uses this mass isotomomer data to obtain the positional isotopomer



^{13}C Metabolic Flux Analysis, Fig. 2 Schematic flowchart of ^{13}C-based metabolic flux analysis

data, which is further used to estimate the fluxes. Software works by minimizing the error between the simulated and experimental mass isotopomer distribution. One fundamental problem of CLE is that it is usually difficult to measure the labeling pattern of many important intracellular metabolites due to their small pool size. To overcome this problem, the labeling patterns of amino acids are often measured, since the amino acids reflect the labeling states of a number of important intermediate metabolites through their precursors. These labeling measurement data provide additional and yet independent constraints on the intracellular fluxes and thus enable a more refined analysis of metabolic fluxes in the complex metabolic network. An overall outline of ^{13}C-based metabolic flux analysis methodology is given in Fig. 2.

Applications of ^{13}C MFA

^{13}C MFA has been widely used in analyzing microbial metabolism and regulation involved in metabolic networks. It is mainly used for analyzing central carbon metabolism but recently has been applied for genome scale networks as well. ^{13}C MFA has good predictability for prokaryotic systems but application of ^{13}C MFA to eukaryotic systems, especially to plant and animal cells, is extremely difficult due to the intracellular compartmentalization. Information of compartmentalization has to be captured by model properly in order to obtain proper fitting of predicted versus experimental data (Wiechert 2001). Similarly application of ^{13}C MFA is obscure when multiple carbon substrates are used. High cost of labeled substrates, need for specialized instrumentations (NMR or MS) for determining

isotopic labeling, and elaborate mathematical and statistical analysis for isotopomer modeling are some additional limitations of ^{13}C MFA, which further restrict its use (Tang et al. 2009). ^{13}C MFA requires excessive experimental data and is quite complicated as compared to other methods of flux analysis; however, for certain cases, ^{13}C MFA is the only solution. For example as stated in the previous section, ^{13}C MFA can resolve parallel pathways, cyclic metabolic pathways and bidirectional fluxes in the metabolic network. ^{13}C MFA has been extensively used in studying metabolism of relevant organisms for biotechnology processes, namely, production of amino acids (lysine, glutamate, phenyl alanine, methionine, etc), vitamin (riboflavin), ethanol, glycerol, antibiotics (penicillin G), and bioplastics (1,3-propanediol) (Iwatani et al. 2008). It has also been used to test the validity of the assumptions made in other methods of flux analysis and about the metabolic network. ^{13}C MFA has been used in studying single gene disruption mutant and establishing robustness in the metabolic network (Blank et al. 2005). ^{13}C MFA has proven to be quite useful in cancer research to study unique metabolic characteristics under diseased condition. It is used to understand flux distribution through important pathways for amino acids and fatty acids biosynthesis in cancer cells (Yang et al. 2008). Other important application of ^{13}C MFA is in identification of bottlenecks in metabolic network of the industrially important organisms. Identified target genes, upon manipulation through genetic engineering techniques, lead to enhanced productivity of the metabolite of interest. ^{13}C MFA has also been used in identifying drug targets for various diseases. By studying flux distribution in pathogens or diseased cells, it might be possible to identify a pathway crucial for the survival of the pathogen inside the host cells. Such pathways can be potential drug targets (Tang et al. 2009). Thus, in spite of being a complex methodology for flux analysis, ^{13}C MFA has wide applicability in different fields. It provides important insights about the characteristics of a metabolic network, for which this technique is indispensable.

Cross-References

- ▶ Flux Balance Analysis
- ▶ Mass Spectrometry, Proteomics, and Metabolomics
- ▶ Metabolic Flux Analysis
- ▶ Spectroscopy and Spectromicroscopy

References

Blank LM, Kuepfer L, Sauer U (2005) Large-scale 13C-flux analysis reveals mechanistic principles of metabolic network robustness to null mutations in yeast. Genome Biol 6(6):R49

Iwatani S, Yamada Y, Usuda Y (2008) Metabolic flux analysis in biotechnology processes. Biotechnol Lett 30(5):791–799

Krömer J, Quek L-E, Nielsen L (2009) 13C-Fluxomics: a tool for measuring metabolic phenotypes. Australian Biochemist 40(3):17–20

Marx A, de Graaf AA, Wiechert W, Eggeling L, Sahm H (1996) Determination of the fluxes in the central metabolism of *Corynebacterium glutamicum* by nuclear magnetic resonance spectroscopy combined with metabolite balancing. Biotechnol Bioeng 49:111–129

Quek L-E, Wittmann C, Nielsen L, Kromer J (2009) OpenFLUX: efficient modelling software for 13C-based metabolic flux analysis. Microb Cell Fact 8(1):25

Tang YJ, Martin HG, Myers S, Rodriguez S, Baidoo EEK, Keasling JD (2009) Advances in analysis of microbial metabolic fluxes via 13C isotopic labeling. Mass Spectrom Rev 28(2):362–375

Wiechert W (2001) ^{13}C metabolic flux analysis. Metab Eng 3(3):195–206. doi:10.1006/mben.2001.0187

Wiechert W, Siefke C, de Graaf AA, Marx A (1997) Bidirectional reaction steps in metabolic networks: II. Flux estimation and statistical analysis. Biotechnol Bioeng 55(1):118–135

Yang C, Richardson A, Osterman A, Smith J (2008) Profiling of central metabolism in human cancer cells by two-dimensional NMR, GC-MS analysis, and isotopomer modeling. Metabolomics 4(1):13–29

Zamboni N, Fischer E, Sauer U (2005) FiatFlux–a software for metabolic flux analysis from 13C-glucose experiments. BMC Bioinformatics 6:209

C Domain

- ▶ Constant (C) Domain

C Gene

- ▶ Constant (C) Gene

C Type Domain

- ▶ Constant (C) Domain

Calibration

Clemens Kreutz[1] and Jens Timmer[2,3,4]
[1]Institute for Physics, Freiburger Center for Data Analysis and Modeling (FDM), University of Freiburg, Freiburg, Germany
[2]Institute for Physics, University of Freiburg, Freiburg, Germany
[3]BIOSS Centre for Biological Signalling Studies and Freiburg Institute for Advanced Studies (FRIAS), Freiburg, Germany
[4]Department of Clinical and Experimental Medicine, Linköping University, Linköping, Sweden

Definition

Calibration is the process of finding the input-output relationship of a system or device (Skoog 2007). Usually, this term is used if an assay or measurement technique is characterized. The calibration curve relates the physical quantity of interest to the experimentally accessible quantities, e.g., the relationship between the concentration of a compound like protein or mRNA to intensities or electric current.

Measurement devices are commonly designed to have a linear calibration curve, e.g., $g(x) = a_0 + a_1 x$. Here, a_0 denotes an offset or background parameter and a_1 and the scaling parameter. The purpose of calibration is then to estimate the parameters of the calibration curve by systematically changing x and evaluating the output $y = g(x) + noise$. In addition, the range of x where the linear relationship holds has to be detected.

Because a mathematical model of a biological system also describes an input-output behavior, estimation of model parameters is termed *model calibration*.

Cross-References

- ▶ Experimental Design
- ▶ Observable
- ▶ Parameter Estimation

References

Skoog D (2007) Principles of instrumental analysis. Thomson, Brooks/Cole

Canalization

Philippe Huneman
Institut d'Histoire et de Philosophie (IHPST), des Sciences et des Techniques, Université Paris 1 Panthéon-Sorbonne, Paris, France

Definition

First recognized by Conrad Waddington, canalization means that many perturbations in the genes will not change the phenotypic output of the process. Waddington in *The Strategy of the Genes* (1957) used the metaphor of "epigenetic landscapes" (Fig. 1a), genes acting like the nails which hold the ropes underpinning a valley (Fig. 1b) in such a way that moving or displacing them does not change much the general landscape. Canalization occurs also with regard to the environment. Phenomena like genetic assimilation mean therefore that new channels have been created. Canalization also implies that organisms with various genotypes may have the same phenotype, which contributes to explain typicality in various species.

Canalization raises two main sets of questions. First, developmental and molecular: what are the mechanisms underlying it (Masel and Siegal 2009)? Second, evolutionary: given that canalization means robustness in the face of either mutational or genetic change, one should ask which came first (Wagner 2005): has canalization evolved as an adaptation in the face of environmental variations? Or has it been selected because of the buffering against mutational perturbation (genetic noise)? Lehner (2010) argued that it is indeed an adaptation for environmental variation, and robustness versus mutational robustness is a by-product.

Besides the evolutionary cause of canalization, biologists also investigate its evolutionary effects. It has been argued that canalization, as a form of robustness, indeed increases evolvability: even if it seems to decrease phenotypic variability and then the opportunities for selection, it appears that canalized developmental systems, because of their robustness across a wide range of variations, allow more evolution, since genotypic variation can accumulate without hampering the production of a viable phenotype.

Canalization,

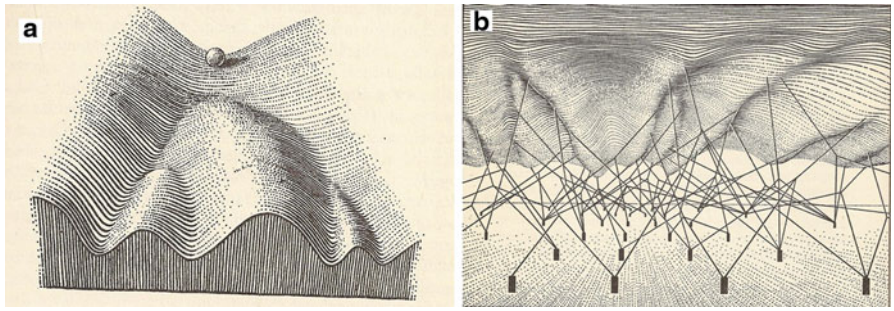


Fig. 1 Canalization (after Waddington). (**a**) Epigenetic landscape, figuring the robustness of developmental fates. (**b**) Genetic underpinning of canalization

Cross-References

► Explanation, Developmental

References

Lehner B (2010) Genes confer similar robustness to environmental, stochastic, and genetic perturbations in yeast. PLoS One 5(2):e9035

Masel J, Siegal M (2009) Robustness: mechanisms and consequences. Tr Gen 25(9):395–403

Wagner A (2005) Robustness and evolvability and living systems. Princeton University Press, Princeton

Cancer

Jingky Lozano-Kühne
Department of Public Health, University of Oxford, Oxford, UK

Synonyms

Malignant tumors; Neoplasms

Definition

Cancer is a generic term for a class of diseases characterized by the rapid uncontrolled growth of abnormal cells which can invade adjacent body parts and spread to other organs. The process of spreading to other organs or other locations of the body is known as metastasis and is a major cause of death from cancer (WHO 2011).

Cross-References

► Cancer Resources
► Neoplasms

References

World Health Organization (2011) Cancer. Fact sheet 297. World Wide Web. http://www.who.int/mediacentre/factsheets/fs297/en/. Accessed 18 May 2011

Cancer and Environmental Influences

C. Vyvyan Howard and Gesa Staats
Centre for Molecular Biosciences, University of Ulster, Coleraine, UK

Definition

Cancer is a multifactorial multistage disease. Environmental influences are implicated in the etiology of cancer at all stages of the life cycle, though particularly during the developmental period. The mechanisms underlying the development of cancer are critical to an understanding of the magnitude of environmental influences.

Characteristics

Background

The environment can be defined as the totality of the biological and nonbiological influences that act upon an organism, a population of organisms, or an

ecological system and that can influence survival. Biological factors include the organisms, their food, and their interactions. Nonbiological factors can be divided into physical and chemical and usually act in combinations. Purely physical influences include all electromagnetic radiations, temperature, and meteorology. Purely chemical influences include soil constitution, water and air quality, and chemical pollution. Organisms respond to environmental changes by evolutionary adaptations.

The etiology of cancer remains unknown, in that a detailed mechanistic description of how cancers arise is lacking. There are many correlations with factors that are associated with an increased or decreased incidence of cancer. Environmental factors can influence cancer incidence and such factors are normally loosely classified into "lifestyle factors" and "external factors." Lifestyle factors are generally regarded to be under the control of the individual and include smoking, drinking, diet, drug taking, exercise level, sexual behavior, and occupation. External factors include background ionizing radiation, ambient chemical pollution, and poor air quality. This entry will concentrate on the effects of chemical pollution on cancer incidence, emphasizing some recent advances in knowledge on the effects of epigenetic influences during development. That is not to ignore the widely studied and highly significant effects of radiation and lifestyle factors, such as tobacco smoking, on cancer incidence. The main tool used to detect alterations in the distribution of disease in human populations is epidemiology. A number of seminal contributions in the eighteenth century contributed to the early development of cancer epidemiology. In 1713, Bernadino Ramazzini (the "father of occupational medicine") reported a relatively high incidence of breast cancer in nuns, linked to a very low incidence of cervical cancer. He pondered whether this might be related to a celibate lifestyle. In 1761, in his book "Cautions Against the Immoderate use of Snuff," John Hill made the first observations linking tobacco use with the development of cancer. Then in 1775, Sir Percival Pott described an increased incidence of carcinoma of the scrotum in chimney sweeps, an occupational disease.

The industrialized world has seen a steady increase in the incidence of cancer throughout the past 150 years. In the UK immediately after the Second World War, the lifetime risk of developing cancer was 1 in 4. By the year 2000, it was less than 1 in 3. The risk of women developing breast cancer in the UK in 1960 was 1 in 20. The latest figures indicate that 1 in 8 women in the UK will now develop breast cancer. These trends are described for a variety of cancers across a large number of countries. Developed countries have become used to high levels of cancer in their populations as "the norm." It has been demonstrated that, for some cancers, the average age of onset within the UK population is decreasing (Newby et al. 2007). This provides strong evidence that an environmental factor is operating.

However, it is a valid question to enquire whether members of traditional preindustrial societies who lived into old age, invariably demonstrate increasing cancer incidence? An examination of the observations made on those societies, whose lifestyles had remained virtually unchanged for hundreds of years, shows a remarkably consistent pattern: an almost total absence of cancer. To investigate this, it is necessary to go back quite a long way into the written record. A key source of information on this topic is a book, by Vilhjamur Stefansson (1960), in which he cites a variety of authors who reported little or no signs of cancer occurring in unindustrialized societies living traditional lifestyles. The correspondents were medically qualified doctors visiting so-called primitive societies in many and widespread communities in Canada, Africa, India, and South America.

In 1915, a report entitled "The Mortality from Cancer Throughout the World" was authored by Frederick L Hoffman, the then chairman of the committee on statistics of the American Society for the Control of Cancer. It analyzed literally thousands of separate reports and all of the data available at that time. A major conclusion was "the rarity of cancer among native man suggests that the disease is primarily induced by the conditions and methods of living which typify our modern civilization." The author went on to explain that "... a large number of medical missionaries and other trained medical observers living for years among native races throughout the world, would long ago have provided a more substantial basis of fact regarding the frequency of occurrence of malignant disease among the so-called uncivilised races, if cancer were met with among them to anything like the degree common to practically all civilised countries"... "Quite to the contrary, the negative evidence is convincing that, in the opinion of

qualified medical observers, cancer is exceptionally rare among the primitive peoples. . ."

It is easy to dismiss such evidence as "anecdotal." However, with the reliably observed continuation of the trend to increased cancer incidence in developed countries, combined with the absence of any evidence at all about substantial cancer incidence rates in preindustrial societies, it seems reasonable to conclude that cancer as a disease is related primarily to industrialization. Other arguments can be brought into play to counter this assertion. For example, average life expectancy is higher in developed countries, and cancer incidence is a function of age (though if one is living in a "soup" of carcinogenic influences from conception to grave, then length of exposure would be expected to be significant, though not necessarily causal in itself). It should be noted though that average life expectancy itself, as a measure, can be deceptive. It has been used to assert that life in earlier times was brutal and short with most adults dying in their third decade. This is a very simplistic explanation. In preindustrial societies, the death rate in infancy was high, but if adolescence was reached then, according to the same medically qualified observers mentioned above, the chances of living a reasonable life span in good health were high and unlikely to end in the development of cancer. These observations are supported by the recognition among most epidemiologists that the major contribution to the rise in average life expectancy in developed countries has been the reduction in infant mortality.

Nature Versus Nurture

Epidemiological studies of monozygous twins are regarded as the golden yardstick for deciding whether a disease is predominantly determined either by genetic or, on the other hand, environmental factors. Lichtenstein et al. (2000) reported a study of the concordance of similar cancers in 45,000 pairs of identical twins. The average rate of concordance was rather low ~10–15%. The conclusion from the study was that environmental factors far outweigh genetic influences in the etiology of cancer. Czene et al. (2002) reported similar conclusions. The concept should not be surprising, because Doll and Peto (1981) showed, in their study of cancer incidence in native Japanese compared to émigré Japanese in Hawaii, that breast cancer rates were four times higher in the emigrants, approaching the incidence prevailing in the indigenous Caucasian inhabitants. The topic has been addressed more recently by Howard and Newby (2004) and Irigaray et al. (2007).

Low-Dose Epigenetic Phenomena and Cancer

There is increasing emphasis on the influence of low-dose intrauterine exposure of the fetus to ubiquitous environmental pollutants on the subsequent risks for cancer in later life. This rather subtle form of environmental chemical exposure, associated with long latency periods, has been prompted by the demonstration in animal models that doses of certain chemicals, several orders of magnitude below the concentration required to produce measureable effects in adults, appear to be able to perturb normal development. The explanation for this is likely to be found in considering the physiological concentrations at which cell signaling molecules, such as hormones, operate. This is typically very low, in parts per trillion. Moreover, dose-response curves tend to be nonlinear and with an inverted U shape. This indicates that there is an optimal concentration at which such cell signaling molecules operate, that is, there can be too little and/or too much, as is typical with receptor-mediated events.

In adult animals, cell signaling molecules are involved in maintaining homeostasis in an intact organism in which most structural and functional elements are relatively stable. In the fetus, on the other hand, cell signaling molecule expression is intimately tied up with control of cell proliferation and apoptosis during windows of developmental activity. Perturbation of the tightly regulated levels of key cell signaling molecules can lead to subtle changes in the phenotype which can be expressed anatomically as tissue dysgenesis or physiologically as functional deficits. Such mechanisms appear to be highly conserved between species, making animal studies potentially more comparable across species than for adult high-dose toxicology.

Hitherto, classical toxicology has mainly been predicated on adult animal studies. Teratological studies have mainly been restricted to the detection of naked eye malformations. Why is this of relevance to the study of environmental causes of cancer? It is known that any child born with a malformation is marginally more likely than average to develop cancer in their lifetime. There is a connection, as yet not fully understood, between the ways that tissues are assembled in

the fetus and the subsequent development of cancers. This process is normally associated with long latencies.

Cell Signaling Disruption in the Fetus and Cancer

The most widely studied group of chemicals capable of affecting the function of natural cell signaling molecules are classed as endocrine-disrupting substances (EDSs). There is general agreement that high-dose therapeutic exposure to synthetic hormones at critical stages of development can cause cancer, for example, intrauterine exposure to diethylstilboestrol (DES) led, after a ~20-year latency period, to clear cell adenoma of the vagina in female offspring. However, a recurring argument against the causal link between environmental exposure to such environmental pollutant EDS and subsequent carcinogenesis, predicated on classical high-dose toxicology, is that pollutants are simply not present in sufficient environmental quantities to lead to a human exposure level that could cause cancer. Another aspect is that many EDS are not genotoxic, which under the somatic mutation theory (SMT) (see entry ▶ Cancer Theories), Boveri (1929), would put a question mark over their ability to cause cancer. Such conclusions, however, are usually based on the logic of cancer induction toxicology performed on adult laboratory animals.

Human evidence of links between exposure to environmental pollutants, at background, and cancer is beginning to emerge, as illustrated in a paper by Hardell and Eriksson (2003) which showed a positive correlation between higher current levels of PCBs, HCB, and chlordanes (long lived pollutants) in mothers with sons who have developed testicular cancer. However, the majority of our current knowledge concerning low-dose fetal exposure to EDS and subsequent cancer development comes from animal studies.

The Endocrine Society has published a Scientific Statement on endocrine-disrupting chemicals (Diamanti-Kandarakis et al. 2009). When considering the role played by EDS in the etiology of breast cancer the report concludes that "*Collectively, these data support the notion that endocrine disruptors alter mammary gland morphogenesis and that the resulting dysgenic gland becomes more prone to neoplastic development*." The study in question, Murray et al. (2007), was the demonstration of the development of carcinoma in situ of the breasts in 33% of fetal rats exposed in the womb to the EDS bisphenol A between gestational day 9 and postnatal day 1, at concentrations 200 times lower than the current regulatory tolerable daily intake. The exposed rats developed such tumors in early adulthood, while there were no such changes in the breasts of the untreated control group. Morphological changes to breast architecture in mice were shown, Markey et al. (2001), at even lower doses.

The Endocrine Society Scientific Statement was less emphatic about the role of EDS in male genital tract tumors. However, the testicular dysgenesis syndrome (TDS) is of relevance. TDS consists of four interlinked conditions: cryptorchidism, hypospadias, oligospermia, and testicular germ cell cancer (TGCC). The incidence of all four of these conditions has been rising sharply over the past decades in many developed countries. TGCC is typically a cancer of younger men, the peak age range for incidence is between 25 and 40. When a child has a cryptorchid testis, then it is either brought down into the scrotum by orchidopexy or removed (orchidectomy). The reason for this is that an intra-abdominal undescended testis is dysgenic and carries a higher risk of developing TGCC.

Skakkebaek et al. (2001) examined bilateral biopsies, one from each testis, in cases of unilateral TGCC demonstrated carcinoma in situ in both testes. The conclusion drawn was that TGCC cancer is a developmental disease. They also demonstrated that hypofertile males were more likely to develop TGCC than average for the population.

Epidemiological evidence for the hypothesis that human exposure to EDCs causes TGCC remains scarce. There are some indications that exposure to PCBs and mono-butyl phthalate may be implicated but the evidence is not conclusive. Migration studies have shown that first generation migrants retain the incidence level of TGCC that prevails in their countries of birth, while second and subsequent generations had rates similar to their adoptive country. This gives strong evidence that TGCC is a developmental condition and that it is mediated via environmental factors.

Mechanisms of Cancer Induction

It is clearly beyond the scope of this entry to discuss all the evidence on EDSs and cancer induction. However, the point of the previous section has been to emphasize that a low-dose mechanism mediated through fetal exposure can and does lead to tissue dysgenesis.

This is essentially a non-genotoxic process. How then does it relate to the somatic mutation theory (see entry ▶ Cancer Theories)? Well, maybe not very much! An alternative hypothesis has been proposed by Sonnenschein and Soto (1999), the tissue organization field theory (TOFT), which provides a more logical explanation of low-dose fetal exposure mechanisms (see entry ▶ Cancer Theories). Under this hypothesis, it is proposed that, as in other life forms, all cells in the body are in a "default" state of proliferation, rather than "quiescence," as previously assumed. In multicellular organisms, where cells must exist in an ordered "society of cells," the desire to proliferate has become continuously suppressed by the evolution of inhibitory juxtacrine and autocrine (i.e., local hormonal) influences. This local control is intimately bound up with local architecture of the tissue at a microscopic level. The balance of the stroma to the parenchyma and the relationship of one cell type to another in 3D space are probably of high significance (Kaufman and Arnold 1996). The TOFT proposes that carcinogens alter tissue interactions, such as those occurring between the stroma and the epithelium, that in turn cause the architectural changes found in carcinomas, which, in turn, facilitates an increased local cell proliferation rate. The control of the microarchitecture of organs by epigenetic factors (external influences), throughout the developmental period, is a delicate process. Some of these factors are known but most are not fully understood. Hormones, which are known to be involved in the control of organogenesis, act in concentrations in the parts per trillion ranges. Xeno-chemicals, which may mimic hormones, are present in the populations of industrialized countries at similar concentrations. From our current knowledge base, they are likely to be inducing tissue dysgenesis in the fetus.

Complexity

Another feature of human exposure to potential carcinogens is the extreme complexity of the mixture of xeno-chemicals to which populations are exposed, which consists of hundreds of chemical groups. If these groups are broken down into individual compounds, then there are tens of thousands of individual chemicals in the mixture. For example, the organochlorine pesticide toxaphene, which in many classifications would be regarded simply as one chemical, has over 62,000 theoretically possible variants if all congeners and enantiomers (mirror image isomers) are taken into account (Vetter and Luckas 1995, 1998). Other groups of organic chemicals also have high numbers of variants. We simply do not possess the tools in toxicology to analyze complex mixtures. Even testing mixtures of two chemicals at a time is quite laborious (Axelrad et al. 2002). There are some bio-monitoring methods being developed for estimating the total dioxin-like activity or estrogen-like activity of mixtures (Murk et al. 1996; Soto et al. 1997), but these do not identify individual components of the mixture. However, many of these components are acknowledged to be carcinogenic or cancer promoters.

Under such circumstances, the power of epidemiology to determine outcome against exposure in human populations is rather weak and requires enormous biomonitoring programs involving hundreds of thousands of participants. Some such programs are getting under way but will not report on cancer outcomes for several decades. In the meantime, Sir Bradford Hill (1960) criteria on causation are worth revisiting. He gave a list of criteria that can be considered when addressing the causation of disease in occupational medicine. Now it is realized that they have much wider significance and applicability to other fields of medicine. It is not essential that all the criteria be fulfilled to establish causation. We discuss each Bradford Hill criteria with respect to the etiology of cancer from environmental pollution, in order of decreasing strength, as previously considered by Nicolopoulou-Stamati, Howard and Gaudet (2004).

1. *Temporal sequence*: The suspected causal agent must appear in the environment prior to the effect it is assumed to be causing. Preindustrialization levels of cancer were low. There has been a steady increase in cancer incidence following the start of industrialization. The incidence of certain hormone related cancers has increased rapidly following the introduction of EDSs into the general environment, and there is universal exposure of the population. This provides strong evidence of causality related to industrial development as measured by this criterion.
2. *Experimental evidence*: There is strong experimental evidence that many of the chemical pollutants in the environment and the food chain are carcinogenic in animal models. To this, we must add radiation exposure as a major influence, where the experimental evidence is overwhelming.

3. *Biological plausibility*: Exposure to environmental pollution commences pre-conceptually and continues throughout all stages of life. This exposure therefore covers vulnerable windows of development. There is very high biological plausibility that exposure to a mixture of known carcinogens would be causally related to the postindustrial increase in cancer incidence. Furthermore, cancer incidence would be age related as this is a measure of the length of exposure to these influences.
4. *Reasoning by analogy*: We can reason that pollutants that are known to cause cancer in animal models are likely to cause cancer in humans. This is indeed the basis of animal testing for the licensing of pharmaceutical and agrochemical agents, and therefore the analogy is widely accepted by governments.
5. *Coherence with biological background and previous knowledge*: We know that there has been environmental contamination with bioactive substances which, prior to their invention and production, had never been part of the environment. Animal detoxification systems are generally well adapted to natural biochemicals that they have coevolved with, which is often not the case with novel xeno-chemicals. Bioaccumulation is an example of such an incompatibility. In addition, the effects of low-dose endocrine-disrupting chemicals on development, a newly discovered mechanism, have to be considered in the context of non-genotoxic mechanisms of carcinogenesis. There is considerable prior biological knowledge to support a coherent theory for such exposures to be able to lead to cancer. The causation of cancer by environmental pollutants is in keeping with this criterion.
6. *Biological gradient (dose-response relationship)*: Dose-response relationships to chemical pollutants have been demonstrated in animal studies, though in the case of EDS they frequently are non-monotonic. Data on humans is weak, with many gaps in knowledge. There are many confounding factors, such as the complexity of the mixture to which the population is exposed. Dose-response relationships to radiation exposure are strong and well documented.
7. *Strength of association*: This is generally a weak criterion for general chemical pollution. Exceptions are, for example, the relationship between asbestos exposure and pleural mesothelioma and radiation and leukemia. However, apart from the undisputed rise in cancer incidence and its temporal sequence considered in association with the rise in environmental pollutant exposure, added to the small number of published reports discussed in the biological gradients section, there is in general weak association. However, this is largely attributable the lack of specificity of association with a complex mixture of influences.
8. *Specificity*: This is the weakest of the associations with the Bradford Hill criteria with respect to the environmental causation of cancer. The reasons for the lack of specificity are clear; humans are exposed to diffuse and multiple influences, including radiation, chemicals, lifestyle choices, and changes in dietary habits. Epidemiology, one of the main tools for making specific associations in human disease, is a relatively weak tool when a mixture is the causal factor under examination, given also that there is very little fetal exposure data. This is particularly true if the effect is to alter the frequency in the population of a common condition. The power of epidemiological studies is further weakened by the fact that there are no true control groups, that is, everybody has some degree of contamination with such pollutants. Therefore, although this criterion cannot be met, there are perfectly good reasons for understanding why it cannot be met, and they should not necessarily be used as a reason for avoiding the use of precaution. However, opponents of the theory of the environmental etiology of cancer adopt this very lack of specificity as the main plank of their arguments.

It appears therefore that five out of the eight Bradford Hill criteria support the argument that environmental pollutants are causally linked to the rise in human cancer incidence. Of the remaining three criteria, one offers some evidence, and the other two are weak. However, the reasons for this weakness are perfectly understandable, and the inability to detect any effect does not preclude its presence but is more likely a commentary on the unsatisfactory nature of the investigatory tools that we have at our disposal.

Conclusions

- The incidence of cancer in preindustrial societies was low.

- Since the beginning of the industrial revolution, there has been an inexorable rise in human cancer incidence, which is still continuing.
- The fact that environmental rather than genetic influences predominate in the causation of cancer has been shown through human twinning studies.
- Low-dose exposure to cell signaling disrupting chemicals during windows of vulnerability in the fetal and infant stages of development, resulting in tissue dysgenesis, appears to be an important influence in adult cancer formation.
- Cancer is a disease associated with industrialization. It is likely that there is a predominantly epigenetic (non-genotoxic mechanism) influence; this means that the process should be reversible if the causative agents can be removed from the environment.
- The proportion of cancers caused by "lifestyle factors" as opposed to "external factors" remains to be determined. However, the current state of knowledge indicates that the statement by the late Sir Richard Doll that only 1–4% of cancers can be attributed to environmental pollution is a gross underestimate. Even using Doll's estimate, the basis of which was never adequately explained would mean that many millions of cancers are already attributable to environmental chemical pollution. The likelihood is that a much higher proportion of human cancers, maybe a majority, is attributable to external factors.

Cross-References

► Neoplasia

References

Axelrad JC, Howard CV, McLean WG (2002) Interactions between pesticides and components of pesticide formulations in an in vitro neurotoxicity test. Toxicology 173:259–268

Bradford-Hill A (1966) The environment and disease: association or causation? Proc R Soc Med 58:295

Czene K, Lichtenstein P, Hemminki K (2002) Environmental and heritable causes of cancer among 9.6 million individuals in the Swedish Family-Cancer Database. Int J Cancer 99:260–266

Diamanti-Kandarakis E et al (2009) Endocrine-disrupting chemicals: an endocrine society scientific statement. Endocr Rev 30(4):293–342

Doll R, Peto R (1981) The causes of cancer. Oxford University Press, Oxford

Hardell L, Eriksson M (2003) Is the decline of the increasing incidence of non-Hodgkin lymphoma in Sweden and other countries a result of cancer preventive measures? Environ Health Perspect 111:1704–1706

Hoffman FL (1923) Cancer and civilization, speech to Belgian National Cancer Congress at Brussels (cited by Stefansson)

Howard CV, Newby JA (2004) Could the increase in cancer incidence be related to recent environmental changes? In: Nicolopoulou-Stamati P et al (eds), Cancer as an Environmental Disease. Kluwer Academic Publishers, pp 171–199

Irigaray P, Newby JA, Clapp R, Hardell L, Howard CV, Montagnier L, Epstein S, Belpomme D (2007) Lifestyle-related factors and environmental agents causing cancer: an overview. Biomed Pharmacother 61:640–658

Kaufman DG, Arnold JT (1996) Stromal-epithelial interactions in the normal and neoplastic development. In: Sirica AE (ed) Cell and molecular pathogenesis. Lippincott-Raven, Philadelphia, pp 403–432

Lichtenstein P, Holm NV, Vrkasalo PK (2000) Environmental and heritable factors in the causation of cancer. N Engl J Med 342:78–85

Markey CM, Luque EH, Munoz De Toro M, Sonnenschein C, Soto AM (2001) In utero exposure to bisphenol A alters the development and tissue organization of the mouse mammary gland. Biol Reprod 65:1215–1223

Murk AJ, Legler J, Denison MS, Giesy JP, van de Guchte C, Brouwer A (1996) Chemical-activated luciferase gene expression (CALUX): a novel in vitro bioassay for Ah receptor active compounds in sediments and pore water. Fundam Appl Toxicol 33(1):149–160

Murray TJ, Maffini MV, Ucci AA, Sonnenschein C, Soto AM (2007) Induction of mammary gland ductal hyperplasia and carcinoma in situ following fetal bisphenol A exposure. Reprod Toxicol 23:383–390

Newby JA, Busby CC, Howard CV, Platt MJ (2007) The cancer incidence temporality index: an index to show the temporal changes in the age of onset of overall and specific cancer (England and Wales, 1971–1999). Biomed Pharmacother 61:623–630

Nicolopoulou-Stamati N, Howard CV, Gaudet BAJ (2004) Re-evaluation of priorities in addressing the cancer issue: consclusions, strategies, prospects. In: Nicolopoulou-Stamati P et al (eds) Cancer as an Environmental Disease. Kluwer Academic Publishers, pp 171–199

Skakkebaek NE, Meyts ER-D, Main KM (2001) Testicular dysgenesis syndrome: an increasingly common developmental disorder with environmental impacts. Hum Reprod 16:972–978

Sonnenschein C, Soto AM (1999) The society of cells. Taylor & Francis, London, UK

Soto AM, Fernandez MF, Luizzi MF, Oles Karasko AS, Sonnenschein C (1997) Developing a marker of exposure to xenoestrogen mixtures in human serum. Environ Health Perspect 105:647–654

Stefansson V (1960) Cancer: disease of civilization? An anthropological and historical study. Hill and Wang, New York

Vetter W, Luckas B (1995) Theoretical aspects of polychlorinated bornanes and the composition of toxaphene

in technical mixtures and environmental samples. Sci Total Environ 160–161:505–510
Vetter W, Scherer G (1998) Variety, structures, GC properties, and persistence of compounds of technical toxaphene (CTTs). Chemosphere 37:2525–2543

Cancer Cell Cycle

▶ Cell Cycle, Cancer Cell Cycle and Oncogene Addiction

Cancer Cenes

▶ Cancer Networks

Cancer Drug Discovery

Reinhard Laubenbacher
Virginia Bioinformatics Institute, Virginia Tech, Blacksburg, VA, USA

Definition

Development of anticancer drugs provides a crucial component of the repertoire of therapeutic interventions, in addition to surgery and radiation therapy. While existing drugs have led to important treatment advances, much remains to be done to discover new drugs that effectively target key molecular mechanisms contributing to cancer pathogenesis. This is only possible through an understanding of the complex nonlinear molecular networks that mediate drug actions and are responsible for the de novo or acquired resistance of tumor cells to many existing drugs. The field of systems biology is focused on precisely such networks and therefore holds the promise of changing the paradigm of drug discovery from one focused on individual protein targets to one that targets networks instead.

Characteristics

As our understanding of the molecular basis of cancer grows, a picture emerges of a complex multi-faceted disease whose pathogenesis involves genetic mutations as well as epigenetic changes, leading to altered gene expression, signaling, and metabolic activity. The enormous individual variation in these changes adds to the challenge of developing effective targeted drugs for even a single known cancer type. Currently, the most common drug treatment for cancer is through the use of low-specificity cytotoxic chemotherapy drugs that affect different aspects of cellular function and whose effects are particularly damaging to fast-dividing cells. In many cases, tumor cells either are or become resistant to such drugs, resulting in limited efficacy. Another, more recent approach has been to develop drugs that target abnormalities specific to tumor cells, such as mutations in specific receptors. Several such drugs are in use that affect intracellular signaling pathways altered in cancer (see, e.g., (Alvarez 2010)). They take advantage of our increased knowledge about the link between genetic lesions and the capabilities normal cells acquire as they undergo the transition to become malignant cells.

Two relatively recent developments in biology research promise a paradigm change in cancer drug development: systems biology and cheap high-throughput sequencing. The fundamental relevance of systems biology to the understanding and treatment of cancer is the insight that genes and proteins do not act in isolation, but rather as nodes in complex interactive networks that include multiple feedback mechanisms and redundancies. The design of effective drugs to battle cancer will depend on the understanding of these networks and of the specific network alterations present in an individual tumor (Goodarzi et al. 2009; Erler and Linding 2010; Klipp et al. 2010). This will allow the discovery of tailor-made combination therapies that target specific network perturbations. High-throughput sequencing has made possible the identification of the genetic changes in an individual patient, which provide crucial information about how to target the network. Thus, systems biology is affecting a paradigm shift away from single proteins as the target of drugs toward the network as a target (Baggs et al. 2010). The result will be what has been called systems pharmacology (Yang et al. 2010), a systems biology approach to drug design. Its hallmark will be combination therapy, guided by an analysis of the complex signaling networks involved in malignant changes (Wu et al. 2010). Mathematical modeling and computer simulation are essential tools both for

the understanding of the nonlinear dynamic networks involved, and for the study of effective control mechanisms (Feala et al. 2010).

The two case studies below illustrate the systems biology approach to cancer drug design. The first, focused on breast cancer, shows how one can use systems biology to address the limited efficacy of existing drugs by identifying new additional targets that can be used to amplify their effect. The second, focused on melanoma, shows how an understanding of signaling networks can lead the way to appropriate choices for combination therapy.

CASE STUDY 1: Breast cancer drugs

There are currently three types of drug therapy available for breast cancer patients: conventional cytotoxic chemotherapy, hormone-based therapy, and targeted therapy based on monoclonal antibodies (Alvarez 2010) The first two drug types are *systemic*, in the sense that their effects are not specifically restricted to cancer cells. For instance, cytotoxic chemotherapy drugs inhibit cell division in some way or target DNA repair mechanisms or interfere with metabolism, and are particularly damaging to fast-dividing cells. While this includes cancer cells, in particular small, fast-growing tumors, it also includes normal cells in, e.g., the bone marrow, intestinal tract, and hair follicles. What fundamentally distinguishes targeted therapy drugs from the others is that they are based on cellular abnormalities specific to some cancer cells, and thereby represent an entirely new and more sophisticated class of chemotherapeutics. For breast cancer cells, an important molecular target that has been identified is the human epidermal growth factor receptor 2 (HER2). Systems biology can play a key role in identifying such targets and understanding the network of interconnected pathways that propagate their signals and their redundancy that helps cancer cells to develop resistance to drugs that interfere with their action. The example of the drug trastuzumab that targets the HER2 pathway and its role in controlling the cell cycle is instructive.

Members of a family of receptor tyrosine kinases, the ErbB family, are frequently overexpressed in epithelial tumors and promote tumor cell proliferation in several cancers, including breast cancer. Four homologous members of the HER receptor family have been identified to date, HER1-4 (also known as ErbB1-4). After ligand binding to ErbB family members 1 (EGFR) or 2 (HER2) the ErbB receptor is activated, followed by a cascade of phosphorylation events that regulate apoptosis and cellular proliferation pathways, mediated by cyclin-dependent kinases. The drug trastuzumab, approved by the U.S. Food and Drug Administration in 1998, is a recombinant humanized monoclonal antibody that binds with high affinity to the extracellular domain of the HER2 receptor, and is applied in patients that exhibit HER2 overexpression. However, at least two thirds of patients are de novo resistant to the effects of this drug. One possible explanation for this resistance on the cellular level could be that cancer cells are able to exploit the redundancy of the signaling network between EGFR and the proteins controlling cell cycle progression, primarily the retinoblastoma protein pRB, to avoid cell cycle arrest. Thus, additional targets within this network need to be identified in order to overcome resistance. In other words, instead of individual proteins, the network has to become the target.

In (Sahin et al. 2009) this is done through the use of a dynamic mathematical model of the network. A computer implementation of the model enables simulation studies that can explore the effect of targeting other proteins in the network for their therapeutic potential. New potential targets were identified, through in silico experiments that were then confirmed in the laboratory or the literature, including cyclin D1 and CDK4, as well as transcription factors c-MYC and ER-a. These could be exploited either instead of or in combination with HER2 to attack the entire network structure.

CASE STUDY 2: Melanoma drugs

Melanoma is a very aggressive cancer with few effective treatment options. However, a recent increased understanding of the genetic events underlying melanoma pathogenesis and the signaling pathways affected by mutations common in melanoma has opened the door to a systems biology approach to the discovery of targeted therapeutics. The review (Ko and Fisher 2010) surveys some of these developments and makes clear the need for combination therapies that target the entire signaling network rather than individual pathways, similar to the case of breast cancer described above.

Several signaling cascades have been discovered as important in melanoma. One is the RAS-RAF-MAPK-ERK growth factor pathway. In virtual all human

melanoma cases this signaling cascade has been altered in some way, in particular through NRAS or BRAF mutations. Several agents targeting proteins in this pathway are available, such as the BRAF inhibitor PLX4032, currently in clinical trials. Another pathway, also subject to RAS signaling, is the PI3K-AKT-mTOR pathway, whose activation has the effect of suppressing apoptosis. Several small molecule inhibitors are available that target mTOR signaling. It is thought that none of the actors in these pathways stand alone in the pathogenesis of melanoma, and it is therefore likely that no single-agent approach to this disease will be successful in affecting a cure. The signaling network including the proteins in the two pathways mentioned here is complex, with feedback loops and redundant pathways, thus, combination therapies targeting several network nodes at the same time hold the most promise for success. Here too, detailed computer models of the network are essential tools for the discovery of effective drug combinations. Probably, the computer model will ultimately have to be enlarged to also take into account other apoptosis pathways as well as growth factor signaling.

Discussion

The experience with cancer drugs that target individual proteins suggests that combination therapies have the greatest promise of being effective. To identify the right combination of targets for a particular tumor it is essential to understand the signaling network that is perturbed through mutations and that mediates the drugs' action. It is the aim of systems biology to provide the link between an organism's genomic sequence and its phenotype through an understanding of the various dynamic networks that connect the two. Thus, a systems biology approach to drug discovery and design of combinatorial therapies is indispensable.

Cross-References

- ▶ Adult T-Cell Leukemia
- ▶ Apoptosis
- ▶ Cdk Inhibitors
- ▶ Cell Cycle Checkpoints
- ▶ Cell Cycle, Cancer Cell Cycle and Oncogene Addiction
- ▶ c-Myc
- ▶ Interaction Networks
- ▶ Network-Based Biomarkers
- ▶ Signal Transduction Pathway
- ▶ Systems Medicine
- ▶ Systems Pharmacology, Drug-Target Networks

References

Alvarez RH (2010) Present and future evolution of advanced breast cancer therapy. Breast Cancer Res 12(Suppl 2):S1

Baggs JE, Hughes ME et al (2010) The network as the target. Wiley Interdiscip Rev Syst Biol Med 2(2):127–133

Erler JT, Linding R (2010) Network-based drugs and biomarkers. J Pathol 220(2):290–296

Feala JD, Cortes J et al (2010) Systems approaches and algorithms for discovery of combinatorial therapies. Wiley Interdiscip Rev Syst Biol Med 2(2):181–193

Goodarzi H, Elemento O et al (2009) Revealing global regulatory perturbations across human cancers. Mol Cell 36(5):900–911

Klipp E, Wade RC et al (2010) Biochemical network-based drug-target prediction. Curr Opin Biotechnol 21(4):511–516

Ko JM, Fisher DE (2010) A new era: melanoma genetics and therapeutics. J Pathol 223(2):241–250

Sahin O, Frohlich H et al (2009) Modeling ERBB receptor-regulated G1/S transition to find novel targets for de novo trastuzumab resistance. BMC Syst Biol 3:1

Wu Z, Zhao XM et al (2010) A systems biology approach to identify effective cocktail drugs. BMC Syst Biol 4(Suppl 2):S7

Yang R, Niepel M et al (2010) Dissecting variability in responses to cancer chemotherapy through systems pharmacology. Clin Pharmacol Ther 88(1):34–38

Cancer Networks

Eric Werner
Department of Physiology, Anatomy and Genetics, University of Oxford, Oxford, UK
Department of Computer Science, University of Oxford, Oxford, UK
Oxford Advanced Research Foundation, Fort Myers, FL, USA

Synonyms

Cancer cenes; Cenes; Developmental cancer networks

Definition

A cancer network is a developmental network (▶ Developmental Control Networks) or ▶ cene that contains one or more self-regenerating loops resulting in endless cell proliferation. The nodes in the network are cell states. The branches denote cell division or jump to a new cell state.

Characteristics

Cancer networks come in two basic forms: cancer ▶ stem cell networks and exponential networks, where cancer stem cell networks are further divided into linear stem cell networks and more general geometric networks:

- Cancer stem cell networks
  - Linear cancer networks
  - Geometric cancer networks
- Exponential networks

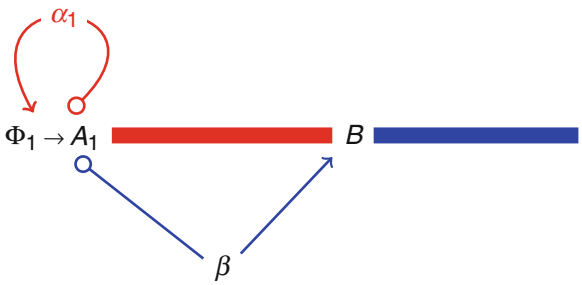
Cancer networks like all ▶ developmental control networks can be further divided into deterministic and stochastic networks. Networks may also involve cell signaling. Based on the topology of their ▶ developmental control networks, there exist a vast number of possible cancer networks. The monograph (Werner 2011b) gives a kind of periodic table of all possible cancer networks together with their salient proliferative and phenotypic properties.

Linear Cancer Networks

A linear cancer stem cell is a self-regenerating cancer cell that generates other terminal progenitor cells. If the network allows stochastic dedifferentiation then these terminal cells may dedifferentiate to cancer stem cells. A linear cancer network G_1 contains one self-regenerating loop and one link to a terminal developmental network. They have many of the characteristics of linear ▶ stem cell networks (Werner 2011b; Fig. 1).

Cancer Meta-Stem Cells

A cancer meta-stem cell or second-order cancer stem cell, G_2, is a cancer stem cell that generates other cancer stem cells. They have many of the characteristics of normal meta-stem cell developmental networks (Werner 2011b). See ▶ Stem Cell Networks (Fig. 2).



Cancer Networks, Fig. 1 A linear, first-order cancer stem cell network. It generates cells B if the condition Φ_1 is met (Werner 2011b). In cancer, the condition Φ_1 may be locked into being satisfied, so that we have endless cell proliferation.

Stem Cell Hierarchy

Stem cell networks form a natural hierarchy based on the number of loops in their networks. For this reason, we distinguish between first-order, second-order, third-order, and further higher-order ▶ stem cell networks.

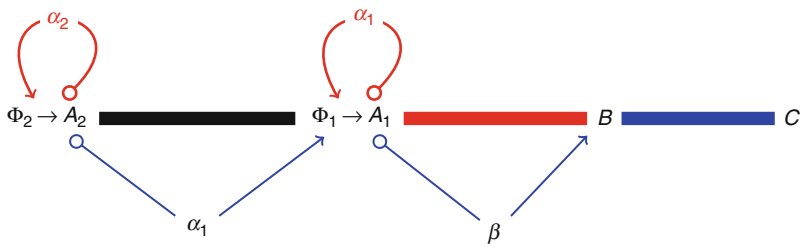
Geometric Networks

We call these networks geometric (control) networks because their proliferative properties are related to properties of geometric numbers (also called the figurative numbers). They are also related to the coefficients of Pascal's triangle (Werner 2011b; Fig. 3).

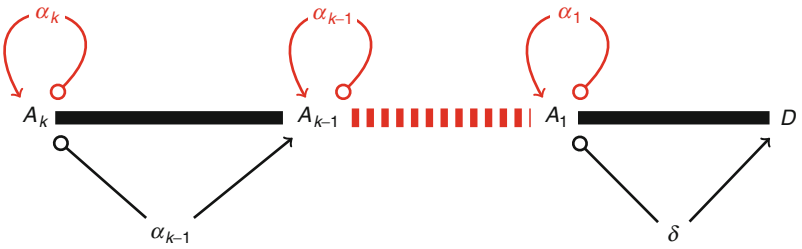
The general case of k loops in cancer ▶ stem cell networks: A higher-order geometric network G_k containing k loops and one link to a terminal developmental network after n rounds of synchronous divisions has an ideal proliferation rate that given by the following formula, when $n > 0$:

$$Cells(n,k) = 1 + n + \frac{n(n-1)}{2} + \frac{n(n-1)(n-2)}{6} + \ldots + \frac{n!}{k!(n-k)!} \quad (1)$$

$$Cells(n,k) = \sum_{i=0}^{k} \binom{n}{i} \quad (2)$$



Cancer Networks, Fig. 2 A cancer meta-stem cell network is a second-order geometric network. It consists of a second-order loop at A_2 linked to a first-order loop at A_1 (Werner 2011b). A cell in state A_2 is a cancer meta-stem cell that generates first-order linear cancer stem cells. In cancer, one or both of the conditions Φ_i may be locked into being satisfied, so that we have endless cell proliferation. In an ideal space with no resistance, synchronous divisions, and when both conditions Φ_i are satisfied, this network generates a *triangle*



Cancer Networks, Fig. 3 A kth-order stem cell network (Werner 2011b). The network contains k loops at control states $A_k \ldots A_1$ and end in a terminal developmental network D

Limited Exponential Growth with Geometric Networks

When the number of rounds of division n is less than the number of loops k, i.e., $0 \leq n \leq k$, then a geometric stem network exhibits exponential growth since the following holds:

$$\sum_{i=0}^{n} \binom{n}{i} = \binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \ldots + \binom{n}{n} = 2^n \tag{3}$$

Metastatic Hierarchy

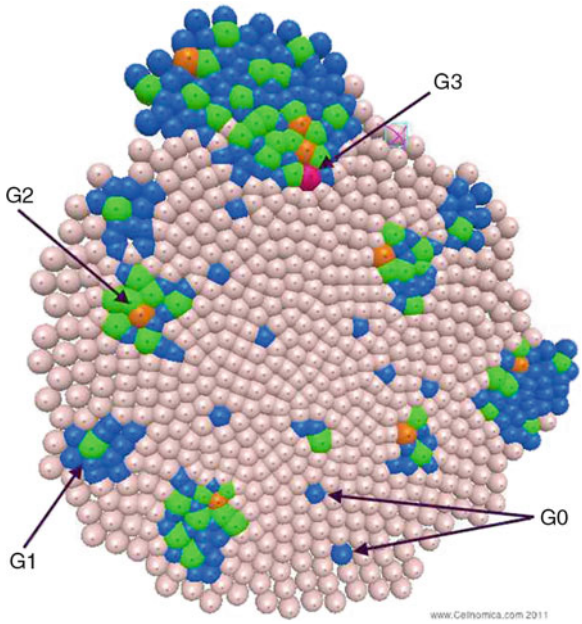
The stem cell network hierarchy generates a hierarchy of cancer metastases (Werner 2011b). A first-order cancer stem cell generates terminal cells. A second-order cancer stem cell generates first-order stem cells which then generate terminal progenitor cells. A third-order stem cell generates second-order stem cells which generate first-order stem cells which generate terminal progenitor cells. Hence, in a given tumor formed by a third-order stem cell, we will see second-order stem cells and first-order stem cells with a majority being terminal or terminal progenitor cells (Fig. 4).

First-order cancer stem cells are slow growing with linear proliferation rates and can be relatively harmless if there is no stochastic dedifferentiation. A second-order stem cell tumor grows more quickly (adding the triangular number) with an ideal proliferation rate given by the following formula, given $n > 0$:

$$Cells(n, 2) = 1 + n + \frac{n(n-1)}{2} \tag{4}$$

A third-order stem cell tumor grows even more quickly (adding the tetrahedral number) with an ideal proliferation rate as follows:

$$Cells(n, 3) = 1 + n + \frac{n(n-1)}{2} + \frac{n(n-1)(n-2)}{6} \tag{5}$$



Cancer Networks, Fig. 4 A metastatic hierarchy with one red G3 stem cell, and the rest are *orange* G2, *green* G1, and *blue* terminal cells growing in a background of light beige cells. For details, see Werner 2011b

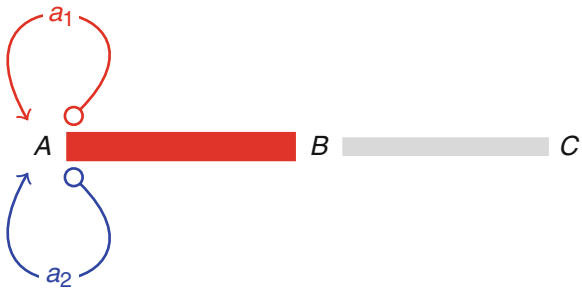
Tumor Properties of Stem Cell Networks

The tumor has interesting properties. Initially, the growth rate is exponential for all cells whose division cycle is less than or equal to the number of loops in the stem cell network. Then, the higher the number of loops, the greater the percentage of cells in the tumor that are actively proliferating. For example, with a tumor generated by a stem cell controlled by a third-order stem cell network, a high percentage of the cells will be proliferating (a ratio that relates the volume of a tetrahedron to the area of the base of that tetrahedron).

Exponential Networks

An exponential network *X* contains two control loops from two daughter cells that lead back to the parent cell (Fig. 5).

As with linear and geometric networks, exponential networks can have a great variety of forms depending on the subnetworks they contain and the other networks that they activate.



Cancer Networks, Fig. 5 An exponential cancer network where a cell in control state A divides into daughter cells that self-loop and differentiate to control state A. For details, see Werner 2011b

Subnetworks and Tangentially Activated Networks

Any loop can include a path that is a subpath of another developmental network. This means the loop has further developmental progeny generated by that subpath. This ability to link to subnetworks and other developmental networks is a basic characteristic of ▶ developmental networks (Werner 2011a, b). These tangentially activated networks can generate the dominant phenotype of a tumor or cancer. For example, with meta-stem cell cancer networks, the dominant phenotype may consist of terminal or terminal progenitor cells that are non-cancerous. Teratoma tumors also may result from a small set of cancer cells with the vast majority being terminal cells of various tissue types.

Conclusion

Cancer networks can exist in a vast space of ▶ developmental networks. The space of all possible cancer networks is described in Werner 2011b.

Cross-References

- ▶ Cene
- ▶ Developmental Control Networks
- ▶ Stem Cell Networks

References

Werner E (2011a) On programs and genomes, arXiv:1110.5265v1 [q-bio.OT]. http://arxiv.org/abs/1110.5265

Werner E (2011b) Cancer networks: a general theoretical and computational framework for understanding cancer, arXiv:1110.5865v1 [q-bio.MN]. http://arxiv.org/abs/1110.5865v1

Cancer Pathology

Barbara J. Davis
Section of Pathology, Tufts Cummings School of Veterinary Medicine Biomedical Sciences, North Grafton, MA, USA

Characteristics

Cancer pathology is the study of structural (morphology), biochemical, molecular, or genotypic alterations and the functional or clinical manifestations associated with *disorders of growth* of tissues. Tumors are defined by disorganized tissue architecture and excessive cell proliferation. A tumor, translated as "swelling," is a general term used in pathology to designate a mass associated with disorders of growth (Fig. 1). The suffix "oma" is used to designate a mass of tissue or tumor. Tumors may be "congenital" or develop after birth.

Congenital tumors are classified as ▶ hamartomas or ▶ choristomas.

Tumors that represent disturbances of growth after birth are classified as ▶ hyperplasia, ▶ dysplasia, or ▶ neoplasia (Fig. 1).

A neoplastic growth may be ▶ benign or ▶ malignant.

A malignant neoplasm (or may also be referred to as a malignant tumor) is generally synonymous with the term "cancer." *That is, the use of the term "cancer" in medicine refers to a malignant neoplasm*. The spread of a malignant cancer is referred to as ▶ metastasis.

Tumors are further defined by their cell and tissue of origin (Fig. 2).

Solid Tumors

Solid Tumors are tissue-based masses and represent an imitation of the tissue and tissue components they are derived from including neoplastic cells, supporting connective tissue and mesenteries, blood vessels, lymphatics, nerves, and immune cells. Solid tumors are subclassified generally based on the tissue and embryological cell of origin into **epithelial, mesenchymal, or mixed tumors**. Embryologically, epithelial tumors are those derived from ectoderm and endoderm. Mesenchymal tumors are derived from mesoderm or neural crest cells. The distinction is important because the biological behavior and the course of treatment vary with cell type.

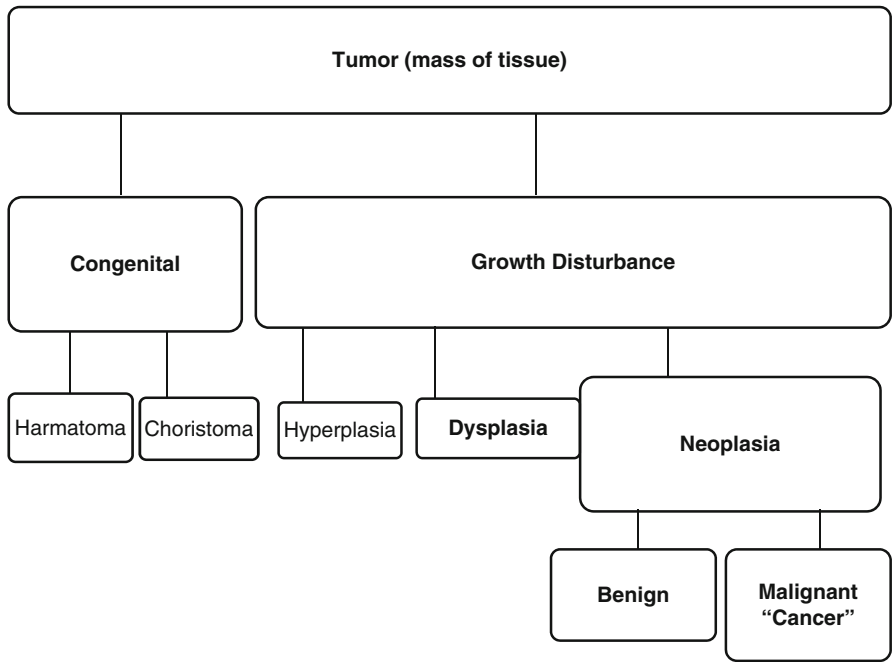
The *nomenclature* of the tumor is then defined by the origin of the cell and whether it is benign or malignant (Table 1).

Benign epithelial tumors are called adenomas, papillomas, polyps; *malignant epithelial cancers* are called carcinomas. For further clarification, the tissue or cell of origin is also included in the name. For example, a malignant cancer of the cells that form the liver parenchyma, or the hepatocytes, is called a hepatocellular carcinoma, while a cancer of the cells that form the bile transport system in the liver, or the bile ducts and cholangioles, is called a cholangiocellular carcinoma.

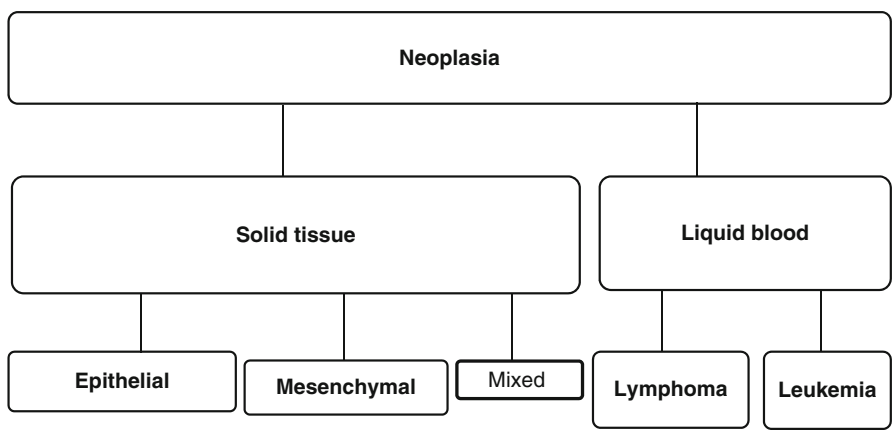
Benign mesenchymal tumors are named with reference to the tissue of origin. For example, a benign mesenchymal tumor of connective tissue is called a "fibroma." A benign tumor of smooth muscle origin is a leiomyoma. *Malignant mesenchymal cancers* are called sarcomas. A malignant cancer of connective tissue is called a fibrosarcoma. A malignant cancer of a smooth muscle cell is a leiomyosarcoma.

Mixed tumors are tumors in which the progenitor cells develop features of both epithelial and mesenchymal components and are named according to tissue and behavior. For example, a tumor of benign mixed tumor of a mammary (breast) gland may be called a fibroadenomas or benign mixed tumor; a malignant mixed tumor is called a carcinosarcoma.

There is also special consideration given to tumors derived from subtypes of cells within tissues such as those from *neuroendocrine cells*. These are more common in the gastrointestinal tract, but also found in respiratory tract, and other sites. Neuroendocrine tumors are called "carcinoids" and may be considered benign, but most are considered malignant or with malignant potential.



Cancer Pathology, Fig. 1 Classification scheme for identification of masses or "tumors"



Cancer Pathology, Fig. 2 Classification of tumors based on origin

Neural crest cells give rise to pigmented melanin-producing cells called melanocytes. A benign tumor of melanocytes is called a "nevus" or melanocytoma. A malignant melanocytic cancer is called a melanoma or malignant melanoma.

Special consideration and nomenclature is also used for tumors of specialized tissue such as the nervous system and the gonads. In the brain, tumors may be classified as neuroepithelial and non-neuroepithelial tumors and have special stains to differentiate their origin. In the gonads, the totipotent germ cells may give rise to tumors of embryonic origin such as teratomas, choriocarcinomas, or even yolk sac tumors.

Hematopoietic tumors

Hematopoietic tumors are also referred to as "liquid tumors" and are of blood-based origin. This category includes cells that make up the immune white blood cell system such as the lymphocytes and granulocytes and red blood cell systems such as erythrocytes and megakaryocytes. Terms such as hyperplasia,

Cancer Pathology, Table 1 Examples of tumor nomenclature by tissue of origin and behavior

		Benign	Malignant
Epithelial		Adenoma	Carcinoma
		Papilloma	
		Polyp	
Of tissue	*Skin*	Papilloma	Squamous cell carcinoma
			Basal cell carcinoma
	Gastrointestinal tract	Polyp or adenoma	Adenocarcinoma
	Bladder	Polyp	Transitional cell carcinoma
	Kidney	Oncocytoma	Renal cell carcinoma
			Clear cell variant
			Papillary carcinoma
		Juxtaglomerlar Tumor	
			Renal pelvis squamous cell carcinoma
	Liver hepatocytes	Hepatoma	Hepatocellular carcinoma
	Liver bile duct	Cholangioma	Cholangiocarcinoma
Mesenchymal		"oma"	Sarcoma
Of tissue	*Connective tissue*	Fibroma	Fibrosarcoma
	Smooth muscle	Leiomyoma	Leiomyosarcoma
	Skeletal muscle	Rhabdomyoma	Rhadomyosarcoma
	Endothelial cells	Hemangioma/angioma	Hemangiosarcoma or angiosarcoma
	Cartilage	Chondroma	Chondrosarcoma
	Bone	Osteoma	Osteosarcoma
CNS	**Neuroepithelial**		
	Astrocytes	Astrocyoma	Anaplastic astrocytoma
	Glial cells		Glioblastoma multiforme
	Oligodendrocytes	Oligodendroglioma	Anaplastic olgiodendroglioma
	Ependymal cells	Ependymoma	Anaplastic epednymoma
	Choriod plexus	Papilloma	Carcinoma
	Embryonic origin		Medulloepithelioma

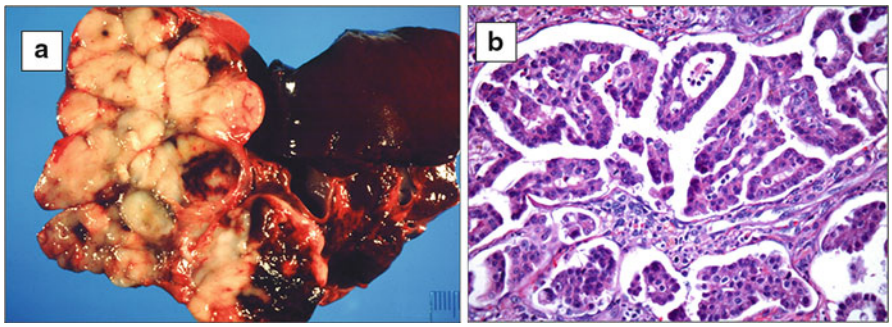
dysplasia, and neoplasia are similarly used to define disorders of altered architecture of these lineages as well. Cancer of lymphoid organs is termed lymphoma. Cancer of circulating bone marrow cells is termed leukemia.

Diagnosis

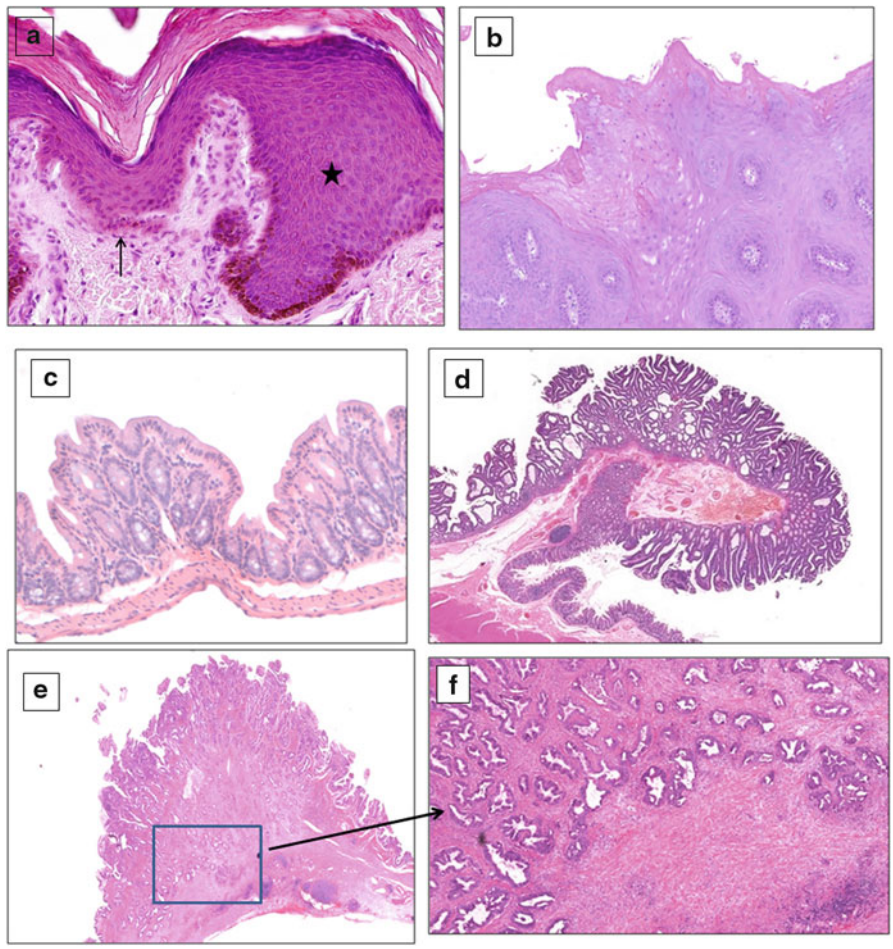
The *diagnosis* of a mass is based on the gross appearance and behavior of the tumor and the histological appearance (Fig. 3).

For microscopic assessment, the tumor is sampled during an in-life surgical procedure called a *biopsy* or after death as an *autopsy*. The preparation of the tumor includes fixation of the mass in a *preservative*; processing the tissue through a series of alcohol dehydration steps to remove water and replace it with a transparent harder material which is usually wax paraffin; *embedding* the tissue in a mold; *sectioning* or "trimming" the mold with the tissue; and then placing the tissue on a glass slide. The tissue will then be stained for visual examination. Preservatives include *10% buffered formalin*, which is the most common routinely used tissue preservative. Other preservatives that may be used are methanol or 80% alcohol, 2% or 4% paraformaldehyde, glutaraldehyde, and "special" fixatives such as Bouin's fixative, which includes picric acid. Tissues may also be *frozen* in a container in liquid nitrogen. The tissue is then stained – hemotoxylin and eosin (or H&E) staining is most commonly used on a routine basis.

The determination of whether the mass is hyperplastic, dysplastic, or neoplastic, of epithelial, mesenchymal, or mixed (or "other") origin, and whether it has features of a benign or malignant tumor is based on the microscopic pattern using criteria including the tissue of origin, level of organization, or

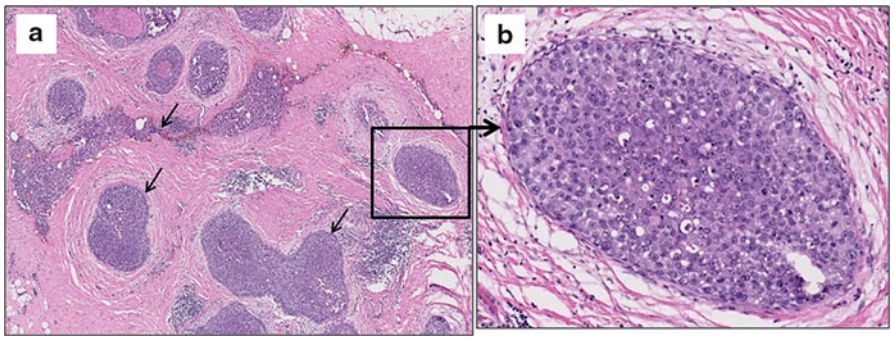


Cancer Pathology, Fig. 3 (**a**) Photograph of liver with a tan irregularly nodular and cavitated mass. A section of mass is dissected, fixed in a preservative (formalin), and processed to make a tissue section and stained for microscopic examination. (**b**) Photomicrograph of a tissue section of the mass in (**a**) stained with hemotoxylin and eosin (H&E). The pattern and location is consistent with a diagnosis of cholangiocellular carcinoma, a neoplasm arising from the biliary epithelium



Cancer Pathology, Fig. 4 Photographs of examples of (**a**) normal (*arrow*) and hyperplastic (*) and (**b**) dysplastic squamous epithelium of mucosa; (**c**) Normal mucosa intestinal epithelium compared to (**c**) adenoma or polyp characterized by disorganized proliferation of epithelium supported by a fibrovascular stalk of connective tissue; and (**d**). Adenocarcinoma characterized by disorganized proliferation of pleomorphic epithelium cells invading into the subjacent submucosa and muscular wall of the intestine (*box*) and (**e**) higher magnification of invading carcinoma

differentiation and level of containment (Fig. 4). The mass is identified and described as to whether it has a connective tissue capsule or not, whether it is compressing adjacent tissue or invading adjacent tissue, and whether the pattern most closely resembling the tissue or cell of origin. For example, neoplastic epithelial cells tend to arrange as glands of acini or tubules, nests or islands. Neoplastic



Cancer Pathology, Fig. 5 Photograph of a section of H&E stained breast tissue diagnosed as carcinoma in situ (**a**) with intraductular but not invasive masses of disorganized pleomorphic epithelial cells that are confined to the basement membrane boxed and shown in (**b**)

mesenchymal cells tend to arrange in interlacing or intersecting bundles, whorls, or palisades. In addition to the overall pattern, the morphology of the cells is also important in the diagnosis. Neoplastic epithelial cells are generally cuboidal, columnar, or polygonal; neoplastic mesenchymal cells are generally elongated or "spindle-shaped." Additional criteria include ▶ anaplasia, cell ▶ pleomorphism, nuclear to cytoplasmic ratios, the appearance of the nucleus, ▶ anisocytosis, ▶ anisokaryosis, ▶ mitotic figures, and their appearance (normal alignment or bizarre). Characteristics such as how fast the tumor grew, whether there were previous tumor diagnoses, or whether the mass represents a "re-growth" of a previously excised tumor are also important in the consideration of the diagnosis.

Benign tumors are considered well differentiated, encapsuled with usually scant and normal mitotic figures, and usually normal mitotic figures. Malignant tumors breech the basement membrane if epithelial in origin and show local invasion, lymphatic or vascular invasion, and destruction of surrounding tissue. A few malignant tumors do not metastasize like basal cell carcinoma or gliomas, and sarcomas tend to be locally invasive. ▶ Carcinoma in situ also has characteristics of malignancy but does not invade the local tissue (Fig. 5). Tumor metastasis may occur as seeding of peritoneal cavity or other spaces (ovarian carcinomas, mesothelioma); spread through lymphatics (most common); or invasion and spread through the vasculature. The liver and lungs are common sites for metastasis because of blood flow to these organs through first pass and capillary beds.

In addition, special enzyme-based stains can be used to reveal cellular components. Staining tissues with antibodies for a colorimetric process, called ▶ immunohistochemistry is also commonly used to further define the tissue or cell of origins.

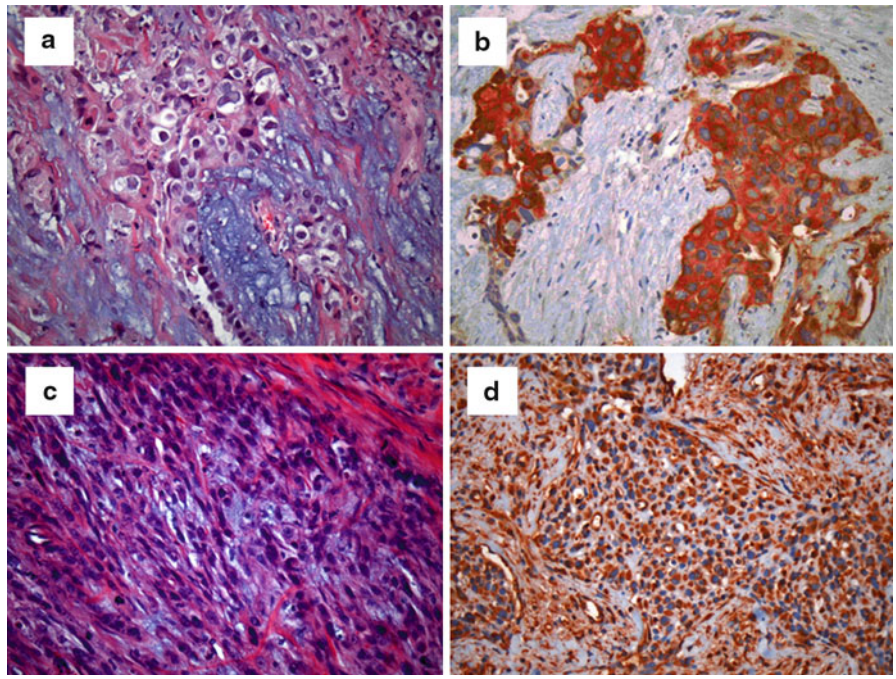
For the diagnosis of solid tumors, IHC is used to differentiate epithelial tumors and mesenchymal tumors based on the type of intermediate filaments. Intermediate filaments are structural proteins that help determine the cell shape. Neoplastic epithelial cells retain cytokeratins and react with antibodies directed at these intermediate filaments (Fig. 6). Neoplastic mesenchymal cells retain vimentin microfilaments. Mixed tumors will show staining for both cytokeratins and vimentin. For hematopoietic tumors, IHC is used routinely to differentiate lymphocytic tumors (lymphoma) derived from B-cells or T-cells (Fig. 7).

IHC methods have also been developed as diagnostic and prognostic indicators such as in breast cancers where the status of estrogen receptor, progesterone receptor, and HER2/neu status is determined prior to and in support of the type of therapy that will be used.

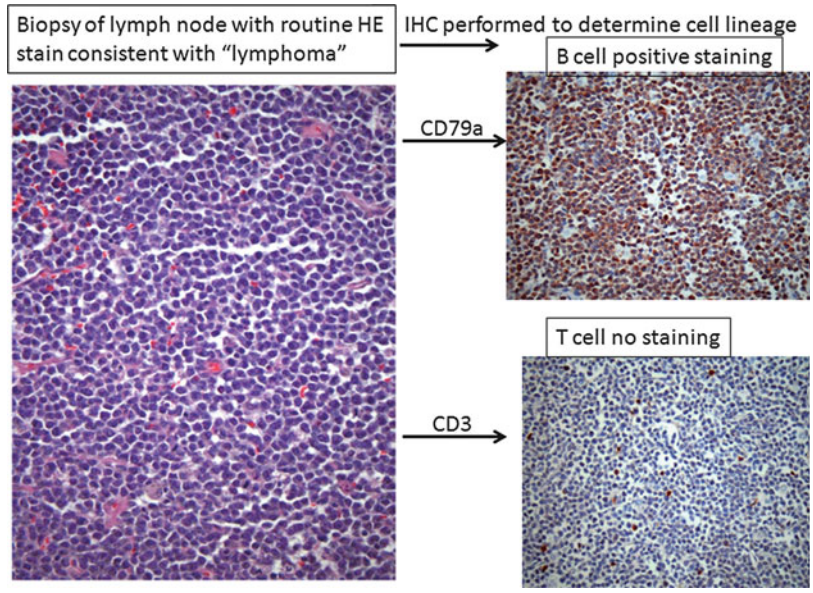
Evaluation of the ultrastructural features of cells via electron microscopy is also valuable in diagnostic evaluation of the origin of some samples that do not have characteristic features or for which enzyme or immunohistochemistry are not diagnostic.

Grading and Staging

A **grade** is the qualitative or semiqualitative determination or assessment of the degree of differentiation of the tumor. It includes a number of mitosis, cellular cytoplasmic, and nuclear features. Grading schemes developed over time for each tumor type when appropriate. Schemes may be in two categories – "high" grade or "low" grade or low, intermediate, or high grade. Grading is subjective and has proven less valuable clinically than staging.



Cancer Pathology, Fig. 6 Photographs of biopsies from (**a**) H&E stained section of a poorly differentiated neoplastic cancer confirmed to be a carcinoma based on identification of (**b**) positive brown cytoplasmic staining in a section stained for cytokeratin intermediate filaments using immunohistochemistry; (**c**) H&E stained section of a intersecting bundles of spindle-shaped cells confirmed to be a sarcoma based on identification of (**d**) positive brown cytoplasmic staining in a section stained for vimentin intermediate filaments using immunohistochemistry



Cancer Pathology, Fig. 7 Photograph of a H&E stained biopsy from an enlarged lymph node consistent with lymphoma and confirmation of subtype of a B-cell origin using immunohistochemistry with antibodies to CD79a, which is expressed on B-cells, or CD3, which is expressed on T-cells but not B-cells

The **stage** of tumor is based on the size of the primary tumor and the extent and spread to regional lymph nodes or blood-borne metastasis. The staging of tumor is designated by the abbreviation "TNM" for Tumor size, regional Node, and Metastasis followed by a numerical assignment for each category. The designations are: T1-T4; N1-N3; and M1, 2.

For carcinoma in situ, the stage would be designated as T0.

References

Rosai J (2011) Rosai and Akerman's surgical pathology, 10th edn. Mosby Elsevier, Edinburgh

Striker T, Kumar V (2009) Neoplasia. In: Aster J, Abbas A, Kumar V, Fausto N (eds) Robbins and Cotran pathological basis of disease, 8th edn. Saunders Elsevier, Philadelphia, pp 259–330

World Health Organization (WHO) Classification of Tumours. Published by Interational Agency for Research on Cancer.

Cancer Resources

Jingky Lozano-Kühne
Department of Public Health, University of Oxford, Oxford, UK

Definition

► Cancer resources provide information about cancer, researches and clinical trials, and tools in cancer research and analysis. Other cancer resources that are intended for patients and caregivers provide information on cancer pain management, support groups, financial assistance, and palliative care. The selected cancer resources described here focus on resources for health professionals, researchers, and students in the field of systems biology. These resources are available in different formats. There are printed publications, online references, databases, and also software packages for cancer data management and research analysis.

Characteristics

Printed Publications and Online References

Basic information about cancer can be found in many medical textbooks and other references. As a reference specific on ► Cancer Systems Biology, the book of Wang (2010) provides not only basic concepts on cancer and systems biology but also applications, information on software tools, data resources, and a significant collection of research updates. This is the first book written specially for computational and experimental biologists. Another reference is that of Cesario and Marcus (2011) on cancer systems biology, bioinformatics, and medicine. This book describes systems approaches to cancer research and spans a variety of topics from laboratory to clinical aspects of cancer. In some publications (e.g., Yan 2010), cancer systems biology is included as a major book chapter.

Several journals (e.g., BMC Systems Biology, Cancer Discovery) provide information on latest results about systems biology and cancer research. BMC Systems Biology (http://www.biomedcentral.com/bmcsystbiol/) is an open access journal published by BioMed Central. It includes original peer-reviewed research articles on the functions of biological systems. The Cancer Discovery journal (http://cancerdiscovery.aacrjournals.org/) is a new cancer information resource that also publishes review articles aside from articles on major advances in cancer research and clinical trials. Monographs and published articles about cancer research are also provided by the International Agency for Research on Cancer (IARC) which is part of the World Health Organization. Publications of the IARC can be freely downloaded from http://www.iarc.fr/.

One comprehensive online resource on cancer is the Website of the National Cancer Institute (NCI, http://www.cancer.gov/) which provides information on the different types of cancer, clinical trials, and cancer statistics. Beyond information, NCI also provides funding and support to cancer researches and also conducts its own laboratory and clinical studies. Another useful Website is the Cancer Systems Biology Resource of the University of Utah (https://cancersystemsbiology.utah.edu/) which provides links to topics such as basic gene and protein information, ► Pathways, tissue distribution and ► Gene Expression information among others.

There are numerous printed and online publications on cancer in addition to the resources described above. Many of them share common references and links as the ones already mentioned. One can also consult the ► Disease Databases and specific cancer (e.g., http://www.breastcancer.org/) sites for searching additional details about cancer.

Cancer Databases

Cancer databases may contain collated information about different types of cancers, cancer genes, or other clinical data related to cancer. As data sources,

they are useful especially for conducting research on analysis methods. One commonly used cancer database is the Cancer Chromosomes Database which will be soon part of the dbVar database, a database of genomic structural variation. Both databases are maintained by the United States National Center for Biotechnology Information (NCBI). The Cancer Chromosome Database is integrated into the NCBI ▶ Entrez system and is actually composed of three sub-databases containing cytogenetic (▶ Cytogenetics), clinical, and/or reference information.

The National Cancer Institute (NCI) provides access to cancer data and also tools for its analysis in their statistics' web page http://www.cancer.gov/statistics/tools. For studies on cancer incidence and population data associated by age, sex, race, year of cancer diagnosis, and geographic areas, their Surveillance, Epidemiology, and End Results (SEER, NCI 2005) data are commonly used. For data that are more specific to genes and pathways, the data portal of the International Cancer Genome Consortium (ICGC, http://www.icgc.org/) would be a useful source. The Cancer Genome Atlas (TCGA, http://cancergenome.nih.gov/) also provides a data portal for researchers on cancer. The data available at TCGA contains clinical information, ▶ Histopathology slide images, and molecular information of high-quality tumor samples. For pathway analysis, the ▶ KEGG Pathway database is a useful resource. For interests in mutation data and sequencing (▶ DNA Sequencing) data, the COSMIC database (Forbes et al. 2008) which can be downloaded from http://www.sanger.ac.uk/genetics/CGP/cosmic/ is a good reference. The Website also provides other cancer-related data and information. For other data on mutations, one can also check the SH2base site (http://bioinf.uta.fi/SH2base/).

Omics data or data in the field of biology ending in "–omics" (e.g., ▶ Pharmacogenomics, ▶ Proteomics) are also commonly used in the study of cancer and systems biology. Omics data enable the study of complex pathways in cancer etiology. Example of online resources that provide information on omics data are the Biodatabase (http://biodatabase.org/) and the Omics World Websites (http://www.omicsworld.com/). The Biodatabase contains links to cancer gene databases, protein, genomic databases, and others, while the Omics World provides links to resources on genomics, genetics, experimental protocols, reagents, instruments, and also software.

There are more data sources available online for cancer research. An attempt to collate and interconnect all these data was done by the National Cancer Institute (NCI) through the cancer Biomedical Informatics Grid (caBIG®, https://cabig.nci.nih.gov/). It was launched by the National Cancer Institute to create a virtual network of interconnected cancer data and people working together in cancer research. The caBIG® community Website has tools and facility for data sharing and also workspaces where members can conduct virtual activities, discussions, and exchange of ideas. Membership to the network is open to the public for free.

Software for Data Management and Analysis

Many of the cancer data providers previously mentioned also provide data management and analysis tools. For example, the SEER Website offers analysis software for the SEER data and other cancer-related databases for studying the impact of cancer in the population. The ICGC Data Portal also runs a freely available software called BioMart which is used for data mining. The caBIG® Website offers quite a comprehensive collection of tools for cancer research (NCI 2009). One of the commonly availed tools in caBIG® is the caArray which is used for microarray (▶ DNA Microarrays) data management system. The caIntegrator is another tool that allows users to set up web portals for integrative search. For integrated genomics, the geWorkbench tools can be used for visualization and analysis of gene expression and sequence data. One can also find tools in caBIG® for molecular analysis, cancer genome-wide association scan (▶ Genome-wide Association Study), clinical trials management, and many more. Currently, more than 40 tools are available in caBIG®.

Other software packages and tools for specialized analysis are also available online. Vital to cancer systems biology research are computational platforms such as software for simulation, analysis, and methodologies for modeling. For visualization, mining, analysis, and modeling of biological networks, an integrated software platform called VisANT (Hu et al. 2009) can be accessed at http://visant.bu.edu/. Additional resources on computational platforms

are also accessible at the Website of the Systems Biology Institute (SBI, http://www.sbi.jp/index.htm) based in Japan. SBI is a nonprofit private research institution that promotes systems biology research and its application to medicine and global sustainability. They have played a major part in the development of the Systems Biology Markup Language (SBML, Hucka et al. 2003) which is used in software packages for graphical editing and simulation of molecular networks.

One software which uses SBML is the CellDesigner (http://celldesigner.org/), a structured diagram editor for drawing gene regulatory and biochemical networks (▶ Gene Regulatory Networks).

Cross-References

- ▶ Cancer
- ▶ Cancer Systems Biology
- ▶ Cytogenetics
- ▶ Disease Databases
- ▶ DNA Microarrays
- ▶ DNA Sequencing
- ▶ Entrez
- ▶ Gene Expression
- ▶ Gene Regulatory Networks
- ▶ Genome-wide Association Study
- ▶ Histopathology
- ▶ KEGG Pathway Database
- ▶ Pathways
- ▶ Pharmacogenomics
- ▶ Proteomics

References

Cesario A, Marcus F (eds) (2011) Cancer systems, biology, bioinformatics and medicine: research and clinical applications. Springer, Dordrecht

Forbes SA, Bhamra G, Bamford S, Dawson E, Kok C, Clements J, Menzies A, Teague JW, Futreal PA, Stratton MR (2008) The catalogue of somatic mutations in cancer (COSMIC). Curr Protoc Hum Genet doi:10.1002/0471142905.hg1011s57

Hu Z, Hung J, Wang Y, Chang Y, Huang C, Huyck M, DeLisi C (2009) VisANT 3.5: multi-scale network visualization, analysis and inference based on the gene ontology. Nucleic Acids Res 37:W115–W121

Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, Kitano H et al (2003) The Systems Biology Markup Language (SBML): a medium for representation and exchange of biochemical network models. Bioinformatics 19:524–531

National Cancer Institute (NCI) (2005) SEER surveillance, epidemiology and end results program. National Institutes of Health, Bethesda

National Cancer Institute (NCI) (2009) The cancer biomedical informatics grid caBIG resource guide. National Institutes of Health, Bethesda

Wang E (ed) (2010) Cancer systems biology, CRC mathematical and computational biology series. CRC Press, Boca Raton

Yan Q (ed) (2010) Systems biology in drug discovery and development: methods and protocols. Humana Press/Springer, New York

Cancer Stem Cell Kinetics

Heiko Enderling
Center of Cancer Systems Biology, St. Elizabeth's Medical Center - CBR 115D, Tufts University School of Medicine, Boston, MA, USA

Synonyms

Cancer-initiating cells; Tumor stem cells; Tumor-initiating cells; Tumor-rescuing units

Definition

Cancer stem cells possess seemingly unlimited proliferation capacity and the ability to give rise to a heterogeneous population of cancer stem cells and mortal non-stem daughters (Al-Hajj et al. 2003). During mitosis, cancer stem cells may undergo either *Asymmetric Division* to simultaneously self-renew and generate one new non-stem cancer cell, or *Symmetric Division* to yield two identical cancer stem cells (Morrison and Kimble 2006; Dingli et al. 2007). Over many generations, a heterogeneous population emerges, within which cancer stem cells are uniquely able to initiate, sustain, and reinitiate tumors after cytotoxic insults and/or at distal sites. Cancer stem cell longevity and self-renewal is, in part, due to upregulation of telomerase that prevents telomere erosion (Blackburn and Gall 1978; Bodnar et al. 1998).

Characteristics

While cancer stem cells can be an infinitesimal subpopulation of the whole tumor (in theory, one cell), the cancer stem cell hypothesis does not exclude cancer stem cells being the dominant phenotype in a tumor or even the trivial case of the cancer stem cell subpopulation being equal to the total tumor population. Let us denote the total tumor population with **T** and the population of cancer stem cells with **C**. Then the cancer stem cell hypothesis can be expressed using set theory as $\mathbf{C} \in \mathbf{T}$. All sets of **C** shown in Fig. 1 are subsets ($\subseteq$) or true subsets ($\subset$) of **T**:

Cancer Stem Cell Hierarchy

It is often argued that tumors follow a unidirectional hierarchy (Fig. 2). Cancer stem cells sit on top of the lineage tree and give rise to mortal non-stem cancer cells with limited proliferation capacity. Let us denote the ith generation non-stem cancer cells with N_i. The kinetics of the cancer stem cell and non-stem cancer cell populations can be described by a term for proliferation (only for cancer stem cells), gain from the previous generation compartment and loss due to advancement into the next generation compartment (only for non-stem cancer cell generations), and a cell death term:

$$\text{cancer stem cells C}: \frac{dC}{dt} = p\delta C - aC$$

1^{st} generation non - stem cancer cells N:

$$\frac{dN_1}{dt} = (1-p)\delta C - \gamma_1 N_1 - b_1 N_1$$

2^{nd} generation non - stem cancer cells N:

$$\frac{dN_2}{dt} = 2\gamma_1 N_1 - \gamma_2 N_2 - b_2 N_2$$

.

.

.

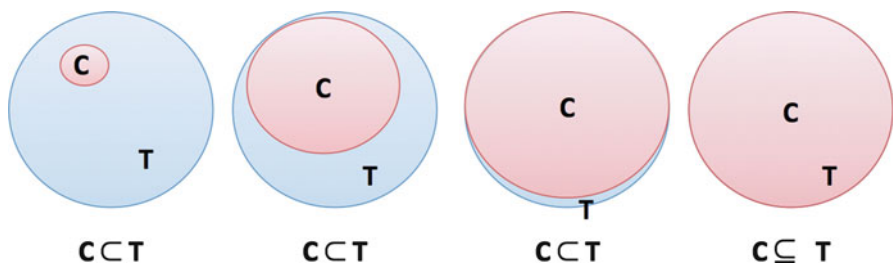
$n-1^{st}$ generation non - stem cancer cells N:

$$\frac{dN_{n-1}}{dt} = 2\gamma_{n-2} N_{n-2} - \gamma_{n-1} N_{n-1} - b_{n-1} N_{n-1}$$

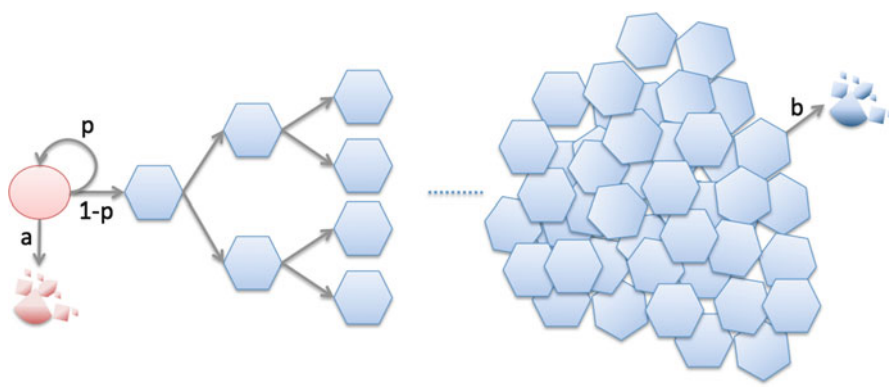
n^{th} generation non - stem cancer cells N:

$$\frac{dN_n}{dt} = 2\gamma_{n-1} N_{n-1} - b_n N_n,$$

where p is the probability of symmetric cancer stem cell division, δ, a and γ_i, b_i are proliferation and apoptosis rates of cancer stem cells and ith generation of



Cancer Stem Cell Kinetics, Fig. 1 Different cancer stem cell (**C**) subsets of a tumor population (**T**)



Cancer Stem Cell Kinetics, Fig. 2 Tumor hierarchy initiated by self-renewing cancer stem cells (*round red cell*, *left*)

non-stem cancer cells, respectively. n is the number of divisions non-stem cancer cells can undergo before cell death (proliferation capacity). The parameters δ and γ_i can be replaced by a function of tumor size f(C + $\sum N_i$) to account for kinetics, including linear, logistic or Gompertzian growth.

Parameter Estimation

It is evident that without cancer stem cells, the population will inevitably die out. Furthermore, tumors can only progress when cancer stem cell proliferation and symmetric division rate is higher than cell death rate, i.e., $p\delta > a$, and any increase in cell death rates a and b_i yields a reduction in tumor size. For $p\delta = a$ and $p\delta > 0$, the tumor is mitotically active at the cellular level but may exhibit macroscopic tumor dormancy. The frequency of cancer stem cells in the population is dependent on the values of probability of symmetric cancer stem cell division p, cancer stem cell death rate a, non-stem cancer cell proliferation and death rates γ_i and b_i, as well as proliferation capacity of non-stem cancer cells n (Enderling et al. 2009; Morton et al. 2011). With adequately chosen parameters, the frequency of cancer stem cells varies enormously (Fig. 1). Experimentation and empirical observations are essential to reliably parameterize the kinetics of all cells in the system.

Cross-References

- ▶ Asymmetric Cell Division
- ▶ Symmetric Cell division

References

Al-Hajj M, Wicha MS, Benito-Hernandez A, Morrison SJ, Clarke MF (2003) Prospective identification of tumorigenic breast cancer cells. Proc Natl Acad Sci USA 100(7):3983–3988

Blackburn EH, Gall JG (1978) A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in Tetrahymena. J Mol Biol 120(1):33–53

Bodnar AG, Ouellette M, Frolkis M, Holt SE, Chiu CP, Morin GB, Harley CB et al (1998) Extension of life-span by introduction of telomerase into normal human cells. Sci (NY) 279(5349):349–352

Dingli D, Traulsen A, Michor F (2007) (A)symmetric stem cell replication and cancer. PLoS Comput Biol 3(3):e53

Enderling H, Hlatky L, Hahnfeldt P (2009) Migration rules: tumours are conglomerates of self-metastases. Br J Cancer 100(12):1917–1925

Morrison SJ, Kimble J (2006) Asymmetric and symmetric stem-cell divisions in development and cancer. Nature 441(7097):1068–1074

Morton CI, Hlatky L, Hahnfeldt P, Enderling H (2011) Non-stem cancer cell kinetics modulate solid tumor progression. Theor Biol Med Model 8(1):48

Cancer Systems Biology

Zhihui Wang and Thomas S. Deisboeck
Harvard-MIT (HST) Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, MA, USA

Definition

Tumors are heterogeneous cellular entities whose growth is dependent on dynamical interactions between cancer cells themselves and the continually varying microenvironment these cells live in (Bissell and Radisky 2001). It is very difficult if not impossible to investigate these dynamic interactions by using wet-lab experiments only because experimental complexity usually restricts observations to single or very limited spatial and/or temporal scales without allowing to reproducibly and quantitatively analyze relationships between these scales. Hence, a more recent paradigm shift within the cancer research community is to recognize and study cancer as a *systems* disease.

Cancer systems biology is an emerging multidisciplinary field that focuses on the systemic understanding of cancer initiation and progression by investigating how individual components interact and collaborate to give rise to the function and behavior of the tumor system as a whole (Deisboeck et al. 2001). In addition to conventional biomedical experiments, a systems approach often involves computational modeling to help decipher the massive amount of data generated today especially in molecular and cell biology. In silico cancer models are necessarily simplified, yet are beginning to adequately represent specific cancer phenomena. In fact, it has been increasingly recognized that such a theoretical approach may help to simulate, predict, and optimize procedures, experiments, and therapies, to test and refine hypotheses. The development of a successful in silico cancer model is expected to be iteratively

conducted, with available experimental data used to guide, support, and shape the model design, and to verify and validate model results.

The discrepancy between the in vitro and the clinical settings has significantly affected the ability of researchers to design and develop successful cancer therapeutics (Khalil and Hill 2005). Hence, one of the most important challenges facing cancer researchers currently is how to better translate in vitro discoveries to clinical application. Data-driven, predictive cancer systems models can help bridge this gap and expedite the development of effective cancer therapeutics. Although still at an early stage, cancer systems biology has begun to provide useful insights into the disease mechanisms involved by providing a systematic and integrative framework for incorporating data, generating testable hypotheses, and guiding further experiments.

Cross-References

► Multilevel Modeling, Cell Proliferation

References

Bissell MJ, Radisky D (2001) Putting tumours in context. Nat Rev Cancer 1:46–54

Deisboeck TS, Berens ME, Kansal AR, Torquato S, Stemmer-Rachamimov AO, Chiocca EA (2001) Pattern of self-organization in tumour systems: complex growth dynamics in a novel brain tumour spheroid model. Cell Prolif 34:115–134

Khalil IG, Hill C (2005) Systems biology for cancer. Curr Opin Oncol 17:44–48

Cancer Theories

Ana M. Soto and Carlos Sonnenschein
Department of Anatomy and Cellular Biology, Tufts University School of Medicine, Boston, MA, USA

Definition

Scientific theories serve the purpose of making the world intelligible. They provide organizing principles and construct objectivity by framing observations and experiments.

Characteristics

John Cairns (1997) rightfully stated: "Biology and cancer research have developed together. Invariably, at each stage, the characteristics of the cancer cell have been ascribed to some defect in whatever branch of biology happens at the time to be fashionable and exciting. . ." This statement pronounced 15 years ago applies to the status of cancer research today, and thus justifies a brief historical note to introduce the two currently competing theories of carcinogenesis, namely, the somatic mutation theory and the tissue organization theory of carcinogenesis.

Historical Perspective

The modern view of ► neoplasms can be traced back to the second half of the nineteenth century, following the advent of the cell theory and the development of modern pathology spearheaded by Virchow, Thiersch and Waldeyer, in Germany. In the view of these German pathologists, the bases of organismic order resulted from embryonic development and were due to interactions of multiple mutually dependent systems within the organism. Neoplasms were viewed as a disorder of these normally interdependent parts (Moss 2003; Sonnenschein and Soto 2008).

All along the twentieth century, the full range of phenotypic potentials available to cells and tissues interacting within the organism was gradually replaced by a more reductionist conceptualization of vital processes (Moss 2003). This shift from the organismic view to a new cell-based and eventually gene-based view started in 1914 when Theodor Boveri claimed in his book entitled *The Origin of Malignant Tumors* that "the problem of tumors is a cell problem" and that cancer was due to "a certain permanent change in the chromatin complex" which, "without necessitating an external stimulus, forces the cell, as soon as it is mature, to divide again" (Boveri 1929). Ever since, cancer was considered as a problem of ► control of cell proliferation due to permanent changes in the "chromatin", which in Boveri's time was already known to contain the heritable material. Boveri's theory represents the precursor of what became to be known later in the twentieth century as the ► somatic mutation theory of carcinogenesis (SMT).

The Somatic Mutation Theory

The premises adopted by the SMT are: (1) cancer is derived from a single somatic cell that, over time, accumulate multiple DNA mutations, (2) the default state of cells in metazoa is *quiescence*, and (3) cancer is a disease of the ▶ control of cell proliferation caused by mutations in genes that affect steps within the cell cycle (Hahn and Weinberg 2003). During the last decades of the twentieth century until now, the SMT became the dominant view in carcinogenesis. From this perspective, cancer was perceived as being a disease caused by mutations in the DNA of a single founder cell and, therefore, cancer has been considered as a clonal disease.

In spite of the dominance of the SMT, until the advent of the ▶ oncogene hypothesis in the 1970s, there were several competing theories to explain cancer; they differed at the level of biological organization in which carcinogenesis occurred. Thus, discussions centered on whether cancer was either a problem of ▶ control of cell proliferation and/or of cell differentiation, or else, of tissue organization. From the 1950s to the 1970s, D.W. Smithers, C. Dawe, J.W. Orr, B. Mintz, G.B. Pierce, and several others offered compelling evidence for carcinogenesis to be considered as a tissue-based phenomenon akin to development gone awry (Soto and Sonnenschein 2012). Notwithstanding, around the 1960s and 1970s, the reductionist view became dominant based on in vitro transformation assays which replaced the animal models that were used until then. Philosophers, historians, and sociologists of biology have advanced explanations about how this view achieved dominance. One of them, Joan Fujimura, proposed a sociological explanation that connected the interests of molecular biologists and the rise of the genetic engineering biotechnology industry into constructing "doable problems" such as "Are there molecular changes in the cellular proto-oncogenes of tumor cells?" Michel Morange proposed, instead, an epistemic explanation whereby the oncogene hypothesis became attractive to researchers because it suggested a straight-forward research program. In this context, the products of oncogenes participated in signal transduction conveying messages from the extracellular milieu through membrane receptors to the nucleus and these messages regulated ▶ cell proliferation (Soto and Sonnenschein 2012).

The Tissue Organization Field Theory

An alternative theory of carcinogenesis, namely, the tissue organization field theory (TOFT) is based on principles similar to those adopted by cancer researchers at the end of the nineteenth century. The TOFT directly challenges the core premises of the SMT by positing that (1) carcinogenesis, like histogenesis (the formation of tissues) and organogenesis (the formation of organs), is a supracellular, emergent phenomenon occurring at the tissue level of biological organization. From this perspective, regardless of its many causes, cancer becomes "development gone awry." An additional fundamental premise of the TOFTs is that (2) as in unicellular organisms and metaphyta, the default states of metazoan cells, are *proliferation* and *motility*. By "the default state of cells is *proliferation*" is meant that the state that cells assume in the presence of abundant building blocks (nutrients) needed to synthesize new cells is *proliferation*. The TOFT stems from a body of literature showing that cells that once belonged to a neoplasm could be "normalized" upon recombination with normal tissue. The SMT precludes this eventuality (Sonnenschein and Soto 2008).

The Hybrid Theories of Carcinogenesis

During over 40 years of dominance by the original and the updated SMT (the oncogene hypothesis), hundreds of oncogenes and dozens of suppressor genes have been described (Hanahan and Weinberg 2011). Originally, Hanahan and Weinberg, proposed what they called the Hallmarks of Cancer, i.e., "...the basic rules governing the neoplastic transformation of normal human cells". These originally six rules postulated that the properties of cancer cells were "...to generate their own mitogenic signals, to resist exogenous growth-inhibitory signals, to evade apoptosis, to proliferate without limits (i.e., to undergo immortalization),to acquire vasculature (i.e., to undergo angiogenesis), and in more advanced cancers, to invade and metastasize." Later on, two extra rules were added, namely: reprogramming of energy metabolism and evading immune destruction (2011). Central to their original and to their updated view remains the idea that carcinogenesis is a *cell-based* problem due to mutations that cause *the founder cancer cell and its progeny* to "proliferate without limits".

In his analysis of the metaphysical presuppositions and scientific practices in cancer research, the philosopher James Marcum (2005) observed an ideological

and practical shift among the supporters of the SMT who adopted concepts of the TOFT, shared by other researchers (Xu et al. 2009). For example, for those who are committed to explaining carcinogenesis through a sub-cellular (mutational) strategy, the causal role of the ▶ stroma in carcinogenesis is transformed into the problem of how mutated genes can affect the interactions of the cancer cells harboring these genes with otherwise normal neighboring cells. This "hybrid" theory has been criticized on the grounds that the premises of the SMT and the TOFT contradict each other (Soto and Sonnenschein 2012).

Systems Biology, Heuristics and Theory Testing

Systems Biology offers an opportunity to explore the ▶ heuristic value and usefulness of the two main theories of carcinogenesis. Mathematical modeling and computer simulation enables researchers the boldness to choose premises and temporarily reject data sets without having to commit prematurely to a program of expensive and time-consuming 'wet' experiments. This exploratory role of Systems Biology may become central to breaking the habit of fixing lacks of fit with ad hoc explanations instead of taking a bold and critical look at the premises that were adopted (Enderling and Hahnfeldt 2011; Soto and Sonnenschein 2012).

In an ultimate analysis, the passage of time will be the final arbiter in deciding which of the two main theories of carcinogenesis has been most useful in organizing principles and constructing objectivity by framing observations and experiments.

Cross-References

- ▶ Cell Proliferation
- ▶ Heuristic Optimization
- ▶ Neoplasms
- ▶ Oncogene
- ▶ Somatic Mutations
- ▶ Stroma

References

Boveri T (1929) The origin of malignant tumors. Williams & Wilkins, Baltimore

Cairns J (1997) Matters of life and death. Princeton University Press, Princeton

Enderling H, Hahnfeldt P (2011) Cancer stem cells in solid tumors: is 'evading apoptosis' a hallmark of cancer? Prog Biophys Mol Biol 106:391–399

Hahn WC, Weinberg RA (2003) Rules for making human tumor cells. N Engl J Med 347:1593–1603

Hanahan D, Weinberg RA (2011) Hallmarks of cancer: the next generation. Cell 144:646–674

Marcum JA (2005) Metaphysical presuppositions and scientific practices: reductionism and organicism in cancer research. Int Stud Philos Sci 19:31–45

Moss L (2003) What genes can't do. MIT Press, Cambridge, MA

Sonnenschein C, Soto AM (2008) Semin Cancer Biol 18:372–377

Soto AM, Sonnenschein C (2012) Is systems biology a promising approach to resolve controversies in cancer research? Cancer Cell Int 12:12

Xu R, Boudreau A, Bissell MJ (2009) Tissue architecture and function: dynamic reciprocity via extra- and intra-cellular matrices. Cancer Metastasis Rev 28:167–176

Cancer-initiating Cells

▶ Cancer Stem Cell Kinetics

Canonical Model

Eberhard O. Voit
The Wallace H. Coulter Department of Biomedical Engineering, Georgia Institute of Technology and Emory University, Atlanta, GA, USA

Definition

A canonical model is a mathematical representation of a biological phenomenon that is constructed according to strict rules and therefore has a regularly structured format. In contrast to ad hoc models, which are composed from a variety of known or alleged functions that seem to be best suited for the specific modeling purpose, each canonical model is obtained from a single type of approximation for all processes within the modeled system. Linear models are canonical, because all processes are represented with linear functions. Nonlinear canonical models include S-systems, Generalized Mass Action systems, Lotka-Volterra systems, and lin-log models. Canonical models are particularly useful if specific mechanisms or process descriptions are not known.

Canonical Network Motifs

Arun S. Konagurthu[1] and Arthur M. Lesk[2]
[1]Clayton School of Information Technology, Monash University, Clayton, VIC, Australia
[2]Department of Biochemistry and Molecular Biology, and Huck Institutes of Genomics, Proteomics, and Bioinformatics, Pennsylvania State University, University Park, PA, USA

Synonyms

3-Cycle; Bifan; FFL; MIM; Network building blocks; SIM; Subgraph patterns

Definition

A system-wide view of complex interactions involved in biological processes finds a natural description as networks or directed graphs. Systematic enumeration of small subgraphs in various biological networks has identified preponderant patterns, which are often described as *motifs*. Many authors accept the following as a canonical set of network motifs: single-input motif (SIM), multiple-input motif (MIM), and feed-forward loop (FFL). There is consensus that motifs are "building blocks" of various biological networks, naturally selected to contribute to their functional architecture. In some networks, especially in those describing genome-wide transcription regulation, specific functional roles have been suggested for motifs (Shen-Orr et al. 2002; Zaslaver et al. 2004; Mangan et al. 2006). For example, SIMs are commonly associated with temporal ordering of gene expression, MIMs with combinatorial gene regulation, and FFLs with filters that do not pass on transient signals. These functions depend not only on the topology of the subgraph, but on the logic at nodes receiving multiple inputs.

Characteristics

There follows a formal description of commonly enumerated subgraph patterns in various biological networks. The description of these patterns and their enumeration algorithms are described in detail in Konagurthu and Lesk (2008a). A study analyzing the preponderance of various subgraph patterns in large natural networks can be found in Konagurthu and Lesk (2008b).

Common Subgraph Patterns in Biological Networks

These patterns represent some of the commonly enumerated subgraph patterns in biological networks: feed-forward loop (FFL), 3-Cycle (3-CYC), single-input motif (SIM), multiple-input motif (MIM), and bifan. (See Fig. 1.)

The following statements formally define these subgraph patterns. The definitions are, in a sense, "orthogonal." That is, our definitions are chosen to avoid ambiguity in classification of any subgraph of a network. (Usage in the literature often allows, for example, MIMs to contain SIMs as subgraphs, leading to difficulty in formulating the problem of motif enumeration.)

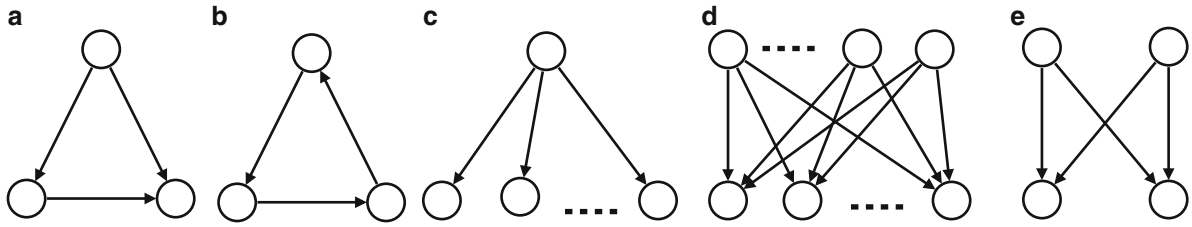
Feed-forward loop: (Fig. 1a) A FFL is a set of three vertices (source, intermediate, and target) with one direct path, and another indirect path through an intermediate vertex, from source to target.

3-Cycle: (Fig. 1b) A 3-CYC is a three-vertex directed cyclic graph.

Single-input motif: (Fig. 1c) A SIM in a large directed graph $G(V, E)$, with a vertex and directed edge (arc) set V and E respectively, is a *bipartite* subgraph $G'(\mathcal{P} \cup \mathcal{C}, E')$ containing two disjoint sets (layers) of vertices $\mathcal{P}$ and $\mathcal{C}$ (denoting parent and child vertex sets, respectively), subject to the following constraints:

1. $|\mathcal{P}|$ (= the number of elements of $\mathcal{P}$) and $|\mathcal{C}| \geq 2$.
2. There is an outgoing edge from the vertex in $\mathcal{P}$ to *every* vertex in $\mathcal{C}$. (E' is the edge set containing these arcs.)
3. There are no arcs between any two vertices in $\mathcal{C}$.
4. There are no incoming edges to any vertex in $\mathcal{C}$ from any other vertex (outside the SIM) in the set of vertices $V - \mathcal{P} - \mathcal{C} \in G$. (Note that this constraint does not prohibit outgoing edges from any vertex in $\mathcal{C}$ to a vertex outside of the SIM.)
5. For the singleton vertex in $\mathcal{P}$, $\mathcal{C}$ is a *maximal* set under the above constraints. (That is, $\mathcal{C}$ cannot be extended by conforming to the above constraints.)

Multiple-input motif: (Fig. 1d) A MIM in $G = (V, E)$ is a bipartite subgraph $G' = (\mathcal{P} \cup \mathcal{C}, E')$ containing two



Canonical Network Motifs, Fig. 1 (**a**) Feed-forward loop; (**b**) 3-Cycle; (**c**) single-input motif; (**d**) multiple-input motif; (**e**) Bifan

disjoint sets (layers) of vertices $\mathcal{P}$ and $\mathcal{C}$ subject to the following constraints:

1. $|\mathcal{P}| \geq 2$ and $|\mathcal{C}| \geq 2$.
2. There are outgoing edges from *every* vertex $p_i \in \mathcal{P}$ to *every* vertex $c_j \in \mathcal{C}$. (E' is the edge set containing these arcs.)
3. There are no arcs between any two vertices in $\mathcal{P}$.
4. There are no arcs between any two vertices in $\mathcal{C}$.
5. Both $\mathcal{P}$ and $\mathcal{C}$ are maximal sets under the above constraints.

Bifans: (Fig. 1e) We note that MIMs arose as a generalization of bifans in the literature. To enforce orthogonality of motif definition, consider a bifan only as a *maximal* 2×2 MIM, where $|\mathcal{P}| = |\mathcal{C}| = 2$. (See section below for more details.)

Maximality and Orthogonality of Motif Definitions

We emphasize that the criterion of *maximality* is crucial to maintain independence (orthogonality) in enumerating the above motifs. In case of SIM, the set $\mathcal{C}$ is maximal, whereas with MIMs, both $\mathcal{P}$ and $\mathcal{C}$ sets are maximal. Without this constraint, it is easy to overcount non-maximal subgraphs of the type MIMs and SIMs. For example, a maximal MIM with p parents and c children contains $[2^p - (p+1)] \times [2^c - (c+1)] - 1$ easily enumerable non-maximal "sub-MIMs," including $\binom{p}{2} \times \binom{c}{2}$ non-maximal bifans. Similarly, a maximal SIM with one parent and c children contains $2^c - c - 2$ non-maximal "sub-SIMs."

The definitions of the SIM exclude the counting of subgraphs of MIMs as SIMs. Without the in-degree constraint (4) on a SIM, every MIM containing p parents and c children would contain $p \times (2^c - (c+1))$ subgraphs that form either a SIM or a subset of a SIM with two or more children (in set $\mathcal{C}$).

Enumeration

Enumerating Feed-forward loops: FFLs can be enumerated in a number of ways. A simple procedure to enumerate all FFLs in a directed graph $G(V, E)$ is as follows: At each vertex $v_i \in V$, for each pair of outgoing edges from v_i to distinct vertices v_j and v_k in G, count a FFL if there is an arc from either v_j to v_k or vice versa.

Enumerating 3-Cycles: Enumerating 3-CYCs is similar to the procedure above. At every vertex v_i, for each pair of outgoing and incoming arcs to-v_j and from-v_k respectively, count a 3-CYC if there is an arc from v_j to v_k. Note that, since a 3-Cycle has an automorphism number of 3 – that is, there are three different permutations of vertices giving an indistinguishable connectivity of 3-CYC – this procedure results in overcounting 3-CYCs by three times. This can be avoided if the 3-CYC list is generated under some canonical ordering of vertex labels.

Enumerating Single-input motifs: For each vertex v_i in the full graph $G = (V, E)$ find the set $C(v_i) \equiv C_i$ of all its children – the set of vertices that are targets of an arc from v_i. Shrink the set C_i by eliminating vertices which have incoming edges from vertices outside $v_i \cup C_i$. The reduced C_i, provided it contains at least two vertices, forms the candidate set from which maximal independent sets of vertices – that is, set of vertices without links between them – are extracted.

Use Bron and Kerbosch algorithm (Bron and Kerbosch 1973) for finding all cliques in graphs as follows. (A clique is a maximal complete subgraph, a subgraph with an edge between every two vertices.) Consider the *undirected complement* of the graph induced by reduced C_i. This is created by deleting all edges originally present, and introducing an undirected edge between any two vertices originally unconnected.

Then the condition that a subgraph of the reduced C_i set has no edges becomes that the corresponding subgraph of the inverted graph is completely connected. Each possible clique (of size ≥ 2 vertices) in the complement graph gives us the disjoint set $\mathcal{C}$ of a SIM for every vertex $v_i \equiv \mathcal{P}$, satisfying all the constraints in its definition.

Enumerating Multiple-input motifs: For each vertex $v_i \in V$ find the set C_i of all its children. Let $P(c)$ be the set of vertices ("parents" of vertex c) which have an incoming edge to c. For each pair $c_1, c_2 \in C_i$, find the maximal subset of nodes that are parents of c_1 and c_2, that is, the intersection of the parents of c_1 and the parents of c_2: $P(c_1) \cap P(c_2)$.

By considering all pairs of elements from the child sets derived from every node, we build up a list of child and corresponding parent sets such that all parents have arcs to all children. Initially, c_1, c_2 and $P(c_1) \cap P(c_2)$ form a child and corresponding parent set (provided that $P(c_1) \cap P(c_2)$ contains ≥ 2 nodes). If another pair c_3, c_4 gives us the same intersection of parents – $P(c_1) \cap P(c_2) = P(c_3) \cap P(c_4)$ – we merge the child sets, updating our list to contain: c_1, c_2, c_3, c_4 and $P(c_1) \cap P(c_2)$ $(= P(c_3) \cap P(c_4))$. We build up maximal child and corresponding parent sets. The process is repeated for all pairs in the parent sets that share the same children.

This step has built up candidate MIMs that are complete as far as parent-child arcs are concerned. The next step is finding subsets of candidate MIMs free of parent-parent and child-child edges.

In the induced *undirected* version of the graph containing parent and child sets of a candidate MIM, consider the parent set and "invert" the edges between vertices in this set by deleting all originally present edges, and introducing an edge between any two nodes originally unconnected. Repeat the same procedure for the vertices in the child set. Preserve the edges (ignoring the direction) linking vertices in the parent set to nodes in the child set – we know by construction that there is an edge between every parent and every child node.

Then the problem of finding MIMs – maximal subsets of parent and child graphs, that contain no interparent edges and no interchild edges but all parent-child edges – is equivalent to finding all cliques in this transformed graph of each candidate MIM.

Note on densely overlapping regulons: Shen-Orr et al. (2002) introduced densely overlapping regulon (DOR). DORs can be thought of as an incomplete MIM, from which some of the arcs between $\mathcal{P}$ and $\mathcal{C}$ are absent.

The method for enumerating MIMs described above can be extended to count DORs. The sets $\mathcal{P}$ and $\mathcal{C}$ are first calculated such that there is an arc between every node in $\mathcal{P}$ and every node in $\mathcal{C}$. Subsequently, relaxing the constraint of having "complete" sets of edges between the two sets, $\mathcal{P}$ and $\mathcal{C}$ can be iteratively extended such that they now include nodes that share at least a certain user-defined threshold on number of edges between the sets. Once extended, the edges between $\mathcal{P}$ and $\mathcal{C}$ can be assumed to be complete, and DORs (free of parent-parent and child-child edges) extracted using the procedure discussed earlier.

References

Bron C, Kerbosch J (1973) Finding all cliques of an undirected graph. Commun ACM 16:575–577

Konagurthu A, Lesk A (2008a) Single and multiple input modules in regulatory networks. Proteins: Struct Funct Bioinf 73:320–324

Konagurthu A, Lesk A (2008b) On the origin of distribution patterns of motifs in biological networks. BMC Syst Biol 2:73+

Mangan S, Itzkovitz S, Zaslaver A, Alon U (2006) The incoherent feed-forward loop accelerates the response-time of the gal system of *Escherichia coli*. J Mol Biol 356:1073–1081

Shen-Orr SS, Milo R, Mangan S, Alon U (2002) Network motifs in the transcriptional regulation network of *Escherichia coli*. Nat Genet 31:64–68

Zaslaver A, Mayo AE, Rosenberg R, Bashkin P, Sberro H, Tsalyuk M, Surette MG, Alon U (2004) Just-in-time transcription program in metabolic pathways. Nat Genet 36:486–491

Canonical Nonlinear Modeling

▶ S-System

Canonical Pathway of microRNA Biogenesis

▶ MicroRNA Biogenesis, Regulation

Capacity

► Disposition

Capacity to Evolve

► Evolvability, Generalized Biology

Capillary Electrophoresis Mass Spectrometry–based Metabolomics

► CE-MS-based Metabolomics

Carbon Assimilation

► Photosynthesis

Carboxyfluorescein Succinimidyl Ester (CFSE)

► Lymphocyte Labeling, Cell Division Investigation

Carcinogens

► Xenobiotics

Carcinoma In Situ

Barbara J. Davis
Section of Pathology, Tufts Cummings School of Veterinary Medicine Biomedical Sciences, North Grafton, MA, USA

Definition

A specific designation for neoplasia derived from epithelial dysplastic changes that occur in all the layers of epithelium but do not breach the basement membrane. They may or may not progress to malignancy.

Cross-References

► Cancer Pathology

Case-based Reasoning

Eyke Hüllermeier, Thomas Fober and Marco Mernberger
Philipps-Universität, Marburg, Germany

Definition

Case-based reasoning (CBR) is a computational problem solving methodology offering a principled approach to knowledge engineering and knowledge-based systems design. Being rooted in cognitive psychology, CBR essentially seeks to formalize a problem solving paradigm upon the basis of a simple rule of thumb, namely, that "similar problems tend to have similar solutions" (Kolodner 1993). More specifically, the idea of CBR is to store and exploit the experience from similar problems in the past, and to adapt then successful solutions to the current situation. Thus, the core of every case-based problem solver is the *case base* or *case library*, which is a collection of memorized "chunks of experience" called cases. Moreover, the concept of *similarity* plays a key role in CBR systems. Inference in CBR can be considered as a specific form of *analogical reasoning*.

Characteristics

CBR is a subfield of artificial intelligence (AI) (see ► Evolutionary Algorithms) with close connections to machine learning, information retrieval (see ► Identification of Gene Regulatory Networks, Machine Learning), databases, semantic web, and ► knowledge management. It combines methods from these and related areas to tackle specific problem tasks, such as diagnosis, planning, product recommendation, and

experience management (Bergmann et al. 2003). From a knowledge-based systems perspective, CBR can be seen as an alternative to the rule-based paradigm for expert systems design. In this regard, it offers a number of advantages, especially from a knowledge acquisition and maintenance point of view. In fact, knowledge in the form of individual cases (often specified as problem/solution tuples (x, y), indicating that y is the best or at least a sufficiently good solution to the problem x) is in general more readily available than generally valid rules, which are difficult to elicit from human experts. Moreover, whereas rule-based inference normally requires a knowledge base to be complete and correct in a logical sense, analogical and similarity-based inference are more tolerant toward the incompleteness and inconsistency of knowledge (Hüllermeier 2007).

Applying CBR principles in a machine learning context, for example, for ▶ classification, comes down to realizing "lazy," instance-based instead of "eager," model-based learning (like rule induction). The nearest neighbor method can be seen as the simplest form of case-based reasoning and is a typical example of the so-called structural approach to CBR. In structural CBR, cases are represented according to a structured vocabulary, that is, an ▶ ontology. The describing features of a case can be organized as flat attribute-value tables, in an object-oriented manner, as graph structures, or by sets of atomic formulas of a logic language. Although the structural approach is most widely used in practice and arguably most relevant for the biological sciences, it is worth mentioning the existence of other types of CBR, such as textual and conversational CBR, which can differ significantly with regard to case representation and reasoning.

By reasoning on the basis of specific cases instead of first principles, CBR is of particular relevance for scientific fields that are characterized by a lack of such principles, and which are more data- than theory-driven. From this point of view, CBR is especially interesting for the biological and life sciences, including systems biology, in which data abounds, analogical/similarity-based reasoning is omnipresent, and theory formation is largely founded on individual case studies. CBR systems have already been developed for a number of problems in this field, such as gene finding, knowledge discovery in sequence databases, protein-structure determination and the planning of crystallization (Jurisica and Glasgow 2004), or more recently the analysis of microarray data (Bangpeng and Shao 2010).

The CBR Cycle

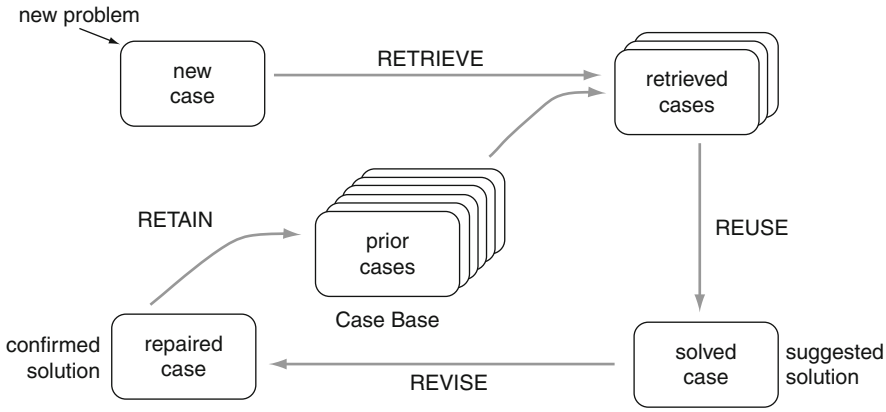
Despite the diversity of concrete implementations, the essentials of the CBR methodology are captured by a surprisingly simple and uniform process model, referred to as the *CBR cycle* (Aamodt and Plaza 1994). It consists of four sequential steps organized around the knowledge of the CBR system: retrieve, reuse, revise, and retain (see Fig. 1).

Problem solving starts when a new problem (the query) must be solved. First, in the *retrieve phase*, one or several cases with the highest similarity to the query are selected from the case base, where similarity is determined by an underlying similarity measure. Since the efficiency of the retrieval step critically depends on the size of the case base, a branch of CBR research deals with methods that improve retrieval efficiency through the use of specific index structures.

In the subsequent *reuse phase*, the solutions of the retrieved cases are adapted according to the requirements of the query. In simple tasks like ▶ classification, where the number of solutions (classes) is limited and typically much smaller than the number of cases, no adaptation is needed at all. On the other hand, for synthetic tasks such as configuration of technical systems or planning, where the solution space is potentially infinite and exceeding the number of available cases, solution adaptation becomes essential. Several techniques for adaptation in CBR, including *transformational* and *generative* adaptation, have been proposed so far (Lopez Mantaras et al. 2005).

In the *revise phase*, the solution determined so far is verified and possibly corrected or improved, for example, through intervention by a domain expert. Feedback can be given in the form of a rating of the result or a manually corrected revised case.

Finally, the *retain phase* adopts the feedback from the revise phase and updates the knowledge. In particular, by adding the revised case to the case base and thus making the new problem solving experience available for future problem solving episodes, a simple form of learning can be realized. However, the continuous growth of the case base causes a utility problem as it may compromise retrieval efficiency. Explicit competence models have therefore



Case-based Reasoning, Fig. 1 The CBR cycle

been developed to enable the selective retention of cases (Smyth and McKenna 2001).

Representing and Structuring Knowledge in CBR

A unified view of the knowledge contained in a (structural) CBR application employs the metaphor of a *knowledge container*, typically referring to four such containers: the vocabulary, the case base, the similarity measure, and the adaptation knowledge.

The vocabulary (often called ontology today) is the basis of all knowledge and experience representation in CBR. The vocabulary defines the information entities and structures (e.g., classes, relations, attributes, data types) that can be used to represent cases, similarity measures, and adaptation knowledge. The case base is the primary form of knowledge in CBR. Traditionally, a case is considered as an instance in the representation space defined by the vocabulary, for example, a vector in an attribute-value representation. From a knowledge representation point of view, vocabulary and case base representations are standard applications of AI and database technology.

Since cases are selected based on their similarity to the current problem, the similarity measure used in a CBR system encodes important knowledge of the domain. It is usually formalized in terms of a function that maps a pair of problem descriptions to a real number, often restricted to the unit interval, with the convention that a high value indicates a high similarity. The semantics of such similarity degrees can be defined via the notions of preference and utility: A higher degree of similarity suggests that a case is more useful for solving the current problem. As a means for modeling similarity functions, the so-called *local–global principle* is widely used: The similarity function is decomposed according to the vocabulary, in such a way that local similarity functions pertaining to individual attributes are used to model the preference according to the corresponding attributes. These local similarities are then aggregated into the global similarity by means of an appropriate aggregation function.

Cross-References

- ▶ Classification
- ▶ Evolutionary Algorithms
- ▶ Identification of Gene Regulatory Networks, Machine Learning
- ▶ Knowledge Management
- ▶ Model Testing, Machine Learning
- ▶ Ontology

References

Aamodt A, Plaza E (1994) Case-based reasoning: foundational issues, methodological variations, and system approaches. AI Communications 7(1):39–59

Bangpeng Y, Shao L (2010) ANMM4CBR: a case-based reasoning method for gene expression data classification. Algorithm Mol Biol 5(1):1–11

Bergmann R, Althoff KD, Breen S, Göker M, Manago M, Traphöner R, Wess S (2003) Developing industrial case-based reasoning applications: the INRECA methodology. Springer, Berlin

Hüllermeier E (2007) Case-based approximate reasoning. Springer, Berlin
Jurisica I, Glasgow J (2004) Applications of case-based reasoning in molecular biology. Artif Intell Mag 25(1):85–97
Kolodner JL (1993) Case-based reasoning. Morgan Kaufmann, San Mateo
Lopez Mantaras R, McSherry D, Bridge D, Leake D, Smyth B, Craw S, Faltings B, Lou Maher M, Cox MT, Forbus K, Keane M, Aamodt A, Watson I (2005) Retrieval, reuse, revision and retention in case-based reasoning. Knowl Eng Rev 20:215–240
Smyth B, McKenna E (2001) Competence models and the maintenance problem. Comput Intell 17(2):235–249

Category Homomorphism

▶ Functor

Category Map

▶ Functor

Causal Bayesian Networks

Daniel Polani
Adaptive Systems Research Group, School of Computer Science, University of Hertfordshire, Hatfield, UK

Synonyms

Causal probabilistic graphical models

Definition

Causal Bayesian networks are an extension of ▶ probabilistic graphical models by the concept of (*causal*) *intervention* which states how the joint probability changes if variables in the network are externally coerced into a given value. This can be seen as modeling the influence of experimental modification of a system as opposed to observing an unperturbed system.

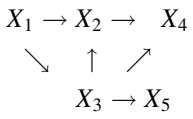
Formally, if one intervenes in a given causal Bayesian network, say by setting a variable X_l to a fixed value x_l, then the joint distribution is changed to

$$p(x_1, \ldots, x_{l-1}, x_{l+1}, \ldots, x_n | \hat{x}_l) := \prod_{\substack{i=1 \\ i \neq l}}^{n} p(x_i | \mathrm{Pa}[x_i])$$

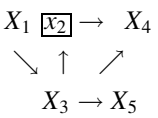
where $\hat{x}_l$ denotes the intervention in the given variable and thus random variables X_i with X_l as parent use the fixed value x_l instead of the parent variable X_l. The definition extends to sets of intervention variables.

Characteristics

Consider, as example, the Causal Bayesian Network (see also ▶ Probabilistic Graphical Model)



Intervening in X_2 by setting it to the value x_2, essentially changes the network to:



whose probability is to be read as:

$$p(x_1, x_3, x_4, x_5 | \hat{x}_2) := p(x_1) p(x_3 | x_1) p(x_4 | x_2, x_3) p(x_5 | x_3)$$

Note that this intervention in X_2 is the same as modifying the original Causal Bayesian network to remove the incoming connections to X_2, to fix the variable's value to x_2, and considering the resulting joint distribution of the other variables.

References

Pearl J (2000) Causality: models, reasoning and inference. Cambridge University Press, Cambridge, UK

Causal Probabilistic Graphical Models

▶ Causal Bayesian Networks

Causal Relationship

Katsuhisa Horimoto
Computational Biology Research Center, National Institute of Advanced Industrial Science and Technology, Koto-ku, Tokyo, Japan

Synonyms

Bayes rule; Bayesian network model; Conditional distribution; Conditional Independence; Correlation relationship

Definition

A causal relationship is when one variable causes a change in another variable. In other words, when an event (the cause) occurs, the second event (the effect) is understood as a consequence of the first. Note that a causal relationship cannot be defined from a joint distribution of observed variables alone, while any correlation (associational) relationship can be defined in terms of the distribution (Pearl 2009).

Cross-References

▶ Bayesian Network Model
▶ Conditional Independence
▶ Correlation Relationship

References

Pearl J (2009) Causal inference in statistics: an overview. Statistics Surveys 3:96–146

Causality

Max Kistler
IHPST, Université Paris 1 Panthéon-Sorbonne, Paris, France

Synonyms

Causation

Definition

Causality is a philosophical concept that expresses a relation or a process linking two (or more) distinct events or facts, which consists in one bringing about the other. Ordinary language contains many expressions denoting causation: x makes y happen, x provokes y, x produces y, x brings y about, x gives rise to y, x induces y, etc. Many verbs express particular types of causation, both in common sense and science: break, bend, link, release, trigger, modify, influence, produce, transfer, catalyze, etc. Many philosophical analyses of the concept of causation have been developed but none has won universal acceptance.

Characteristics

Causality (or causation) has always been a focus of philosophical controversy. One debate, which is today as lively as ever, is about the correct analysis of the concept: What is a causal relation, and what are its terms: objects, events, or facts? However, since the Renaissance, this debate coexists with a more fundamental debate on whether causality is a legitimate notion in science at all. I start with the latter issue, and then present the most influential and relevant accounts of causation.

Contemporary debates about the legitimacy of the concept of causation in science have been structured by Russell's 1912 essay "On the notion of cause." Here are two of his reasons for doubting that science looks for or analyzes causal relations or processes.

1. The "principle of causality" says that "the same causes have the same effects." In other words,

there are strict lawful regularities where instances of a first type of event, A, are invariably followed by instances of a second type of event, B. There are two reasons for doubting that such laws exist in advanced sciences. The first is that events recur only insofar as they are conceived vaguely: Once an event is described in a quantitatively precise manner, as required by advanced science, the probability of its recurrence tends toward zero. However, this claim is more plausible for macroscopic events such as throwing bricks than for microscopic events, such as the oxidation of a carbon atom yielding a carbon dioxide molecule. The second reason holds for both macro and micro events: A type of event can only recur if it is narrowly construed: Abstracting away from the circumstances, which will never exactly recur, the event must be circumscribed to a narrow space-time region. The problem is that the narrow localization of the first event A prevents the regularity of its being followed by a second event B from being strict: The circumstances being unspecified, nothing precludes them from containing factors that interfere with the processes leading from A to B. The regularity that As are followed by Bs may be only a "*ceteris paribus* law" (see definition).

2. The second reason is that science looks, and should look, for laws rather than for causes: The laws discovered in advanced sciences have the form of functional equations, in particular differential equations. Russell gives two reasons for why such laws cannot be considered to be causal: First, functional determination is independent of the direction of time, whereas causality is essentially directed from past to future. In mechanics, the same laws, together with the description of a system at time t, can be used to calculate future and past states of the system. The same is true for reversible biochemical reaction equations (these are exceptional, most biologically relevant reactions being irreversible). Moreover, laws expressing the functional dependence of different quantities in the equilibrium state of a system, such as the ideal gas law $pV = nRT$ or the equation indicating the equilibrium constant of a biochemical reaction, cannot be causal because the dependence holds at a given instant of time, whereas causality requires that the cause determines an effect that takes place *later*. Second, laws relate certain properties of objects and events, not particular events that have many properties, whereas causes and effects are particular events.

It is controversial whether Russell was right (1) on fundamental physics and on (2) whether all other sciences, including biology, would eventually come to resemble fundamental physics in abandoning causal concepts (Price and Corry 2007). Independently of the issue of that debate, it has often been argued that the concept of causation is required to make sense, not only of many common sense judgments but also of experimental and applied science.

Several analyses of the meaning of causal statements, both in science and common sense, have been elaborated in philosophy. Until the mid-twentieth century, the deductive-nomological (DN) analysis (definition) was generally accepted as an adequate account of both scientific explanation and causation. However, it has now been widely abandoned because it abolishes the distinction between causal and noncausal scientific explanations. Constitutive explanations seem to be noncausal: The conformation of a macromolecule can be explained in terms of its constituents and the laws applying to the constituents in virtue of their properties, such as the distribution of electric charges. Mechanical explanations can also be seen to be noncausal, insofar as they do not provide the cause of some phenomenon, event, or fact, but show how the parts of a complex system are articulated to guarantee the functioning of the whole system, i.e., its evolution from an initial state to an end state.

Here is a brief sketch of the guiding ideas of the most influential recent accounts of causation. They are grounded on the following concepts respectively: (1) counterfactual dependence, (2) process, (3) probability raising, (4) intervention or manipulation.

1. The counterfactual approach develops the intuition that causes make a difference to their effects. Its leading hypothesis is that the meaning of the statement that event c is the cause of event e can be analyzed in terms of counterfactual conditionals such as: "If c had not occurred, e would not have occurred." David Lewis (1973) has elaborated a semantic analysis of the truth conditions of such counterfactuals. Assume that events c and e have occurred in the actual world. Then "c causes e" is true if and only if the closest of all non-actual

possible worlds in which c does not occur is such that e does not occur in it either. Causation cannot be directly analyzed in terms of counterfactual dependence. Such an analysis fails, e.g., in situations of redundant causation, which are frequent in biology. Assume that a chemical reaction R can be catalyzed by each of two enzymes, E1 and E2, and take a particular situation in which enzyme E1 was active but not E2. In this situation, there is no counterfactual dependence of R on E1, for even if E1 had not catalyzed the reaction, E2 would have. This type of situation is also a case of "preemption": The influence of E2 on R is preempted by the influence of E1. To say that E1 preempts E2 from causing R means that E2 would have caused R in the absence of E1, but that E1 has blocked E2's influence on R. The problem for the counterfactual analysis is that the presence of the preempted cause E2 removes the counterfactual dependence of R on its cause E1. This problem can be solved by allowing that the causal influence from E1 to R runs through a chain of intermediate steps. To cope with other problematic situations, the counterfactual analysis has been modified in several other ways. Furthermore, probabilistic processes have been taken into account: What is counterfactually dependent (directly or indirectly) on the cause is the *probability* of the effect, rather than the effect. Finally, our intuitive concept of causation makes a difference between causes and background conditions. This can be made explicit in counterfactuals: "c rather than c* caused e rather than e*" can be analyzed as "if c* had occurred instead of c, then e* would have occurred instead of e."

2. Another intuition has it that causation requires the existence of a continuous process that stretches from cause to effect. Salmon (1984) has combined Russell's (1948) analysis of causal processes in terms of world-lines (as they have been introduced in special relativity by Minkowski diagrams) and Reichenbach's (1991) idea that causal processes are characterized by their capacity to transmit marks. The paradigm case of this analysis is the propagation of a radio signal, which is a local modification of a structure spreading along a continuous world-line from a sending event to a receiving event. The use of tracers in biomedical research makes sense in the framework of such process accounts: A tracer allows to track causal influence along a "causal line," which is a world-line consisting of a series of contiguous regions of space-time, which are connected by the transmission of the tracer.

A scientist who undertakes uncovering the causal structure of a complex situation may follow two strategies: She may either search for statistical correlations among the variables characterizing the situation, or experimentally manipulate these variables. Each of these research strategies – analysis of observations and experimentation – provides the central idea of a philosophical strategy for analyzing the concept of causation.

3. The main hypothesis of probabilistic analyses of causality (Eells 1991) is that a variable (or factor, or "event" in the language of probability theory) A causes a variable B if and only if $P(B|A) > P(B|\neg A)$. Such theories are usually intended as analyses of generic causation (definition), as opposed to singular causation (definition). However, $P(B|A) > P(B|\neg A)$ is insufficient to justify the judgment that A causes B. It must also be excluded that A and B are effects of some common cause C, which is called a screening factor (definition). Probabilistic analyses can also be adapted to account for situations in which a factor *diminishes* the probability of one of its effects: If factor F, which *enhances* the risk of a disease M, is positively correlated to factor S, which *diminishes* the risk of M, it is possible that $P(M|F) = P(M|\neg F)$ or even $P(M|F) < P(M|\neg F)$, because the negative influence of S on M cancels out (first case) or even overcompensates (second case) the positive influence of F on M.
4. The most important recent analysis of causation is intended as a model of another fundamental research strategy for the discovery of causes in complex situations, as it is used in sciences such as economics, sociology, epidemiology, or systems biology: To find out whether one variable X has a causal influence on another variable Y in the system, one holds fixed all variables Z that cause Y but are not caused by X, and then manipulates X by an intervention (definition) (Woodward 2003). X is a cause of Y if and only if this change of the value of X also changes the value of Y. Ideally, this strategy leads to the discovery of the complete network of influences among all variables

characterizing the system. It can be expressed either by structural equations, which represent the value of each variable as a function of all variables that have a direct causal influence on it, or equivalently by oriented graphs, in which variables are represented by nodes and influences by edges connecting these nodes. However, this analysis cannot be used to discover causal influences from scratch: In order to find out whether X causally influences Y, one must already presuppose knowledge about which other variables influence Y.

By requiring that causes of Y which are not effects of X must be held fixed during the intervention on X, the problem of preemption (mentioned above) is avoided: In the situation sketched above, in which reaction R is caused by enzyme E1 but could also have been caused by "back-up" enzyme E2, the value of E2 is set to 0, because E2 does not intervene in the actual situation. If E2 is "blocked," then manipulation of E1 shows that E1 causes R: The value of variable R is a function of E1.

The interventionist analysis also correctly analyzes situations in which causes diminish the probability of their effect, if a distinction between "total cause" and "contributing cause" is introduced. If F enhances the probability of M but also the probability of S, which in turn diminishes the probability of M, F may not be a *total cause* of M, if the positive (direct) and negative (indirect, through S) effects of F on M just cancel out. However, by holding fixed S, it is possible to discover that F is nevertheless a *contributing cause* of M.

Initially developed to analyze generic causation (among variables), the interventionist framework has recently been adapted to the analysis of token causation: This "actual causation" is modeled by representing influences among particular values of the variables in the network.

In comparing and evaluating such accounts, one must bear in mind that philosophical analyses may pursue different goals: The counterfactual analysis is, e.g., intended to be an a priori analysis of our common sense concept of causation. This implies that it is intended to be applicable to physically impossible situations, such as fictions describing magical interactions. Process accounts have no such ambition. Their aim is to discover the "real essence" of causality as it is in the real world.

References

Quoted in Text

Eells E (1991) Probabilistic causality. Cambridge University Press, Cambridge
Lewis D (1973) Counterfactuals. Basil Blackwell, Oxford
Price H, Corry R (eds) (2007) Causation, physics and the constitution of reality: russell's republic revisited. Oxford University Press, Oxford
Reichenbach H (1991) The direction of time. University of California Press, Berkeley
Russell B (1948) Human knowledge: its scope and limits. Allen and Unwin, London, 5th edition, 1966
Salmon W (1984) Scientific explanation and the causal structure of the world. Princeton University Press, Princeton
Woodward J (2003) Making things happen. Oxford University Press, Oxford

Recent Books and Overview Articles

Beebee H, Hitchcock C, Menzies P (eds) (2009) The Oxford handbook of causation. Oxford University Press, Oxford
Schaffer J (2007) The metaphysics of causation. In: Zalta EN (ed) The Stanford encyclopedia of philosophy, http://plato.stanford.edu/entries/causation-metaphysics/
Sosa E, Tooley M (eds) (1993) Causation. Oxford University Press, New York

Causation

▶ Causality

Causation, Generic and Singular

Max Kistler
IHPST, Université Paris 1 Panthéon-Sorbonne, Paris, France

Definition

Generic causation is a relation between two "factors," "properties," or "events" (in the language of probability theory) represented by variables. The meaning of the statement that smoking (A) (generically) causes lung cancer (B) can be analyzed according to several proposals: In the probabilistic framework, it means that, within a given population, belonging to set A raises the probability of belonging to set B. In the

interventionist framework, it means that an intervention that changes the value of variable A, for example, by making some nonsmokers smokers, changes the value of variable B. It is not possible to draw any inference from the existence of a relation of generic causation between factors A and B in a given population to causal processes at the level of individuals possessing properties A and B (or belonging to sets A and B), in other words to the existence of relations of *singular causation*: In a given population, A (smoking) may cause B (having lung cancer), John may smoke (John is A) and have lung cancer (John is B), and yet, John's smoking may not be the (or even a) cause of his cancer. The cause of John's having cancer may be the fact that he has inhaled asbestos dust.

Cross-References

▶ Causality

Causation, Origination

Niall Palfreyman
Biotechnology and Bioinformatics, Weihenstephan-Triesdorf University of Applied Science, Freising, Germany

Definition

Causation is the physical process assumed by ▶ causality to link one occurrence, the cause, to a second occurrence, the effect, which is understood to occur as a consequence of the cause. A central challenge for the study of causation is *origination*: How are events in the physical world originally caused?

Characteristics

Aristotle distinguishes in *Metaphysics V 2* between four types of cause:

- *Material cause* is the intrinsic physical constitution of an entity: "*The charge of an electron causes it to react to an electric field*."
- *Efficient cause* is the primary exogenous source of change in an entity or system: "*An earthquake caused the tsunami*."
- *Formal cause* is the relational organization which is responsible for properties of a system: "*The charge distribution on a water molecule causes water to be wet*."
- *Final cause* is the purpose or meaning which motivates an action: "*Hunger caused me to hurry home*."

Dynamical systems theory describes causation in terms of 6systems of first-order differential equations $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, t)$, where $\mathbf{x}$ is a vector of state variables of a dynamical system S, t is time, and $\mathbf{f}$ is the set of structural relations which determine the dynamic behavior $\dot{\mathbf{x}}$ of S. This suggests several kinds of formal causation:

- *Upward causation*: Change Δx_i in an individual state variable causes a change $\Delta\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x} + \Delta x_i)$ in the collective macro-behavior of S.
- *Downward causation*: A collective change $\Delta\mathbf{x}$ in all variables causes a change $\Delta\dot{x}_i = f_i(\mathbf{x} + \Delta\mathbf{x})$ in a specific micro-behavior.
- *Reciprocal causation*: A change $\Delta x_{i|j}$ in either of two variables x_i or x_j causes a change $\Delta\dot{x}_{i|j} = f_{i|j}(\mathbf{x} + \Delta x_{j|i})$ in the other's behavior.

These distinctions are of fundamental importance for understanding the origination of phenotypic novelty, environmental effects, and disease. For example, one approach to carcinogenesis seeks efficient, upward causes in the accumulation of genetic mutations within a single originator cell (somatic mutation theory), while another seeks formal, downward cause from tissue-organization fields (tissue-organization field theory).

Physicalism traditionally reduces explanation to material and efficient causes. This reductive program stems from the explanatory benefits of Descartes' and Newton's partitioning of the world into two domains: the *physical* domain of material and efficient cause operating on inert matter, and the *epistemic* domain of final causes operating through human agents. The epistemic domain was thus bracketed out of physical science, and formal cause assumed a twilight role as action-at-a-distance fields.

Physicalism achieved its crowning triumph in the theory of relativity. The special theory is founded on efficient time-like interaction between observers, while in the general theory, geometry becomes the material cause of dynamics. However, mass-energy

equivalence and nonlinearity of the gravitational field equations elevated the relativistic field concept to a new status: formal cause as the bearer of material properties. This inclusion of formal cause into the physicalist fold was consolidated by quantum field theory, which shifts emphasis from particle-bound properties and interactions as material and efficient causes to the formal cause of nonlocal field configurations.

This trend is paralleled in biology, where living systems were traditionally explained in terms of efficient cause: Mayr's proximate developmental causes and ultimate evolutionary causes. Yet this scheme is incomplete on two counts:

- *Developmental systems theory* (DST) (▶ Developmental Systems Theory) stresses the formal reciprocity of gene-organism-environment relations, which complicates views of gene or environment as efficient developmental causes (Oyama et al. 2001).
- Acknowledging gene-plus-environment as efficient evolutionary cause still neglects the role of the organism as the formal confluence of both – particularly with regard to the origination of phenotypic and genotypic novelty (Müller and Newman 2003).

These issues have shifted the focus of biological explanation toward formal cause. Systems biology in particular, in studying the influence of structural and dynamical organization on ontogeny, claims an unambiguous allegiance to formal cause.

The rehabilitation of formal cause has prompted a search for the origination of action in formal cause. A system-level behavior B of a deterministic dynamical system S is *emergent* if it is sufficiently irreducible to micro-behaviors b of S's components as to constitute an originator of efficient cause. Attempts to formalize this notion of irreducibility have led to introduction of the term *supervenience*: B *supervenes* on b if b logically determines B; B is then emergent if it supervenes on no set of micro-behaviors.

A litmus test for coherence of the emergence concept is its relation to downward causation, which permits an emergent behavior $B(t)$ of S at time t to influence a micro-behavior b of S's own components. For example, a weasel's cells generate a white winter coat which is a formal cause of its survival, and thus also of the very cells which generated it. To claim that coloration is non-supervenient on these cellular behaviors seems to invoke a contradictory circular causality.

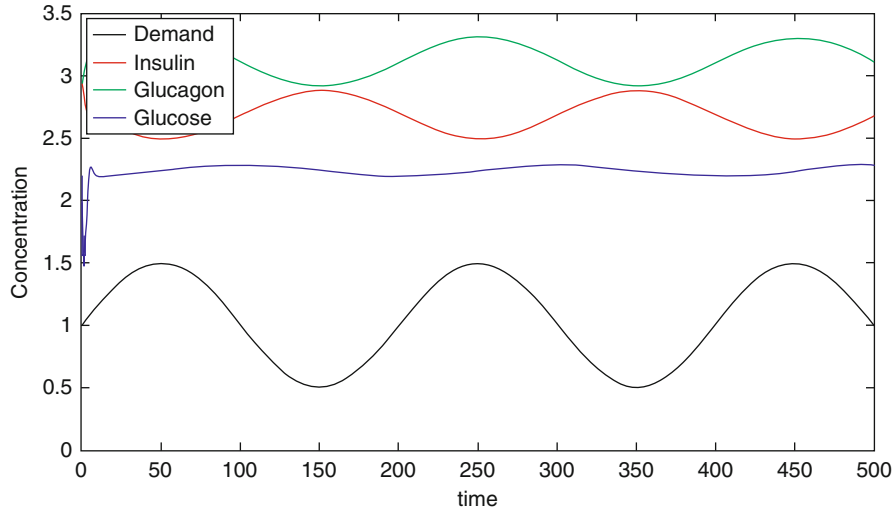
Kim (in Bedau and Humphreys 2008) interprets downward causation in two ways: *synchronic* ($B(t)$ conditions $b(t)$ at the *same* instant t) and *diachronic* ($B(t)$ conditions $b(t + \Delta t)$ at a *later* instant). He argues that synchronic downward causation is logically incoherent, but acknowledges (footnote 37, p. 153) that diachronic downward causation in nonlinear dynamical systems may be feasible.

Pursuing this idea further, Soto et al. (2008) argue for diachronic downward causation. They propose a materialistic non-physicalist ontology in which systems of natural events operate historically to generate ontological novelty (novel qualities and novel structures). Thompson (2007) maintains that diachronic downward causation is coherent in cases where $B(t)$ is non-separable into individual micro-behaviors by reason of its dependence on interactions between these behaviors. This form of emergence is clearly irreducible to individual micro-behaviors, yet even Thompson (pp. 429–30) seems doubtful of its adequacy as a basis for ontological origination. The problem is that dynamic co-emergence cannot escape determination by its component micro-behaviors because even in a nonlinear dynamical system, it still supervenes on the formal organization of the structural relationships **f**.

Seeking a locus of origination in formal cause descends philosophically from the later Wittgenstein, who sought to define meaning in terms of an irreducible closed set of formal language-usage relations, or *language game*. Yet as Putnam (1980) demonstrated, a closed formal system, by abandoning dependence on external context, loses all causal relevance to that context. By definition, it is then difficult to see how such a system can originate from its context.

Appeal to external context suggests a way to emancipate formal cause from internal organization by simply opening S up to interaction with its context. The weasel's coat-change is conditioned in part by the changing autumnal day-length, and so does *not* supervene on the weasel's internal organization. Yet mere openness is an inadequate basis for origination. Inserting a coin into a coffee-machine is a contextual interaction which results in a cup of coffee, yet we consider the coin, not the machine, the efficient cause of this activity, because a small variation, or defect, in the coin prevents coffee-making from occurring.

By contrast, the weasel's color-change is *self-organizing* (▶ Self-Organization): It proceeds even



Causation, Origination, Fig. 1 Self-organized buffering of blood-glucose level against varying demand

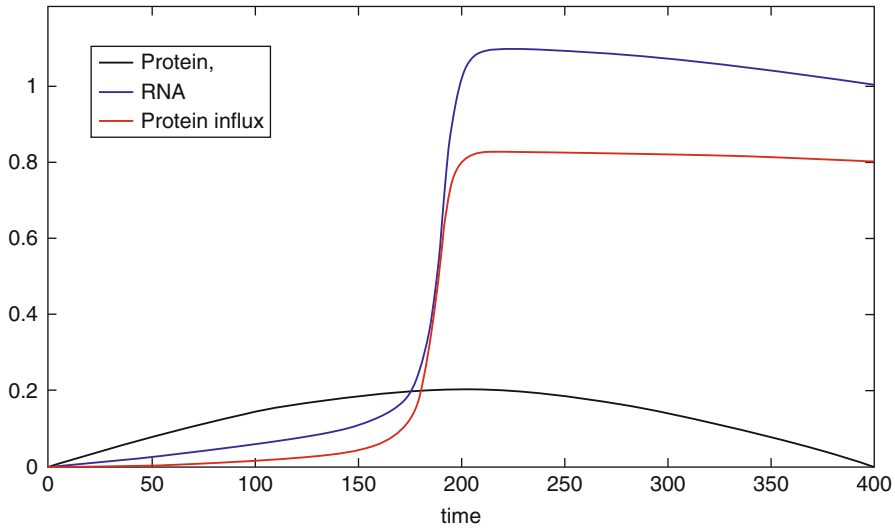
if clouds obscure the day-length. Self-organizing systems generate reliably similar outcomes from dissimilar initial conditions; Fig. 1 illustrates self-organized buffering of blood-glucose against variations in demand in the mammalian insulin-glucagon system. Swenson (2010) uses the term *autocatakinetic* (ACK) to describe systems which self-organize using contextual resources. The weasel is an ACK system which buffers its coat-change against fluctuations in its context, so its winter coat is neither supervenient on the weasel's internal structure nor is it determined by its context. So is the weasel *qua* organism the causal origin of the coloration change?

This is the question of final cause, and goes to the heart of physical-epistemic dualism. In his *Critique of Judgment*, Kant argued for a distinction between physical and biological explanation, characterizing physical systems as *heteronomous* (*other-governed*), and living organisms as the *autonomous* (*self-governing*) originators of final cause. Autonomous systems are certainly ACK, and ACK systems have the capacity to assert their autonomy against the dictates of context; however, Kant's definition of autonomous systems (▶ Autonomy) as final causes depends on their ability to construct meaning from environmental cues, enabling them to predict and respond to future events.

ACK systems can potentially construct meaning, since they necessarily exhibit ▶ complex behavior, bifurcating into qualitatively distinct system behaviors $B(t)$ in response to small quantitative changes in some critical parameter P. If these changes in P stem from S's context, they constitute environmental cues which S amplifies into vigorous responses $B(t)$ (see Fig. 2). However, for Kant, meaning is also essentially personal: S must *itself* determine the dynamics of its own meaning-constructions.

Since S's behavior $\dot{\mathbf{x}}$ is determined by its structure $\mathbf{f}(\mathbf{x}, t)$, we may ask who determines this structure. We might assume the weasel's coloration change to be meaningful, relying as it does on an interpretation of day-length in terms of future snow. However, the structures generating this dynamic arise from propensities of the weasel species-niche system, so the interpretation "snow" does not originate in this individual weasel. The status of organisms as final causes therefore hangs on one question: *Do individual organisms determine their own structure?*

To attribute this determination to efficient cause would be to accept Cartesian dualism; however, the ability of complex open systems to amplify minimal contextual variation makes their context a rich source of internal dynamical variation. If this variation is dynamically coupled to the very structures which generate it, any consequent stabilization of those structures would necessarily be compensated by a corresponding increase in ▶ entropy production by the system. Swenson's (2010) proposed Law of Maximum Entropy Production (LMEP) stipulates



Causation, Origination, Fig. 2 Complex response of a protein-RNA switch to varying protein influx

that such variation is subject to differential selection on its ability to maximize entropy production. LMEP can therefore be expected to drive *dynamic stabilization* (Thompson 2007), the differential selection of behaviors on their ability to stabilize their structural base.

Structures which generate stable, ordered dynamics are thus an expected feature of LMEP selection, raising the question: Who determines this choice of structure? Sober and Wilson (1998) analyze this question in terms of multilevel *targets* of selection – groups of individuals (for example, cells or organisms) whose collective stability depends on interaction with each other, but not with individuals outside the group. On this account, autonomous systems may be understood as ACK systems which determine their own structure not efficiently, but by presenting a unitary target to dynamic stabilization.

In summary, current scientific understanding appears compatible with an account of causation which resolves the physical-biological dualism of Kant. In particular, concepts of final cause, emergence, and origination are compatible with autonomous (dynamically stabilizing, ACK) dynamical systems. These owe their autonomy to their engagement as a systemic unity in *meaning-making*: the dynamic stabilization of meaningful responses through the praxis of meaning-construction. LMEP suggests that autonomous systems are an expected feature of the natural world.

Cross-References

- ▶ Autonomy
- ▶ Causality
- ▶ Complex Behavior
- ▶ Developmental Systems Theory
- ▶ Entropy
- ▶ Self-Organization

References

Bedau MA, Humphreys P (eds) (2008) Emergence. MIT Press, Cambridge, MA
Müller GB, Newman SA (eds) (2003) Origination of organismal form. MIT Press, Cambridge, MA
Oyama S, Griffiths PE, Gray RD (eds) (2001) Cycles of contingency. MIT Press, Cambridge, MA
Putnam H (1980) Models and reality. J Sym Logic 45:464–482
Sober E, Wilson DS (1998) Unto others. Harvard University Press, Cambridge, MA
Soto AM, Sonnenschein C, Miquel P-A (2008) Acta Biotheor 56:257–274
Swenson R (2010) Selection is entailed by self-organization and natural selection is a special case. Biol Theory 5(2):167
Thompson E (2007) Mind in life. Belknap, Cambridge, MA

CD4+ T Cell Response

- ▶ Th1 Response

CD8+ Cytotoxic T Lymphocytes (CTL)

Joo Chuan Tong
Data Mining Department, Institute for Infocomm Research, Singapore, Singapore
Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, Singapore

Definition

CD8+ cytotoxic T lymphocytes are a subgroup of T lymphocytes whose main function is to induce the apoptosis of infected (e.g., viruses, bacteria, parasites, or fungi), mutated (e.g., cancer), or foreign (e.g., transplants) cells.

Cross-References

► TAP Translocator, In Silico Prediction

References

Murphy K, Travers P, Walport M (2007) Janeway's immunobiology, 7th edn. Garland Science, New York, pp 364–368

CDK Inhibitors

Livia Pérez-Hidalgo and Sergio Moreno
Instituto de Biología Molecular y Celular del Cáncer, CSIC/Universidad de Salamanca, Salamanca, Spain

Definition

CDK inhibitors are proteins that suppress CDK-cyclin protein kinase activity in the G1 phase of the cell cycle and promote G1 arrest in response to environmental or intracellular stimuli (► Cyclins and Cyclin-dependent Kinases).

Oscillations in CDK-cyclin activity are accurately regulated in order to ensure the correct sequence of cell cycle events (► Cell Cycle). This regulation occurs at several levels, including cyclin binding to CDKs, CDK phosphorylation, and CDK-cyclin activity inhibition carried out by CDK inhibitors (CKIs). CKIs play a key role in G1 regulation (Sherr and Roberts 1999). In late G1, at the restriction point in mammals and Start in yeasts, cells must decide to progress into a new cell cycle or to enter a quiescent state. After this point, cells become insensitive to mitogenic or antimitogenic signals, and are committed to complete the new cell cycle (Morgan 2007).

Characteristics

Two Families of CKIs Regulate Cell-Cycle Progression and Differentiation

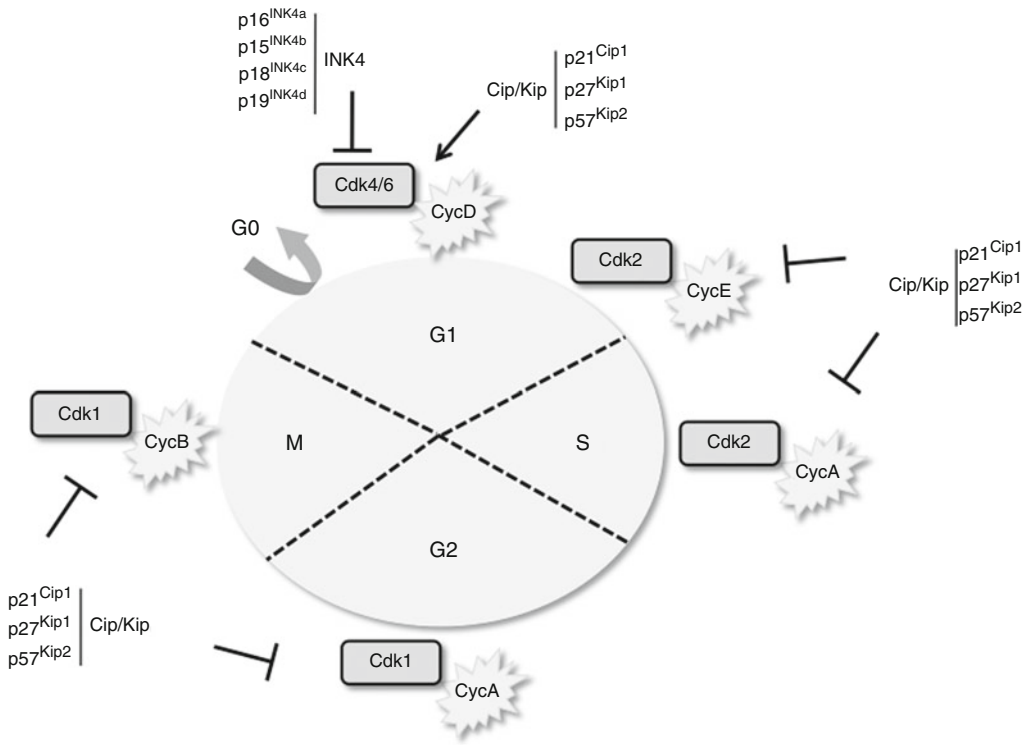
In mammals, CKIs are included in one of two families according to their structure and CDK targets:

1. The INK4 (*IN*hibitors of CD*K4*) family of CKIs includes $p16^{INK4a}$, $p15^{INK4b}$, $p18^{INK4c}$, and $p19^{INK4d}$. They bind CDK4 and CDK6 and form binary complexes with these kinases, preventing CDK interaction with cyclin D. The four INK4 proteins share a similar structure composed of multiple ankyrin repeats (Cánepa et al. 2007).
2. The Cip/Kip (*C*DK-*i*nteracting *p*rotein/*K*inase *i*nhibitor *p*rotein) family of CKIs comprises $p21^{Cip1}$, $p27^{Kip1}$, and $p57^{Kip2}$. These proteins have both a positive and a negative role on CDK activity: they act as cofactors required for CDK4/6-cyclin D complex assembly but inhibit CDK2-cyclin E, CDK2-cyclin A, and CDK1-cyclin B activity. Cip/Kip inhibitors share a conserved N-terminal domain, which binds to CDKs and cyclins, but differ in the rest of the sequence (Besson et al. 2008; ► Cell Cycle of Mammalian Cells) (Fig. 1).

In other organisms, proteins functionally related to mammalian CKIs also play key roles in cell-cycle regulation and in promoting cell-cycle exit during differentiation processes (See Table 1; De Clercq and Inzé 2006).

In *Drosophila melanogaster*, there are two Cip/Kip-related CKIs, Dacapo and Roughex. Dacapo regulate the exit from the cell-cycle during embryonic development, and Roughex contributes to the establishment of G1 through downregulation of CDK-cyclin A complexes.

The *Caenorhabditis elegans* genome encodes two CKIs, cki-1 and cki-2, belonging to the Cip/Kip family of CDK inhibitors, and both are required for cell-cycle exit into quiescence during development.



CDK Inhibitors, Fig. 1 Regulation of cell-cycle progression by CKIs in mammalian cells. Two families of CDK inhibitors in metazoans regulate G1 progression: Cip/Kip proteins and INK4 proteins

In *Xenopus*, Cip/Kip-related protein $p27^{Xic1}$ is involved in cell-cycle exit and differentiation in myogenesis and neurogenesis (▶ Cell Cycle of Early Frog Embryos).

Yeasts' CKIs share no sequence homology with metazoan CKIs, but they are functionally and structurally related. In the budding yeast *Saccharomyces cerevisiae*, two CKIs regulate cell-cycle progression, Sic1 and Far1. Sic1 is required for the establishment of a G1 phase and to delay entry into S phase until ▶ prereplicative complexes are built. In the absence of Sic1, cells show defects in G1 progression, and this impairs S phase dynamics. Cells start DNA replication from fewer origins, extending S phase and entering mitosis with unreplicated chromosomes that inefficiently separate during anaphase. As a consequence, sic1-defective cells show increased genomic instability. All these defects are rescued by delaying CDK activation and S phase, indicating that the primary function of Sic1 is at the end of G1 (Lengronne and Schwob 2002). Although the primary function of Sic1 appears to be at this point of the cycle, Sic1 is also required at mitotic exit, contributing with its accumulation, together with Clb2 degradation, to the irreversibility of this transition (▶ Cell Cycle Dynamics, Irreversibility). Far1 is induced in response to pheromones, leading to the G1 arrest that precedes mating (▶ Cell Cycle, Budding Yeast).

In the fission yeast *Schizosaccharomyces pombe*, a single Sic1-related inhibitor, Rum1, carries out the functions in cell proliferation and differentiation of budding yeast Sic1 and Far1, respectively. Rum1 regulates progression through the G1 phase of the cell cycle and is essential in processes that require lengthening the G1 phase, such as the responses to pheromones and nutrient starvation. Also, Rum1 is required to prevent CDK1 (Cdc2) activation at Start until the critical cell size is reached (Labib and Moreno 1996). (▶ Cell Cycle, Cell Size Regulation; ▶ Cell Cycle, Fission Yeast) (Fig. 2).

CDK Inhibitors, Table 1 CDK inhibitors

Species	Name	Target	Function
Human	$p16^{INK4a}$	CDK4/6-cyclin D inhibition	Prevention of G1/S transition. Induction of senescence.
	$p15^{INK4b}$	CDK4/6-cyclin D inhibition	Prevention of G1/S transition. Mediator of TGFβ cell-cycle inhibitory effects.
	$p18^{INK4c}$	CDK4/6-cyclin D inhibition	Prevention of G1/S transition. Exit from the cell cycle and cell differentiation.
	$p19^{INK4d}$	CDK4/6-cyclin D inhibition	Prevention of G1/S transition. Exit from the cell cycle and cell differentiation.
	$p21^{Cip1}$	CDK4/6-cyclin D assembly and CDK2-cyclin A/E inhibition	DNA-damage-induced cell cycle arrest in G1 and G2.
	$p27^{Kip1}$	CDK4/6-cyclin D assembly and CDK2-cyclin A/E inhibition	Cell-cycle exit into quiescence.
	$p57^{Kip2}$	CDK4/6-cyclin D assembly and CDK2-cyclin A/E inhibition	G1 arrest in embryonic development. Maintenance of endocycles.
Drosophila	Roughex	CDK1-cyclin A	Exit from mitosis. Establishment of G1 phase.
	Dacapo	CDK2-cyclin E	Cell-cycle exit during embryonic development. Establishment of endocycles.
Xenopus	$p27^{Xic1}$	CDK2-cyclin E/A	Establishment of G1 phase in somatic cell cycle. Differentiation in muscle and neural cells.
C. elegans	cki-1	CDK2-cye1(cyclin E)	Developmental cell-cycle quiescence.
	cki-2	CDK2-cye1(cyclin E)	Developmental cell-cycle quiescence.
Fission yeast *S. pombe*	Rum1	Cdc2(CDK1)-Cig2/Cdc13	Regulation of G1 progression. G1 arrest in response to pheromones or nitrogen starvation.
Budding yeast *S. cerevisiae*	Sic1	Cdc28(CDK1)-Clb5/6	Establishment of the G1 phase.
	Far1	Cdc28(CDK1)-Cln1/2/3	G1 arrest and differentiation in response to mating pheromones.

Role of Mammalian CKIs in Tumorigenesis and Phenotypes of CKI Knockouts

Cip/Kip Inhibitors

According to its function limiting cell-cycle progression, $p27^{Kip1}$-deficient mice display cell proliferation defects: increased body size, multiple organ hyperplasia, and pituitary tumors. Unlike other tumor suppressor genes, $p27^{Kip1}$ is rarely mutated in cancer but appears frequently downregulated by increased degradation, decreased transcription or cytoplasmic mislocalization. Besides, the simultaneous inactivation of other ▶ tumor suppressor genes enhances the tumor-prone phenotype of $p27^{Kip1}$ knockout (▶ Cell Cycle, Cancer Cell Cycle and Oncogene Addiction).

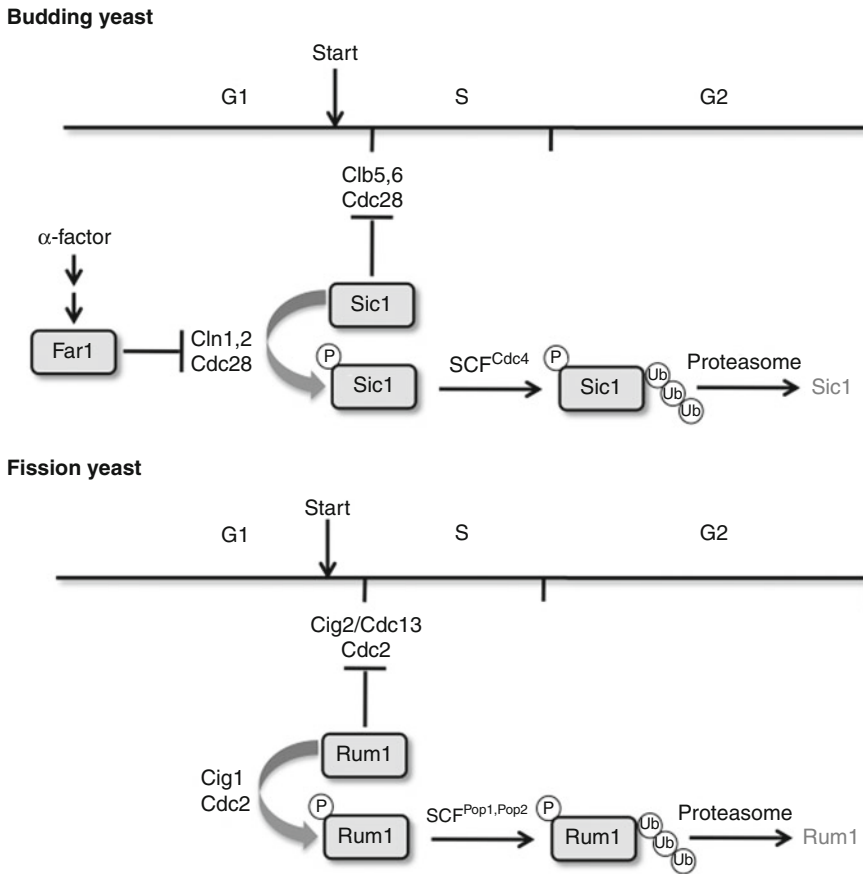
In contrast to $p27^{Kip1}$, $p21^{Cip1}$ knockout mice do not display a hyperproliferative defect. They show increased predisposition to tumor development in combination with mutations in other tumor suppressor genes. As with $p27^{Kip1}$, mutations in $p21^{Cip1}$ gene are rare in tumors. In addition, cells lacking $p21^{Cip1}$ fail to arrest in response to DNA damage (see below).

$p57^{Kip2}$ is required for embryonic development, and $p57^{Kip2}$-null mice have major developmental defects and mostly die perinatally. This inhibitor shows a tissue-specific pattern of expression both in the adult and during embryogenesis.

INK4 Inhibitors

Members of the INK4 family display different patterns of expression: $p16^{INK4a}$ and $p15^{INK4b}$ are not expressed during embryonic development but are induced in response to oncogenic stress, and $p18^{INK4c}$ and $p19^{INK4d}$ are expressed during embryogenesis and at later stages in adults, carrying out their function during organismal development. Mice deficient in any of these proteins are viable.

$p16^{INK4a}$ is an important tumor suppressor that is frequently mutated or downregulated in many different forms of human cancer in contrast to the other three *INK4* genes, which are rarely mutated in cancer. It is not generally expressed during fetal development but accumulates in adults during ▶ senescence. In the



CDK Inhibitors, Fig. 2 Regulation of the cell cycle by CKIs in budding and fission yeast. *Top*: In budding yeast, Sic1 delays entry into S phase through binding to Cdc28-Clb5,6 complexes. At G1/S transition, Sic1 is phosphorylated by Cdc28-CIn1,2, thereby inducing ubiquitylation of Sic1 by SCF^{Cdc4} and proteasome-dependent degradation. In response to pheromones, Far1 induces cell-cycle arrest in G1 prior to cell fusion by inhibiting Cdc28-Cln1,2 complexes. *Bottom*: In fission yeast, Rum1 regulates progression through G1 and induces a cell-cycle arrest in the absence of nutrients and in response to pheromones by inhibiting the activity of Cdc2-Cig2 and Cdc2-Cdc13 complexes. Phosphorylation of Rum1 by Cdc2-Cig1 targets it to ubiquitylation by $SCF^{Pop1,Pop2}$, prior to degradation by the proteasome

absence of $p16^{INK4a}$, $p15^{INK4b}$ has an essential function as a tumor suppressor.

The loss of $p18^{INK4c}$ and $p19^{INK4d}$ leads to male reproductive defects, and mice lacking $p19^{INK4d}$ also display progressive hearing loss. In addition, inactivation of $p18^{INK4c}$ predisposes to cancer development and is a haploinsufficient tumor suppressor in mice. By contrast, $p19^{INK4d}$ null mice do not display tumors nor are more susceptible to cancer development in response to carcinogens.

Function of $p21^{Cip1}$ in the DNA-Damage Response

$p21^{Cip1}$ increases following DNA damage in a p53-dependent manner, and it makes an essential contribution to the DNA-damage response in human cells regulating multiple processes, such as cell cycle progression, ▶ apoptosis, and transcription (Cazzalini et al. 2010).

Regulation of cell cycle progression occurs by inhibition of CDK2-cyclin E and CDK2-cyclin A complexes, required for phosphorylation of ▶ retinoblastoma, but also by binding to PCNA (proliferating cell nuclear antigen), which interferes with PCNA function during DNA replication.

Another part of the response of $p21^{Cip1}$ to DNA damage is transcriptional regulation, where $p21^{Cip1}$ has both a positive and a negative role. This regulation occurs through different mechanisms. Inhibition of

CDK activity prevents retinoblastoma phosphorylation, thus leading to inhibition of E2F-dependent transcription. Besides, $p21^{Cip1}$ associates with the DNA-binding proteins E2F1, STAT3, and MYC, suppressing their transcriptional activity, and derepresses p300-CREBBP targets.

In contrast to the role of p53 in the induction of apoptosis in response to DNA damage and oxidative stress, $p21^{Cip1}$ protects cells from apoptosis, although in some cases it may show a pro-apoptotic role. Since an active cell cycle is required for apoptosis, $p21^{Cip1}$ may prevent apoptosis through its role in suppressing cell-cycle progression, but also it may inhibit apoptosis directly, blocking pro-caspase-3 processing or preventing stress-induced apoptosis by binding to ASK1/MEKK5 or JNK1/SAPK.

Role of CKIs in Endocycles

CKIs play an essential role in the control of CDK activity in endocycling cells, which undergo multiple rounds of DNA replication without an intervening mitosis (Ullah et al. 2009).

$p57^{kip2}$ triggers genome ▶ endoreplication during differentiation of trophoblast stem (TS) cells into trophoblast giant (TG) cells. It is expressed at the end of S phase, preventing entry into mitosis and inducing the accumulation of cells in G2. As CDK activity is low, ▶ APC^{Cdh1} is activated instead of APC^{Cdc20}. One of the targets of APC^{Cdh1} is geminin, an inhibitor of the licensing factor Cdt1. The destruction of geminin and other targets, such as cyclin A, creates a window in G2 with low CDK activity, which permits the formation of prereplicative complexes. $p57^{kip2}$ disappears after phosphorylation by CDK2-cyclin E and degradation by the 26S ▶ proteasome, to allow the onset of S phase. Oscillations of $p57^{kip2}$ and cyclin E protein levels maintain the endocycles. In *Drosophila*, endocycles in nurse cells are maintained through oscillations in the levels of Cyclin E and the CKI Dacapo.

Regulation of CKIs by Mitogenic and Antimitogenic Signals

In mammalian cells, in response to mitogenic signals and activation of the Ras-MAP kinase pathway, cyclin Ds are synthesized, and stable cyclin Ds-CDK complexes containing the $p27^{kip1}$ inhibitor as a cofactor form, thus taking $p27^{kip1}$ from CDK2-cyclin E/A complexes. This leads to the activation of these complexes and the G1/S transition. Mitogenic stimulation also inhibits, through PI3K and AKT, GSK3β phosphorylation of cyclin Ds, thereby inhibiting their nuclear export and enhancing the nuclear accumulation of active cyclin Ds-CDK complexes. Inhibition of GSK3β also stimulates expression of cyclin Ds through regulation of the transcription factors AP-1 and Myc.

Transforming growth factor (TGFβ) antimitogenic signal induce a G1 block, causing synthesis of CKIs $p15^{INK4b}$ and $p21^{Cip1}$. $p15^{INK4b}$ levels increase through the Smad complex. Smad proteins inhibit expression of Myc, and, after Myc has been repressed, expression of $p15^{INK4b}$ increases (Morgan 2007).

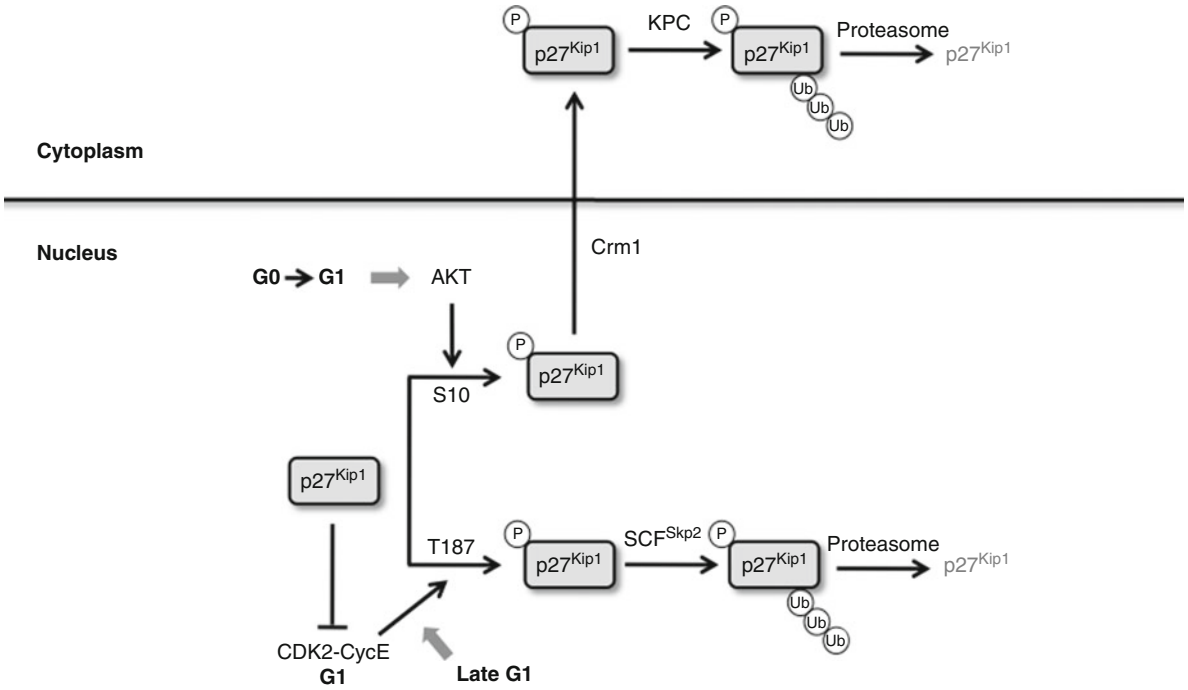
Degradation of CKIs at the G1/S Transition

Irreversible entry into the cell cycle at Start is controlled by positive ▶ feedback loop that generate irreversible CDK activation (▶ Cell Cycle Dynamics, Irreversibility). Proteolytic degradation of CKIs at G1/S, triggered by a starter kinase, is essential for Start in yeast and mammalian cells (Morgan 2007; Novak et al. 2007).

In human cells, different pathways lead to $p27^{kip1}$ phosphorylation, ubiquitylation, and proteasome-mediated destruction at G1/S:

1. In late G1, $p27^{kip1}$ is phosphorylated at T187 by CDK2-cyclin E and CDK2-cyclin A complexes. This phosphorylation targets $p27^{kip1}$ for ubiquitylation by SCF^{Skp2} (▶ Skp1/Cul1/F-Box-containing Complex (SCF)) and subsequent degradation by the 26S proteasome. The interaction between $p27^{kip1}$ and Skp2 is enhanced by additional factors, such as Cks1 (CDK subunit 1). Also, interaction between CDK2-cyclin A and T187-phosphorylated $p27^{kip1}$ stimulates $p27^{kip1}$ ubiquitylation by SCF^{Skp2}.
2. Phosphorylation of $p27^{kip1}$ at S10 by AKT promotes its nuclear export through interaction with the nuclear export protein Crm1. The stability of cytoplasmic $p27^{kip1}$ is regulated by the E3 ligase KPC (Kip1 ubiquitylation-promoting complex), which targets it to degradation by the proteasome (Fig. 3).

In addition, G1/S transition in cycling cells and after DNA-damage repair requires degradation of $p21^{Cip}$. $p21^{Cip1}$ is degraded at the proteasome by ubiquitin-dependent and ubiquitin-independent



CDK Inhibitors, Fig. 3 Regulation of $p27^{Kip1}$ by phosphorylation and proteasome degradation. CDK2-cyclin E phosphorylates $p27^{Kip1}$, promoting its ubiquitylation by SCF^{Skp2} and degradation at the proteasome. Phosphorylation of $p27^{Kip1}$ by AKT promotes $p27^{Kip1}$ translocation to the cytoplasm and KPC-dependent ubiquitylation

mechanisms. Phosphorylation by CDK2-cyclin E at S130 promotes SCF^{Skp2}-dependent degradation, which occurs in part after translocation of phosphorylated $p21^{Cip1}$ to the cytoplasm. $p21^{Cip1}$ reacummulates again in G2, and it is degraded at the G2/M transition by the proteasome after ubiquitylation by the E3 ligase APC^{Cdc20}. In addition, the $CRL4^{Cdt2}$ E3 ubiquitin ligase promotes $p21^{Cip1}$ degradation during S phase and in response to damage (Lu and Hunter 2010).

In yeasts, CKI degradation is initiated by phosphorylation of the CKI by CDK-cyclin complexes different from those inhibited by the CKI:

In fission yeast, phosphorylation of Rum1 at T58 and T62 by CDK1-Cig1 is required for Rum1 ubiquitylation by the $SCF^{Pop1,Pop2}$ and degradation by the 26S proteasome, thus leading to CDK1-Cig2/Cdc13 activation.

In budding yeast, G1/S Sic1 destruction by the proteasome depends on phosphorylation at multiple CDK consensus sites (S/T)PX(K/R) by CDK1 (Cdc28)-Cln1/2 complexes. Once phosphorylated, Sic1 binds to SCF^{Cdc4}, which ubiquitylates Sic1 and targets it for destruction by the proteasome, leading to CDK1-Clb5/6 activation and the onset of S phase.

Cross-References

- ▶ Anaphase-Promoting Complex (APC/C)
- ▶ Apoptosis
- ▶ Cell Cycle
- ▶ Cell Cycle of Early Frog Embryos
- ▶ Cell Cycle of Mammalian Cells
- ▶ Cell Cycle Transition, Principles of Restriction Point
- ▶ Cell Cycle, Budding Yeast
- ▶ Cell Cycle, Cancer Cell Cycle and Oncogene Addiction
- ▶ Cell Cycle, Cell Size Regulation
- ▶ Cell Cycle Dynamics, Irreversibility
- ▶ Cell Cycle, Fission Yeast
- ▶ Cyclins and Cyclin-dependent Kinases
- ▶ Endoreplication
- ▶ Feedback Loops

- ▶ Prereplicative Complex
- ▶ Proteasome
- ▶ Retinoblastoma
- ▶ Senescence
- ▶ Skp1/Cul1/F-Box-containing Complex (SCF)
- ▶ Tumor Suppressor Gene

References

Besson A, Dowdy SF, Roberts JM (2008) CDK inhibitors: cell cycle regulators and beyond. Dev Cell 14:159–169

Cánepa ET, Scassa ME, Ceruti JM, Marazita MC, Carcagno AL, Sirkin PF, Ogara MF (2007) INK4 proteins, a family of mammalian CDK inhibitors with novel biological functions. IUBMB Life 59:419–426

Cazzalini O, Scovassi AI, Savio M, Stivala LA, Prosperi E (2010) Multiple roles of the cell cycle inhibitor p21(CDKN1A) in the DNA damage response. Mutat Res 704:12–20

De Clercq A, Inzé D (2006) Cyclin-dependent kinase inhibitors in yeast, animals, and plants: a functional comparison. Crit Rev Biochem Mol Biol 41:293–313

Labib K, Moreno S (1996) rum1: a CDK inhibitor regulating G1 progression in fission yeast. Trends Cell Biol 6:62–66

Lengronne A, Schwob E (2002) The yeast CDK inhibitor Sic1 prevents genomic instability by promoting replication origin licensing in late G1. Mol Cell 9:1067–1078

Lu Z, Hunter T (2010) Ubiquitylation and proteasomal degradation of the p21(Cip1), p27(Kip1) and p57(Kip2) CDK inhibitors. Cell Cycle 9:2342–2352

Morgan DO (2007) The cell cycle: principles of control. New Science, London

Novak B, Tyson JJ, Gyorffy B, Csikasz-Nagy A (2007) Irreversible cell-cycle transitions are due to systems-level feedback. Nat Cell Biol 9:724–728

Sherr CJ, Roberts JM (1999) CDK inhibitor: positive and negative regulators of G1-phase progression. Genes Dev 13:1501–1512

Ullah Z, Lee CY, Depamphilis ML (2009) Cip/Kip cyclin-dependent protein kinase inhibitors and the road to polyploidy. Cell Div 4:10

cDNA Microarrays

- ▶ DNA Microarrays

CDR-IMGT

- ▶ Complementarity Determining Region (CDR-IMGT)

Cell Compartment

- ▶ Organelle and Functional Module Resources

Cell Cycle

Attila Csikász-Nagy
The Microsoft Research - University of Trento Centre for Computational and Systems Biology, Trento, Italy

Introduction

Cell cycle is the sequence of events whereby a growing cell duplicates all of its components and divides them between two daughter cells. This biological process is essential for reproduction, thus, it is a basic principle of life. There is a long history of cell cycle research and the discoveries of the key regulators of the underlying machinery have been recognized by Nobel Prize awards. The cell cycle was also an early target of mathematical modeling and pioneering work in systems biology. One of the first systems biology approaches realizing the experiment-model-experiment loop was concerned with the cell cycle regulation of budding yeast cells as well as some of the earliest system-level measurements of this system. Further research on cell cycle regulation in various organisms included diverse systems biology techniques. This work has made considerable contribution to our understanding of key principles of cell cycle regulation. Nevertheless, there are still many open questions that could be answered by a combination of computational and experimental approaches.

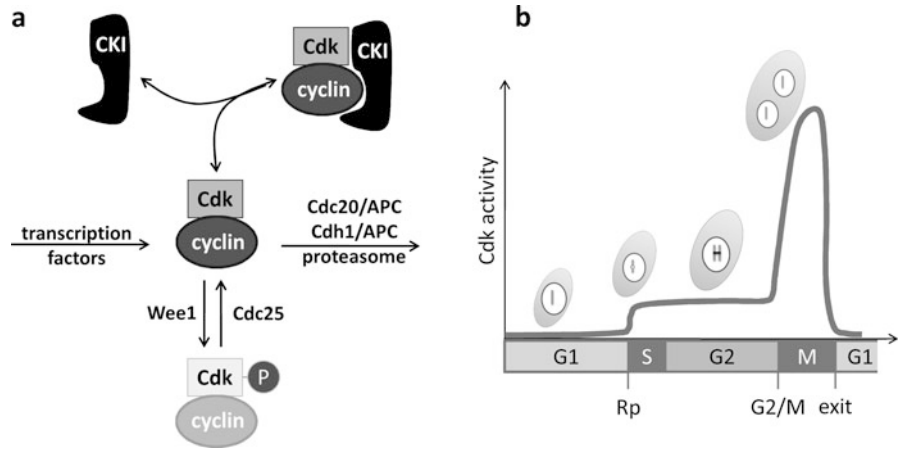
In this essay, we cover the basics of cell cycle physiology, molecular biology and the modeling the underlying processes and mechanisms. This discussion serves as a context for all articles on the cell cycle. Here, we cover historical discoveries both on the experimental and theoretical sides, explaining what led to a Nobel Prize in cell cycle research and how mathematical models of the cell cycle contributed to the success of systems biology.

Biology of the Cell Cycle

Adequate cell cycle regulation is needed to ensure that each daughter cell inherits accurate copies of the mother's DNA, thus cell cycle separates the duplication and the separation of chromosomes in time. During S-phase of the cell cycle the ▶ DNA replication takes place while later in M-phase, the chromosome segregating ▶ mitosis occurs. These two phases are separated by two gap phases, leading to a full cell cycle with G1-S-G2-M phases in this order (Morgan 2006). At the end of M-phase, the cell nucleus divides, which is followed by ▶ cytokinesis when the two daughter cells separate. The gap phases are also needed for proper homeostasis of cells: the newborn daughter cells should have a cell size more or less similar to the birth size of the mother cell to keep the cell size close to constant between generations. The G1 and G2 phases are used to extend the required time to pick up this extra cell growth that is needed to duplicate cell size before cell division. This balance between cell growth and division in many cell types is established by active cell size regulation (▶ Cell Cycle, Cell Size Regulation): either in G1 or G2 phase the cells check if their size is large enough to, so that the cell can proceed to the next cell cycle stage (S or M-phase respectively). Proper cell cycle physiology (▶ Cell Cycle, Physiology) also requires that the S and M phases happen in the correct order. This is achieved through ▶ cell cycle checkpoints. The above mentioned G1 and G2 phase size checks occur together with a check on the successful completion of the earlier mitosis (at the G1 to S checkpoint) or DNA replication (at the G2 to M checkpoint). These checks ensure that the DNA is replicated only after it is successfully segregated and cells start mitosis only after DNA replication is completed (Hartwell and Weinert 1989; Lukas et al. 2004). A third checkpoint exists in M-phase, where the cell ensures that chromosome segregation starts only when all chromosomes are properly bound to the separation machinery (Ciliberto and Shah 2009). Several other checkpoints exist in which cells monitor external and internal conditions and proceed only if all checks are passed. An example is the budding yeast morphogenesis checkpoint, where cells check if the bud formation is properly initiated, before mitosis starts (Sia et al. 1998).

These checkpoints work through a cross-talking network of signaling pathways that talk to the core cell cycle machinery to halt progress as long as problems remain. Metabolic pathways inform the cell cycle (▶ Cell Cycle Signaling, Metabolic Pathway) about the presence or absence of nutrients in the environment. Stress pathways (▶ Modeling Approaches of Cell Cycle Regulation by Signaling Pathways for External Stress) monitor other external and internal conditions, such as osmotic pressure, oxidative state of the cells (▶ Cell Cycle Signaling, Hypoxia), DNA (▶ Cell Cycle Arrest After DNA Damage), and spindle damage (▶ Cell Cycle Signaling, Spindle Assembly Checkpoint). All of these pathways control activities of key proteins in the core cell cycle machinery.

The central members of this core cell cycle machinery are the complexes of ▶ cyclins and cyclin-dependent kinases. These complexes can induce the phosphorylation of various target molecules that signal to downstream cell cycle processes. Such phosphorylation events control the initiation of DNA replication and mitosis directly by the effect of cyclin-dependent kinases or through the regulation of centrosomal and mitotic kinases. The cyclin-dependent kinase (Cdk) can phosphorylate substrates only if a cyclin molecule is bound to it. This cyclin is responsible for the substrate recognition of the complex (Bloom and Cross 2007). As the name suggests, most cyclins appear only periodically during the cell cycle, their production and degradation can be regulated by both external and internal signals of the cell. The activity of Cdk/cyclin complexes are regulated throughout the cell cycle by control of the synthesis or degradation of cyclins (Zachariae and Nasmyth 1999), by binding the complex to stoichometric ▶ Cdk inhibitors (CKI), or by controlling the kinase activity of Cdk through phosphorylation of its regulatory sites (Lindqvist et al. 2009) (Fig. 1a). These Cdk regulators ensure that Cdk activity is low in G1 phase, it appears at the G1/S transition point (also called Start or Restriction point (▶ Cell Cycle Transition, Principles of Restriction Point)), it increases in G2 phase and reaches its maxima at the G2/M transition (▶ Cell Cycle Transitions, G2/M), and it sharply drops at mitotic exit (▶ Cell Cycle Transitions, Mitotic Exit) (Fig. 1b). Checkpoints freeze Cdk activity, thus, the subsequent cell cycle transition cannot be initiated until the checkpoint signal is not removed: DNA damage does not allow the sharp increase in Cdk activity and the spindle checkpoint inhibits the drop in Cdk activity, ensuring that the upcoming cell cycle transitions are inhibited.



Cell Cycle, Fig. 1 *Regulation of Cdk activity during the cell cycle*. (**a**) Possible ways Cdk activity is regulated: CKI – Stochiometric Cdk inhibitor acts in G1, where Cdh1/APC complex induces cyclin degradation through the proteasome. Synthesis of cyclin could be also regulated by transcriptional factors. Wee1 and Cdc25 are the most important kinase/phosphatase pair that regulates Cdk in G2 phase and Cdc20/APC induces cyclin removal at the end of the cell cycle. (**b**) Temporal pattern in Cdk activity during the cell cycle and cell cycle transitions (*Rp* – Restriction point, exit – mitotic exit). In S phase, the DNA replicates, in G2 two copies are present and Cdk activity is low but not zero. Sharp increase in Cdk activity induces mitosis, where the nuclei divide and the drop in the activity triggers cell division

Crucial functions of the core cell cycle machinery in eukaryotic cells are the following:

- The alternation of DNA replication (S-phase) and mitosis (M-phase) is essential in normally proliferating somatic cells. This is achieved by changing Cdk activity in each phase, and ensuring that the cycle proceeds only in one direction. Incidental (or regulated) drop in Cdk activity in G2 phase can induce ▶ endoreplication, when two rounds of S phase happen without mitosis and two nuclear divisions occur in ▶ meiosis, when the DNA content is halved. These processes normally occur only during development and embryogenesis, they are avoided in normal somatic cells by careful control of Cdk activity.
- Checkpoint mechanisms stop the cell cycle, by freezing Cdk activity, if some previous event has not been properly completed and ensure that repair mechanisms are initiated to "solve" the problem. In higher eukaryotes, similar pathways eventually induce the suicide of the cell if the repair fails to avoid the spreading of the mutant, which could lead to tumor formation (▶ Cell Cycle, Cancer Cell Cycle and Oncogene Addiction).
- Growth in cellular mass could be rate-limiting, thus, the DNA replication-division cycle is "slowed down" by inserting gap phases (G1- and G2-phase). Such size control (▶ Cell Cycle, Cell Size Regulation) clearly works in yeast cells, but its role in higher eukaryotes is still not fully understood.

The basic controls are conserved in all eukaryotic organisms from yeast to human. Indeed, human Cdks can rescue Cdk mutants of yeast cells and many other core molecules are similarly highly conserved. Although the cell cycle regulatory molecules of prokaryotes (▶ Cell Cycle, Prokaryotes) are highly dissimilar, their interaction network and the resulting cell cycle dynamics are very similar to the yeast system.

Historical Discoveries of Cell Cycle Research

Systematic analysis of cell cycle mutants in the 1970s by Lee Hartwell and Paul Nurse led to the discovery of Cdk while cyclin was discovered by Tim Hunt. These researchers received the Nobel Prize in 2001 for their breakthroughs in understanding cell cycle regulation (Nasmyth 2001).

The first results toward the identification of a biochemical component governing cell cycle progression came from investigations on cell extracts from frog eggs (▶ Cell Cycle of Early Frog Embryos). Studies in the early 1970s described the existence of

a *maturation promoting factor* (MPF) whose function was related to induction of quiescent cells into division (Masui and Markert 1971). At the same time, yeast genetic approaches lead to the discovery of the most important genes of cell cycle regulators: Hartwell identified cell division cycle (Cdc) genes in the budding yeast *Saccharomyces cerevisiae* (▶ Cell Cycle, Budding Yeast) through a genetic screen of temperature-sensitive mutants (Hartwell et al. 1973). Paul Nurse isolated temperature-sensitive mutants of the rod shaped fission yeast, *Schizosaccharomyces pombe* (▶ Cell Cycle, Fission Yeast), and found cell-cycle-blocked long cells, but also noticed unusually small cells that seemed to be able to proliferate normally (Nurse 1975). He named this mutant "wee" (meaning small in Scottish) and realized that the related genes must have an important role in controlling proper cell division. Nurse also figured out that one of the genes with wee phenotype can be mutated in another domain to produce cell-cycle-blocked elongated cells, discovered that this gene (Cdc2) is homologues to Hartwell's most important budding yeast gene, and found the human version of it (▶ Cell Cycle of Mammalian Cells). Tim Hunt contributed to the previous work of Hartwell and Nurse when he studied the control of mRNA translation and found proteins whose abundance oscillated (Evans et al. 1983). He referred to these periodically degraded molecules as "cyclins." Later discoveries showed that the complex of cyclin and Cdc2 (generally now called Cdk) is responsible for the MPF activity in frog embryos. After these breakthrough discoveries, several cell cycle regulators and their functions have been identified in various organisms. We learned that similar molecules control the cell cycle of plants, worms, and flies and began to understand the intricacies of their regulation in the different organisms. Indeed, not only cyclin and Cdk proteins are conserved among eukaryotes, but most of the cell cycle regulator proteins as well as their interactions (Nurse 1990). Here, we present the generic features of the cell cycle regulatory network (Csikasz-Nagy et al. 2006) together with the regulators in most important cell cycle test organisms (Fig. 2 and Table 1).

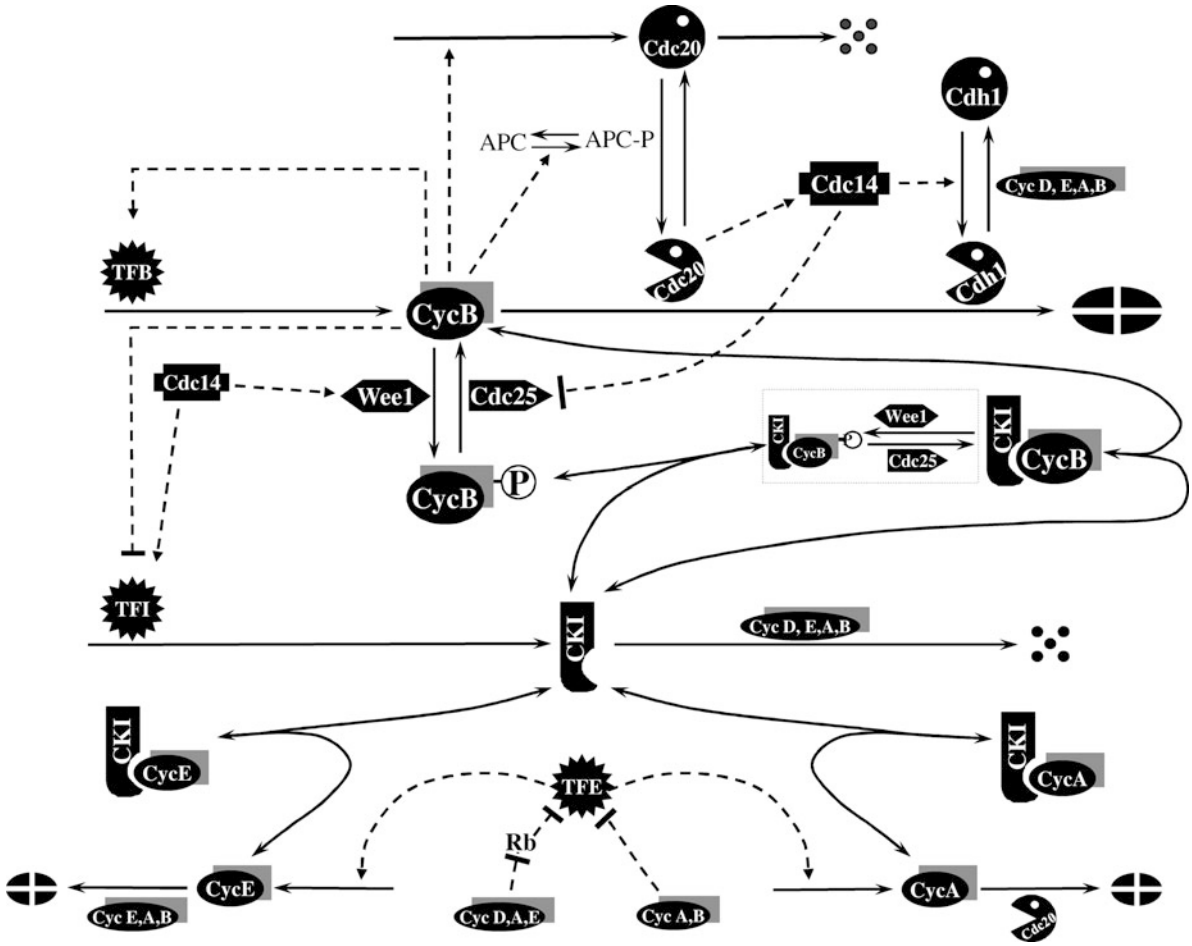
These interactions can drive the periodic appearances of the various cyclins. CycD is present throughout the entire cell cycle, CycE, and CycA appears at the G1/S transition restriction point and CycB activity increases at G2/M transition and drops at mitotic exit. The system-level interactions depicted in Fig. 2 cannot be easily understood by reductionistic thinking. Mathematical and computational models can help to understand such wiring diagrams and test if the current knowledge on molecular interactions is sufficient to describe experimental observations (Csikasz-Nagy 2009). Figure 3 shows simulation results of the wiring diagram of Fig. 2. It explains how the waves of the different cyclins overlap with the inactivation of Cdk/cyclin inhibitors (CKI, Cdh1 and Wee1).

Initial Steps in Mathematical Modeling of the Cell Cycle

Mathematical modeling of the cell cycle has been started already when scientists were only guessing the mechanisms that could drive the processes of the cell division cycle. Early efforts considered the phenomenon as a black box but could still discover some of the rules that determine the observed physiology of cells. From the 1960s, we can find mathematical models that explain some key aspects of cell cycle regulation from phenomenological observations on cell size and cell cycle time distributions (Koch and Schaechter 1962). The great interest in the newly discovered chemical oscillations in the 60s made considerable contributions to research on theoretical physical chemistry and later to mathematical biology as well. Researchers, such as Albert Goldbeter, John J. Tyson, Arthur Winfree, and others, realized that the theories and tools to analyze chemical oscillations might help to understand biological oscillations. They started to investigate calcium oscillations, circadian clocks, and many other biological oscillators, among them the oscillations that drive cell division. As the first experimental results on cell cycle regulation appeared (see above), theoreticians started to create models to predict how Cdk controls the cell cycle. Some of these models led to a better understanding of the basic dynamical properties of the cell cycle regulatory network (Csikasz-Nagy 2009).

Predictive Models of the Cell Cycle

In the early 1990s, several hypotheses emerged concerning the origin of cell cycle oscillations. Theory tells us that for oscillations the system needed to



Cell Cycle, Fig. 2 *Generic cell cycle regulatory network. Solid arrows* depict molecular transitions, molecules "sitting" on *arrows* and *dashed arrows* represent activations; *dashed blunted end lines* stand for inhibition effects. *Gray squares* behind cyclins (CycA-D) represent their constantly bound Cdk partners

contain a negative feedback loop (auto-inactivation), some nonlinearity, and either a delay or a positive feedback loop (auto-activation) (Novak and Tyson 2008). For the cell cycle oscillations, Albert Goldbeter proposed a molecular cascade mechanism that produces a limit cycle originating from a pure negative feedback of Cdk and its degradation machinery (Goldbeter 1991). John J. Tyson and Bela Novak predicted a more complex mechanism producing a relaxation oscillator that moves around a hysteresis loop as a result of a combination of positive and negative feedbacks (Novak and Tyson 1993). The Novak-Tyson model proposed that the positive feedback loop between Cdk/CycB and Cdc25 and the double negative (thus also positive) loop between Cdk/CycB and Wee 1 create a situation where cyclin levels needs to reach a critical value to induce an abrupt activation of Cdk/CycB.

Nonlinear positive feedback loops can create multistability (Ferrell 2002) when, in a given parameter range, the system can exist in more than one stable states and only the history of the system determines in which of these states we will find it (this is the concept of hysteresis). Transitions from one stable state to another can happen very abruptly when one of the controlling parameters reaches a critical value. To set back the system to the original state, the same parameter has to decrease to a much lower value. This can ensure that transitions between the two stable states are quite robust, for small fluctuation these transitions are irreversible. Now we understand the importance of hysteresis and irreversible switches (▶ Cell Cycle Dynamics,

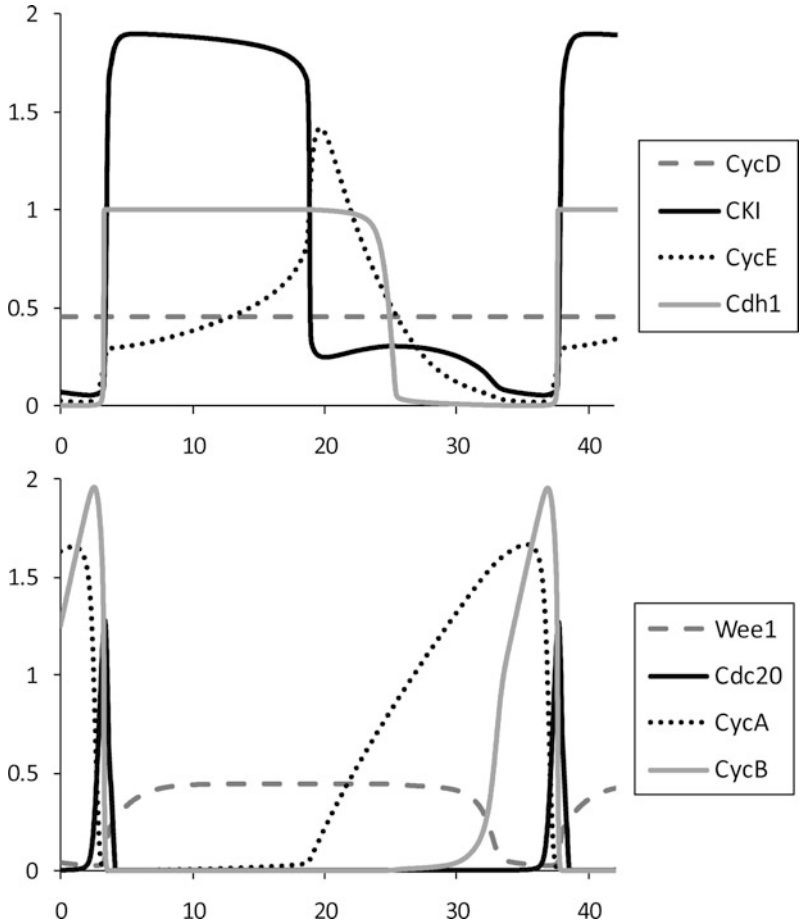
Cell Cycle, Table 1 List of key cell cycle regulators in the most important cell cycle research model organisms

Function	Name in Fig. 2	Budding yeast	Fission yeast	*Xenopus laevis*	Mammalian
Starter CDK/cyclin complex	CycD	Cdc28/Cln3	Cdc2/Puc1	Cdk4,6/CycD	Cdk4,6/CycD1–3
(G1/S transition inducers)	CycE	Cdc28/Cln1,2	-	Cdk2/CycE	Cdk2/CycE1,2
S-phase promoting factor (SPF)	CycA	Cdc28/Clb5,6	Cdc2/Cig2	Cdk1,2/CycA	Cdk1,2/CycA1,2
M-phase promoting factor (MPF)	CycB	Cdc28/Clb1,2	Cdc2/Cdc13	Cdc2/CycB	Cdk1/CycB1,2
Stoichiometric kinase inhibitor	CKI	Sic1	Rum1	Xic1	$p27^{Kip1}$
Mitotic cyclin degradation regulator (with APC)	Cdh1	Cdh1 (Hct1)	Ste9 (Srw1)	Fzr	Cdh1
Cyclin B, Cyclin A degradation regulator (with APC)	Cdc20	Cdc20	Slp1	Fizzy	$p55^{Cdc}$
Retinoblastoma protein	Rb	Whi5	-	RBL1,2	Rb1
Cyclin E, Cyclin A transcription factor	TFE	Swi4/Swi6 Mbp1/Swi6	Cdc10/Res1	XE2F	E2F1–3
CDK/cyclin B inhibitor kinase	Wee1	Swe1	Wee1	Xwee1	Wee1
CDK/cyclin B activator phosphatase	Cdc25	Mih1	Cdc25	Xcdc25	Cdc25C
Phosphatase working against the CDKs	Cdc14	Cdc14	Clp1/Flp1	Xcdc14	Cdc14

Irreversibility) in the cell cycle: the cell must complete one state before starting another event, avoiding the same phase to repeat twice (Novak et al. 2007).

As a result of the positive feedback in the Novak-Tyson model, the removal of some CycB cannot inactivate Cdk immediately – showing hysteresis and bistability in the cell cycle (▶ Cell Cycle Dynamics, Bistability and Oscillations). This prediction was independently experimentally verified by two groups (Pomerening et al. 2003; Sha et al. 2003), and it was also discovered that by removing the negative effect of Wee1 on Cdk (thus disrupting the positive feedback loop), the oscillations became less robust and their amplitude was reduced (Pomerening et al. 2005). These insights highlighted the importance of positive feedbacks in cell cycle regulation.

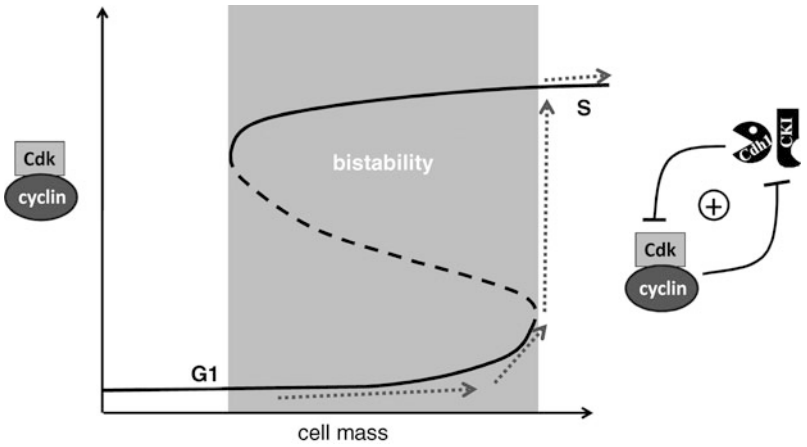
Hysteresis and multistability was also proposed by mathematical modeling and verified later experimentally for the cell cycle regulation of yeast cells. The two landmark papers by Katherine C. Chen and colleagues from the Tyson lab dealt with the cell cycle of budding yeast cells. The models that were created in this work are now the influential mathematical models of cell cycle regulation (Chen et al. 2004; Chen et al. 2000). These models are based on an extensive literature data collection on more than 100 budding yeast cell cycle mutants. The final model can simulate the findings of almost all of these experiments. In their first paper, Chen et al. (2000) examined the molecular events regulating the START transition of the cell cycle. The model accounts for many details of the cells' physiology (viability-lethality, cell size, lengths of different cell cycle phases) in wild-type and ~50 mutant cells. The authors proposed that G1 and S/G2/M phases are alternative states and cells switch between them during the G1/S transition, when CDK activity abruptly increases. The bistability of the system is given by the antagonism between CDK/cyclin complexes and their inhibitors CKI (Sic1 in budding yeast) and Cdh1 (Fig. 4). The authors also proposed an experiment which could verify the existence of this bistability. In 2002, Fredrick R. Cross and his group performed experiments following these suggestions and confirmed the prediction (Cross et al. 2002). In this groundbreaking study, Cross went much further and carried extensive experimental tests of various properties of the model while also performing some crucial measurements to help further model development. This milestone study was one of the first that was fully devoted to test a mathematical model, and could therefore be viewed as one of the first true systems biology studies. Two years later, the measurements and corrections by Cross were implemented into a new version of Chen's model (Chen et al. 2004). It was also extended to cover detailed regulation of mitotic exit. The resulting model was tested against the behavior of 131 mutants and marginally failed only on 11 of them – highlighting the aspects of the model that lacked sufficient experimental evidence. This model also predicted the presence of a mitotic exit regulatory phosphatase that was later experimentally identified (Queralt et al. 2006).



Cell Cycle, Fig. 3 *Simulation of a "generic" cell cycle model.* Temporal fluctuations in molecular levels of the most important cell cycle regulators (notations as in Table 1 and Fig. 2). Plots show that the level of CycD is not fluctuating during the cell cycle, still its activity together with the increasing CycE dependent Cdk activity can induce CKI degradation and activation of further CycE and CycA production (after inactivation of Rb and activation of E2F, as noted on Fig. 2). Cdk/CycA and Cdk/CycE induce Cdh1 inactivation and drive the cell into S-phase. Wee1 keeps Cdk/CycB inactive in G2. Later, as the autocatalytic loops of Cdk/CycB activation turn on, Wee1 gets inactivated and the increase in Cdk/CycB activity induces M-phase. At the end of mitosis, when the spindles are properly aligned, Cdc20 activation is not halted anymore and the active Cdc20/APC induces CycB degradation and CKI and Cdh1 activation to finish the cell cycle. The drop in Cdk/CycB activity induces cell division

Cell Cycle Modeling Techniques

The most common approach to describe a system composed of biochemical reactions is to translate the interactions of a molecular wiring diagram into ordinary differential equations (ODEs) (▶ Cell Cycle Modeling, Differential Equation). These can be either solved analytically or, in the case of larger models, numerically by computational means. The solutions of ODEs can be compared with existing experimental data, thus some of the kinetic rate constants of the model can be defined. ODEs have been used for many years to create mathematical models of eukaryotic cell cycle regulation, since this was the technique that was used to investigate chemical and physical oscillators. Thus, all the developed tools of dynamical systems theory could be used to understand ODE models of cell cycle regulation. Maybe to most frequently used technique is bifurcation analysis (▶ Cell Cycle Model Analysis, Bifurcation Theory), which follows the steady state solutions of the system of differential equations. These steady states can identify



Cell Cycle, Fig. 4 *Bistability in the budding yeast cell cycle*. The signal-response curve (*left*) shows how the steady state Cdk/Cyc activity depends on cell mass (*solid and dashed lines* stand for stable and unstable steady states respectively). In early G1 phase only one steady state exists, with low Cdk/Cyc activity. As the cell grows the second stable state appears as well, but the original G1 steady state still remains until the cell mass reaches a critical value where this disappears and only the high Cdk activity state remains. As the cell grows (*dotted lines*) Cdk gets abruptly activated at the critical cell mass what drives the cell into S-phase. Small fluctuations cannot drive the system back to G1 phase. Unless the cell mass sensing pathway activity drops below the lower boundary of the bistabile zone (*gray background*), the cell will stay in the high Cdk activity S/G2/M state. The bistability is a result of an antagonistic, double negative (thus positive) feedback loop, where Cdk inhibits Cdh1 and CKI activity, while Cdh1 induces cyclin degradation and CKI inhibits Cdk/Cyc complexes (Fig. 1a)

the phases of the cell cycle and cell cycle transitions could be associated with bifurcations. Indeed Fig. 4 is a bifurcation diagram where cell mass is a bifurcation parameter and Cdk activity is the state variable.

The major problem with most biological systems is the fact that very few rate constants can be measured directly. Parameter sensitivity analysis (▶ Cell Cycle Models, Sensitivity Analysis) is used to identify the parameters that have the highest influence on model behavior. This method could help to reduce the number of parameters that need to be measured or estimated. Several techniques exist to estimate the parameters of a particular model, from manual parameter "twiddling" based on "educated guesses" to highly sophisticated parameter inference using software tools (Panning et al. 2008; Schmidt and Jirstrand 2006). Almost all of these have been used in various cell cycle models. Metabolic control analysis is mostly used to identify key reactions of metabolism, but recent studies showed that similar technique could be used to identify key reactions of the cell cycle as well (▶ Cell Cycle Signaling, Metabolic Pathway).

Our knowledge about the regulatory network of the cell cycle and its interactions with other cellular pathways is constantly growing. As a consequence, the complexity of mathematical models is growing as well. Detailed analysis and parameterization of such large systems can be difficult when the number of ODEs is large. Simplification by logical modeling has been proposed to overcome this problem (▶ Cell Cycle Modeling Using Logical Rules). Logical modeling has a long tradition in biology and recently some applications to cell cycle research also appeared. These models are based on Boolean algebra, where the activity of each component is represented by two states: ON and OFF, providing a method which is computationally less expensive. Behaviors that originate from the topology of the system can be investigated this way. In their pioneering study, Li et al. (2004) showed that the network of the budding yeast cell cycle regulation is robustly designed, as 86% of all possible initial states lead the system to the same stable cell cycle path. This result underlines that the biochemical network of cell cycle regulation can function in a parameter-insensitive way.

Fluctuations in the average behavior of a cell population can be described by deterministic ODE models or logical models, but the behavior of individual cells cannot be simulated that way. Stochastic modeling approaches (▶ Cell Cycle Modeling, Stochastic Methods) are

becoming more and more popular as they provide a means to analyze single cells and find relevant results in cooperation with novel single-cell experimental techniques such as quantitative flow cytometry and fluorescence microscopy. Some cell cycle models are already using stochastic simulation techniques, such as Langevin-type stochastic ODEs or the exact stochastic simulation algorithm (Gillespie 2007). The effect of noise in nonsymmetrical cell division on cell size distribution on the population level has been also investigated. Much more detailed stochastic models of cell cycle regulation are appearing. Such models take into account measured molecular levels, protein, and mRNA half-lives and several other noisy processes. Most of these models are based on earlier ODE models that are "unpacked" into elementary reactions to be able to run stochastic simulations on them. A successful approach along these lines uses Petri nets (▶ Cell Cycle Modeling, Petri Nets) which allow to move between deterministic and stochastic simulations and to follow molecule numbers instead of concentrations. Some other modeling concepts from computer science are adopted and adapted to model biological systems. Rule-based modeling and especially various process algebras (▶ Cell Cycle Modeling, Process Algebra) are able to handle combinatorial complexity caused by complex formation and various protein modifications. These allow an easier handling of multisite phosphorylation and multi-component complex formations, which are both relevant issues for cell cycle regulation.

System Level Cell Cycle Experiments

Many models are parameterized by fitting data on physiological observations: cell sizes, cell cycle phase distributions, and viabilities of mutants – after painstaking literature mining for these data. The advanced large-scale measurements and the related databases provide modelers data to help them formulating larger, more detailed, and more precise models in the near future.

Results collected over the last 40 years, including recent system-level experiments have extended our knowledge and now allow us to construct much more detailed interaction maps of key cell cycle regulatory steps (▶ Cell Cycle Transition, Detailed Regulation of Restriction Point). Classical flow cytometry was first used in 70s to determine DNA content and with that, the cell cycle stage of cell populations. Now, with the help of fluorescence proteins, it could be used in a much more sophisticated way to determine molecule number changes during the cell cycle (▶ Cell Cycle Analysis, Flow Cytometry). Fluorescence proteins and advanced microscopy techniques also allow analysis of single cells and even localization of single molecules during the cell cycle (▶ Cell Cycle Analysis, Live-Cell Imaging). The periodic oscillations in transcription during the cell cycle was an early target of transcriptomics research (▶ Cell Cycle Analysis, Expression Profiling) indeed, one of the first genome-scale microarray experiments was performed on the budding yeast cell cycle. Protein level fluctuations are classically followed by Western blots, but now large-scale mass spectrometry analysis is widely used to follow protein level fluctuations during the cell cycle and to indentify the phosphorylation states of Cdk-regulated proteins (Aebersold and Mann 2003). Advanced ChIP–chip techniques help to reveal the transcriptional network of cell cycle regulation (Bahler 2009) and, essentially, all molecular and systems biology methods are used in cell cycle research. The collected data is stored in various databases on protein interactions, transcriptional interactions, post-translational modifications, etc (▶ Cell Cycle Data Analysis). There are some cell cycle specific databases that contain information on periodically transcribed genes during the cell cycle (Gauthier et al. 2010) and that introduce related mathematical models as well (Alfieri et al. 2008). Furthermore, literature mining tools are helping us in searching for specific experimental results on our favorite proteins. Large-scale gene perturbation experiments are also being used to investigate the cell cycle: studies on single- or double gene-deletion strains (Hillenmeyer et al. 2008) and phenotype analysis of protein overexpressing cells are all giving important results to cell cycle research (▶ Cell Cycle Analysis, Systematic Gene Overexpression).

All these resources are being used to support the development of computational models of cell cycle regulation. The phenotypes of the deletion and overexpression mutant strains provide physiological constrains for such models, while time-course measurements of mRNA and protein levels provide data to be fitted by the simulation of the models.

Models Beyond the Cell Cycle

Cell cycle of individual cells in multicellular organisms (such as humans) is controlled by the environment as well as it is influenced by its neighboring cells. Cell–cell interactions make the modeling of cell cycle in a cellular tissue a multilevel problem (▶ Multilevel Modeling, Cell Proliferation). Not only the internal molecular interactions of each individual cell, but their incoming and outgoing signals to their neighbors need to be tracked. Cell cycle modules could be introduced into agent-based models to simulate how cell cycle is controlled in individual interacting cells. Furthermore, the core cell cycle engine should properly regulate important downstream processes, such as DNA replication and cell division. To carry out a proper function, it is interlocked with several other pathways. Various environmental effects (stress, light, temperature, etc.) can influence progression through cell division cycles, and several drugs, external signaling molecules, and metabolites could also affect cell proliferation (Csikasz-Nagy 2009). Many of these upstream and downstream pathways of cell cycle regulation have been investigated in great detail by systems biology techniques. The yeast osmotic pressure sensing pathway is one of the most investigated pathways; several experiments and computational models are dealing with it (Klipp et al. 2005). Computational models of the mammalian system are focusing on the regulation of the cell cycle by the low oxygen induced hypoxia pathway, on the Src and NFκB pathways, as well as on the interactions between DNA damage, apoptosis, and the cell cycle. A recent discovery on the connection between the circadian clock and the cell cycle in mammalian cells (▶ Cell Cycle, Coupled with Circadian Clock) also initiated some computational modeling work along these lines. Likewise, the role of micro RNAs in relation to cell cycle regulation are also being investigated by computational modeling (▶ Cell Cycle Regulation, microRNAs). Also, some of the downstream processes induced by the cell cycle machinery have been studied by mathematical modeling. The regulation of cell growth, cell division, DNA replication and several aspects of mitosis are all described in some detail in the literature (Lygeros et al. 2008; Mogilner et al. 2006), but none of these are connected to a model of the core cell cycle regulatory network. Thus, we still need to wait to see the results describing the crosstalk between the various upstream and downstream pathways and the core cell cycle machinery.

Summary

A full systems biology workflow of *experiment* → *database* → *model* → *experiment* was applied to understand the role of positive feedback in the cell cycle regulation. To realize the entire loop took more than 10 years in the case of the frog egg studies, but it took only 4 years in the case of the budding yeast cell cycle control. We are hoping that with the emergence of systems biology, stimulating interactions between experimentalists and theoreticians could lead to more detailed, more precise and realistic models. For instance, we will need more sophisticated experimental techniques to measure protein activity fluctuations in time – desirably in individual cells. At the same time, better modeling and parameter estimation software is needed to deal with the increasing amounts of data and to accommodate new types of measurements on individual cells. Even though we have learned a great deal about cell cycle regulation in the last 40–50 years, there are still gaps in our knowledge that need bridged, and experimentalists and theoreticians need to work together to accomplish this.

Cross-References

- ▶ CDK Inhibitors
- ▶ Cell Cycle Analysis, Expression Profiling
- ▶ Cell Cycle Analysis, Flow Cytometry
- ▶ Cell Cycle Analysis, Live-Cell Imaging
- ▶ Cell Cycle Analysis, Systematic Gene Overexpression
- ▶ Cell Cycle Arrest After DNA Damage
- ▶ Cell Cycle Checkpoints
- ▶ Cell Cycle Data Analysis
- ▶ Cell Cycle Dynamics, Bistability and Oscillations
- ▶ Cell Cycle Dynamics, Irreversibility
- ▶ Cell Cycle Model Analysis, Bifurcation Theory
- ▶ Cell Cycle Modeling Using Logical Rules
- ▶ Cell Cycle Modeling, Differential Equation
- ▶ Cell Cycle Modeling, Petri Nets
- ▶ Cell Cycle Modeling, Process Algebra

- Cell Cycle Modeling, Stochastic Methods
- Cell Cycle Models, Sensitivity Analysis
- Cell Cycle of Early Frog Embryos
- Cell Cycle of Mammalian Cells
- Cell Cycle Regulation, microRNAs
- Cell Cycle Signaling, Hypoxia
- Cell Cycle Signaling, Metabolic Pathway
- Cell Cycle Signaling, Spindle Assembly Checkpoint
- Cell Cycle Transition, Detailed Regulation of Restriction Point
- Cell Cycle Transition, Principles of Restriction Point
- Cell Cycle Transitions, G2/M
- Cell Cycle Transitions, Mitotic Exit
- Cell Cycle, Budding Yeast
- Cell Cycle, Cancer Cell Cycle and Oncogene Addiction
- Cell Cycle, Cell Size Regulation
- Cell Cycle, Coupled With Circadian Clock
- Cell Cycle, Fission Yeast
- Cell Cycle, Physiology
- Cell Cycle, Prokaryotes
- Cyclins and Cyclin-Dependent Kinases
- Cytokinesis
- DNA Replication
- Endoreplication
- Lymphocyte Dynamics and Repertoires, Biological Methods
- Meiosis
- Mitosis
- Modeling Approaches of Cell Cycle Regulation by Signaling Pathways for External Stress
- Multilevel Modeling, Cell Proliferation

References

Aebersold R, Mann M (2003) Mass spectrometry-based proteomics. Nature 422:198–207

Alfieri R, Merelli I, Mosca E, Milanesi L (2008) The cell cycle DB: a systems biology approach to cell cycle analysis. Nucleic Acids Res 36:D641–D645

Bahler J (2009) Global approaches to study gene regulation. Methods 48:217

Bloom J, Cross FR (2007) Multiple levels of cyclin specificity in cell-cycle control. Nat Rev Mol Cell Biol 8:149–160

Chen KC, Csikasz-Nagy A, Gyorffy B, Val J, Novak B, Tyson JJ (2000) Kinetic analysis of a molecular model of the budding yeast cell cycle. Mol Biol Cell 11:369–391

Chen KC, Calzone L, Csikasz-Nagy A, Cross FR, Novak B, Tyson JJ (2004) Integrative analysis of cell cycle control in budding yeast. Mol Biol Cell 15:3841–3862

Ciliberto A, Shah JV (2009) A quantitative systems view of the spindle assembly checkpoint. EMBO J 28:2162–2173

Cross FR, Archambault V, Miller M, Klovstad M (2002) Testing a mathematical model for the yeast cell cycle. Mol Biol Cell 13:52–70

Csikasz-Nagy A (2009) Computational systems biology of the cell cycle. Brief Bioinform 10:424–434

Csikasz-Nagy A, Battogtokh D, Chen KC, Novak B, Tyson JJ (2006) Analysis of a generic model of eukaryotic cell-cycle regulation. Biophys J 90:4361–4379

Evans T, Rosenthal ET, Youngblom J, Distel D, Hunt T (1983) Cyclin: a protein specified by maternal mRNA in sea urchin eggs that is destroyed at each cleavage division. Cell 33:389–396

Ferrell JE (2002) Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability. Curr Opin Cell Biol 14:140–148

Gauthier NP, Jensen LJ, Wernersson R, Brunak S, Jensen TS (2010) Cyclebase.org: version 2.0, an updated comprehensive, multi-species repository of cell cycle experiments and derived analysis results. Nucleic Acids Res 38:D699–D702

Gillespie DT (2007) Stochastic simulation of chemical kinetics. Annu Rev Phys Chem 58:35–55

Goldbeter A (1991) A minimal cascade model for the mitotic oscillator involving cyclin and cdc2 kinase. Proc Natl Acad Sci USA 88:9107–9111

Hartwell LH, Weinert TA (1989) Checkpoints: controls that ensure the order of cell cycle events. Science 246:629–634

Hartwell LH, Mortimer RK, Culotti J, Culotti M (1973) Genetic control of the cell division cycle in yeast: V. Genetic analysis of cdc mutants. Genetics 74:267–286

Hillenmeyer ME, Fung E, Wildenhain J, Pierce SE, Hoon S, Lee W, Proctor M, St Onge RP, Tyers M, Koller D, Altman RB, Davis RW, Nislow C, Giaever G (2008) The chemical genomic portrait of yeast: uncovering a phenotype for all genes. Science 320:362–365

Klipp E, Nordlander B, Kruger R, Gennemark P, Hohmann S (2005) Integrative model of the response of yeast to osmotic shock. Nat Biotechnol 23:975–982

Koch AL, Schaechter M (1962) A model for statistics of the cell division process. J Gen Microbiol 29:435–454

Li F, Long T, Lu Y, Ouyang Q, Tang C (2004) The yeast cell-cycle network is robustly designed. Proc Natl Acad Sci USA 101:4781–4786

Lindqvist A, Rodriguez-Bravo V, Medema RH (2009) The decision to enter mitosis: feedback and redundancy in the mitotic entry network. J Cell Biol 185:193–202

Lukas J, Lukas C, Bartek J (2004) Mammalian cell cycle checkpoints: signalling pathways and their organization in space and time. DNA Repair (Amst) 3:997–1007

Lygeros J, Koutroumpas K, Dimopoulos S, Legouras I, Kouretas P, Heichinger C, Nurse P, Lygerou Z (2008) Stochastic hybrid modeling of DNA replication across a complete genome. Proc Natl Acad Sci USA 105:12295–12300

Masui Y, Markert CL (1971) Cytoplasmic control of nuclear behavior during meiotic maturation of frog oocytes. J Exp Zool 177:129–145

Mogilner A, Wollman R, Civelekoglu-Scholey G, Scholey J (2006) Modeling mitosis. Trends Cell Biol 16:88–96

Morgan DO (2006) The cell cycle: principles of control. New Science Press, London
Nasmyth K (2001) A prize for proliferation. Cell 107:689–701
Novak B, Tyson JJ (1993) Numerical analysis of a comprehensive model of M-phase control in *Xenopus* oocyte extracts and intact embryos. J Cell Sci 106(Pt 4):1153–1168
Novak B, Tyson JJ (2008) Design principles of biochemical oscillators. Nat Rev Mol Cell Biol 9:981–991
Novak B, Tyson JJ, Gyorffy B, Csikasz-Nagy A (2007) Irreversible cell-cycle transitions are due to systems-level feedback. Nat Cell Biol 9:724–728
Nurse P (1975) Genetic control of cell size at cell division in yeast. Nature 256:547–551
Nurse P (1990) Universal control mechanism regulating onset of M-phase. Nature 344:503–508
Panning T, Watson L, Allen N, Chen K, Shaffer C, Tyson J (2008) Deterministic parallel global parameter estimation for a model of the budding yeast cell cycle. J Global Optim 40:719–738
Pomerening JR, Sontag ED, Ferrell JE Jr (2003) Building a cell cycle oscillator: hysteresis and bistability in the activation of Cdc2. Nat Cell Biol 5:346–351
Pomerening JR, Kim SY, Ferrell JE Jr (2005) Systems-level dissection of the cell-cycle oscillator: bypassing positive feedback produces damped oscillations. Cell 122:565–578
Queralt E, Lehane C, Novak B, Uhlmann F (2006) Downregulation of PP2A(Cdc55) phosphatase by separase initiates mitotic exit in budding yeast. Cell 125:719–732
Schmidt H, Jirstrand M (2006) Systems biology toolbox for MATLAB: a computational platform for research in systems biology. Bioinformatics 22:514–515
Sha W, Moore J, Chen K, Lassaletta AD, Yi CS, Tyson JJ, Sible JC (2003) Hysteresis drives cell-cycle transitions in *Xenopus laevis* egg extracts. Proc Natl Acad Sci USA 100:975–980
Sia RA, Bardes ES, Lew DJ (1998) Control of Swe1p degradation by the morphogenesis checkpoint. EMBO J 17:6678–6688
Zachariae W, Nasmyth K (1999) Whose end is destruction: cell division and the anaphase-promoting complex. Genes Dev 13:2039–2058

Cell Cycle Analysis, Expression Profiling

Jürg Bähler and Samuel Marguerat
Department of Genetics, Evolution and Environment and UCL Cancer Institute, University College London, London, UK

Synonyms

Expression profiling; Transcriptome profiling; Transcriptomics

Definition

Gene expression (or ▶ Transcriptome) profiling refers to the simultaneous measurement of the activities of most or all genes present in a cell, a tissue, or a whole organism. The activity of a given gene is determined by the number of RNA transcripts (expression level) derived from this gene under a specific condition. Using cells that are synchronized for the cell cycle (▶ Cell Cycle, Synchronization), gene expression profiling enables the large-scale identification of genes that are periodically expressed as a function of the cell cycle (▶ Cell Cycle-Regulated Gene Expression). This approach thus provides global data of gene regulation during the cell cycle, which supports a systems-level understanding of cell-cycle control.

Characteristics

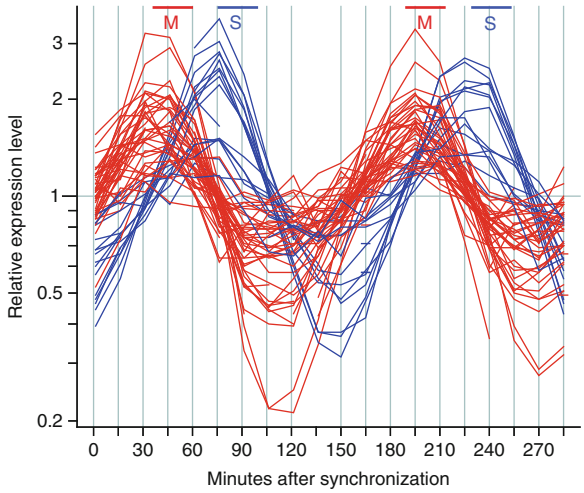
Methods for Gene Expression Profiling

Gene expression profiling has been enabled by the development of large-scale technologies such as ▶ DNA microarrays and, more recently, next-generation sequencing-based approaches such as ▶ RNA-seq. These genome-wide approaches are quite straightforward experimentally, and the biggest challenge is often to extract biologically meaningful information from the huge data sets generated. Computational data mining is therefore a central aspect of gene expression profiling, including a range of specialized approaches such as ▶ clustering.

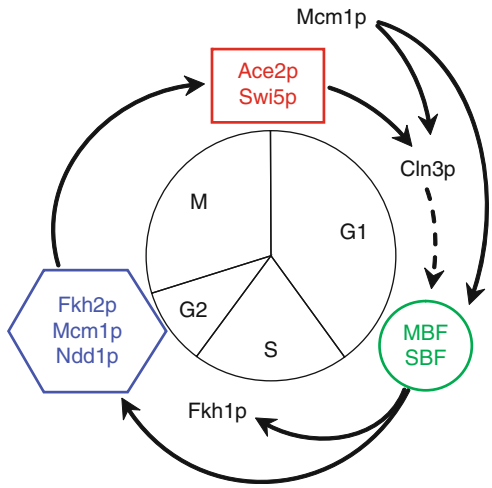
Besides methods for gene expression profiling to measure ▶ transcriptome levels during the cell cycle, complementary large-scale technologies are available for global insight into ▶ cell-cycle-regulated gene expression. For example, ▶ chromatin immunoprecipitation (ChIP) combined with ▶ DNA microarrays (ChIP-chip) or sequencing (ChIP-seq) provide a means to determine the genome-wide binding sites of transcription factors controlling gene expression. Moreover, recent developments in ▶ proteomics are now allowing the global measurement of protein levels and post-translational modifications as a function of the cell cycle.

Gene Expression Profiling During the Cell Cycle

Periodic control of transcription seems to be a universal feature of the cell-division cycle.



Cell Cycle Analysis, Expression Profiling, Fig. 1 Expression profiles of fission yeast transcripts peaking in levels before M-phase (*red*) or S-phase (*blue*). Transcriptomes were analyzed by microarrays every 15 min after cell-cycle synchronization for two full cell cycles. Data from Rustici et al. (2004)



Cell Cycle Analysis, Expression Profiling, Fig. 2 Serial regulation of transcription factors during the budding yeast cell cycle. Each group of transcription factors (specified in colored frames) regulates transcription factors acting in the next cell-cycle phase (shown in circle). Continuous and hatched lines indicate transcriptional and post-translational regulation, respectively

Gene expression profiling of proliferating cells after ► cell-cycle synchronization have identified hundreds of periodically expressed genes, in organisms ranging from bacteria to humans (Breeden 2003). Periodically expressed genes often play specific roles in the cell cycle, and their expression levels typically peak just before the phase during which they function. Particularly rich genome-wide data are available for budding and fission yeasts, which provide complementary models for cell-cycle-regulated ► gene expression (Bähler 2005). About 10–20% of all yeast genes are periodically expressed during the cell cycle, many of them in three transcriptional waves that roughly coincide with major cell-cycle transitions: initiation of DNA replication, entry into mitosis, and exit from mitosis (Fig. 1).

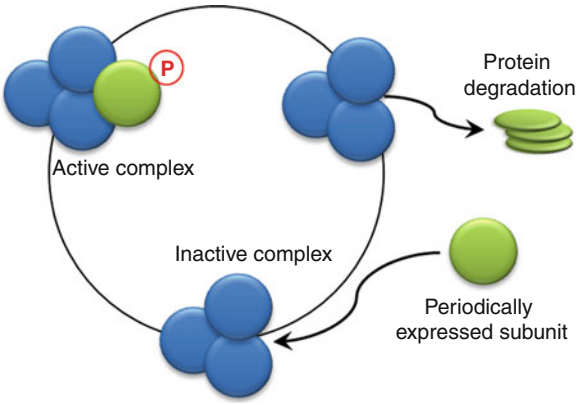
Control of Cell-Cycle-Regulated Gene Expression

Periodic gene expression is integrated with cell-cycle progression by multiple feedback loops at both transcriptional and post-translational levels (Bähler 2005; Futcher 2002). Some cell-cycle transcription factors are functionally conserved from yeast to human. For example, proteins of the forkhead family, such as Fkh2 in yeast and FoxM in human, control genes whose transcript levels peak before M-phase. Other transcriptional regulators, such as MBF/SBF in yeast and E2F in human, are not conserved at the sequence level but play analogous roles in controlling genes whose transcript levels peak before S-phase.

Cell-cycle-regulated gene expression often shows serial regulation, whereby transcription factors functioning during one cell-cycle phase up-regulate transcripts that encode downstream transcription factors functioning during the next phase (Fig. 2) (Futcher 2002). Such a network of sequentially expressed transcription factors may act as a cell-cycle oscillator that regulates the periodic gene expression program independently of, and in addition to, cyclin-dependent kinases (Simmons Kovacs et al. 2008).

Budding and fission yeasts show similarities as well as intriguing differences in their ► transcriptional regulatory networks, revealing evolutionary plasticity of gene expression control (Bähler 2005): Regulatory circuits between conserved transcription factors have been rewired, and periodic transcription of many orthologous genes is not conserved, except for a core set of genes that may be universally regulated during the eukaryotic cell cycle and may play key roles in cell-cycle progression.

Although protein complexes involved in the cell cycle are largely conserved among eukaryotes, their



Cell Cycle Analysis, Expression Profiling, Fig. 3 Just-in-time assembly of protein complex during the cell cycle, controlled by a single cell-cycle-regulated subunit (*green*). Post-translational regulation of the periodically expressed subunit, such as phosphorylation (*red*) and protein degradation, often serve as extra layers of control

regulation has changed considerably during evolution (de Lichtenberg et al. 2007). Most protein complexes contain both periodically and constitutively expressed subunits, but the periodically expressed subunits differ between organisms. Moreover, the periodically expressed subunits, which may control complex activity by just-in-time assembly, tend to be coordinately controlled at transcriptional as well as post-translational levels, highlighting that multiple regulatory layers are often integrated to drive cell-cycle progression (Fig. 3).

Metabolic Cycle

An intriguing special case for periodic gene expression is the metabolic cycle. When continuously grown under nutrient-limiting conditions, budding yeast exhibit highly periodic cycles of respiratory bursts, and microarray studies revealed that over 50% of the genes are expressed periodically during these metabolic cycles. The cell-division cycle is gated by this metabolic cycle, with DNA replication being segregated away from the oxidative phase when cells are actively respiring. This temporal compartmentalization ensures that DNA is replicated when oxidative stress is lowest, in order to minimize genome damage. The coordination between metabolic and cell-division cycles may be related to the coupling between circadian and cell-division cycles in other organisms (Chen and McKnight 2007).

Cross-References

- Cell Cycle-Regulated Gene Expression
- Cell Cycle, Synchronization
- Chromatin Immunoprecipitation
- Clustering
- DNA Microarrays
- Proteomics
- RNA-Seq
- Transcriptional Regulatory Network
- Transcriptome

References

Bähler J (2005) Cell-cycle control of gene expression in budding and fission yeast. Annu Rev Genet 39:69–94

Breeden LL (2003) Periodic transcription: a cycle within a cycle. Curr Biol 13:R31–R38

Chen Z, McKnight SL (2007) A conserved DNA damage response pathway responsible for coupling the cell division cycle to the circadian and metabolic cycles. Cell Cycle 6:2906–2912

de Lichtenberg U, Jensen TS, Brunak S, Bork P, Jensen LJ (2007) Evolution of cell cycle control: same molecular machines, different regulation. Cell Cycle 6:1819–1825

Futcher B (2002) Transcriptional regulatory networks and the yeast cell cycle. Curr Opin Cell Biol 14:676–683

Rustici G, Mata J, Kivinen K, Lió P, Penkett CJ, Burns G, Hayles J, Brazma A, Nurse P, Bähler J (2004) Periodic gene expression program of the fission yeast cell cycle. Nat Genet 36:809–817

Simmons Kovacs LA, Orlando DA, Haase SB (2008) Transcription networks and cyclin/CDKs: the yin and yang of cell cycle oscillators. Cell Cycle 7:2626–2629

Cell Cycle Analysis, Flow Cytometry

James W. Jacobberger
Case Comprehensive Cancer Center, Case Western Reserve University, Cleveland, OH, USA

Synonyms

Cell cycle analysis, Phase fraction analysis; Cytometry, cytofluorimetry, microfluorimetry, analytical cytology; Flow cytometry, flow microfluorimetry

Definition

► Cell cycle analysis commonly denotes a group of analytical processes used to estimate the fraction of cells that are resident in cell cycle phases, states, or stages by cell-based measurement methods. Generally, the measurements are performed with optical instruments connected to photosensitive components (photomultiplier tubes, photo diodes, and charge coupled devices [CCD]) and associated electronics to digitize light that is emitted (fluorescence, phosphorescence) or scattered (not absorbed) by individual cells. The digitization process converts a uniform pulse of light to a usable number that represents the quantity of some biochemical entity (usually, DNA), "reported" by the fluorescent label.

Characteristics

Instrumentation

► Flow cytometers generally employ optical elements (lenses, fiber optics, filters, and mirrors) to direct focused beams of monochromatic or a restricted band of light from either lasers or arc-lamps to a spot within a flowing stream that contains cells that move, single file, under forces of laminar flow. Cells, in suspension, are introduced to flow cytometers from tubes and plate wells. Laser scanning cytometers employ moving stages, laser light sources, and modulated mirrors to scan cells that are stationary with respect to an optically transparent substrate that is scanned by a moving beam spot in one direction while the microscope stage moves in the orthogonal direction. The single existing commercial imaging flow cytometer uses the same principles as ordinary flow cytometers but uses a CCD to capture several images of each cell moving through the interrogation region and integrates data on a per pixel basis. A variant of flow cytometry, "cell sorting," employs a flow cytometer with inline mechanisms for undulating and charging the fluid stream to create charged droplets containing one or more cells as the stream exits a nozzle. Charged plates then deflect the droplets in one of 2–4 directions based on computation that occurred upstream at the interrogation point. This latter technology, while useful in cell cycle research, is not required for cell cycle analysis. Some major sources of commercial instrumentation are listed in Table 1.

Key Features of Cytometry

There are two principle features of cytometry that distinguishes it from other approaches to cell biology. The first is that the data are a set of measurements with each element a measurement on a single cell. The value of that distinction is that for any given asynchronous population, under specific conditions, cells can be observed at some frequency at all possible states for the system. For example, proliferating cells will be measured at all cell cycle phases or states related to the specific cell type and environment. An element that is related to this is that in multivariate cytometry, all measurements are correlated. This means that if measurement *a* represents state 1 in the system, that state is also characterized with respect to measurement *b*. The second feature of cytometry is that it is quantitative. In the design and manufacture of the instruments, emphasis is on obtaining quantitative measurement with high precision – i.e., if two particles, one containing 1*x* and one containing 2*x* of fluorescent dye molecules, emphasis is placed on minimizing error when 1*x* + error and 2*x* + 2error represents reality. In cell cycle analysis, it is not uncommon to measure cells with one genome (G0/G1) and cells with two genomes (G2/M) at a G2/G1 ratio of 1.95 or better and coefficients of variation (CV) of 1–4%.

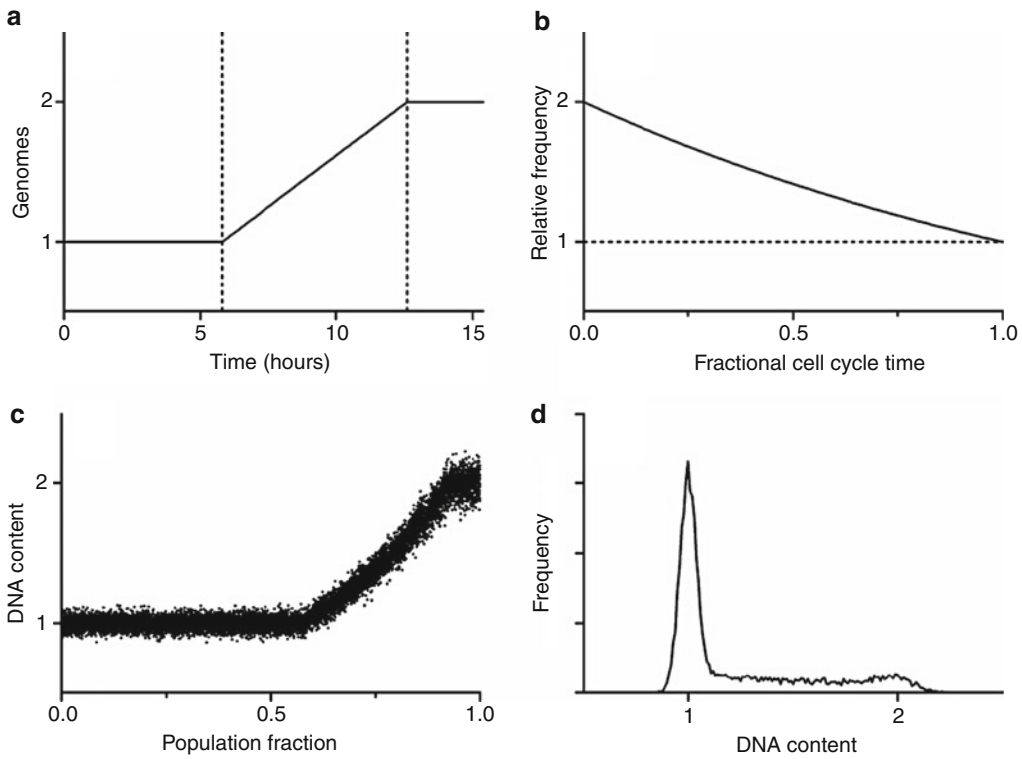
Basis for Cell Cycle Analysis

Figure 1a shows a simulated expression profile for DNA or genome content for a cycling mammalian somatic cell with a cell cycle time (Tc) of 15.4 h. The time spent with one genome is the length of the G1 phase of the cell cycle – in this case, 5.6 h. The time spent synthesizing a second genome is the length of S phase (6.8 h), and the length of time with two genomes is 2.8 h (G2 + M). S phase is defined by genomic DNA synthesis. Figure 1b shows a simulated age distribution for a proliferating population on a fractional cell cycle time (Tc) basis. This distribution accounts for binary division at the end of the cycle. If the expression profile (A) is transformed to account for overrepresentation in the population due to binary division; and normally distributed random numbers

Cell Cycle Analysis, Flow Cytometry, Table 1 Commercial research cytometers[a]

Instrumentation	Company	City, state/country	Web address
Flow cytometers/Cell sorters	BD biosciences	San Jose, CA	bdbiosciences.com
	Beckman coulter	Miami, FL	coulterflow.com
	i-Cyt/Sony	Champaign, IL	i-cyt.com
Flow cytometers	Partec company	Münster, Germany	partec.com
	Millipore	Billerica, MA	millipore.com
	Invitrogen	Carlsbad, CA	invitrogen.com
	Accuri	Ann Arbor, MI	accuricytometers.com
Laser scanning cytometers	Compucyte	Westwood, MA	compucyte.com
Imaging flow cytometers	Amnis	Seattle, WA	amnis.com

[a]This list is not comprehensive



Cell Cycle Analysis, Flow Cytometry, Fig. 1 Relationship between cell cycle-related DNA expression and cytometry of DNA content. **Panel (a)** represents DNA expression over the life of a cell with the values G1 = 5.8 h, S = 6.8 h, and G2 + M=2.8 h, which are values that we have previously measured for NIH3T3 cells growing in 10% serum at a density of 87–121 cells/cm^2 (Disalvo et al. 1995). **Panel (b)** shows the age distribution of a population of asynchronously growing cells, accounting for the generation of two cells at birth, plotted as a function of the fractional cell cycle time. See (Walker 1954) for the logic. **Panel (c)** shows simulated DNA content measurements that would be measured for cells programmed to express DNA as in A. To generate these data, the X axis of A is transformed as Xt = (2*exp(-0.6931*Xi/Tc))*(X/Tc), Tc = cell cycle time (15.4), Xt = transformed X, Xi = original X. Xts are unitless fractions of Tc. **Panel (d)** is a histogram of the data in Panel (c)

are generated along the adjusted "expression" curve to simulate cytometric measurements, we get Fig. 1c. If we then create a histogram from the simulated data (Fig. 1d), one obtains a simulated DNA content distribution for a proliferating cell population with an average cell cycle time of 15.6 h. This model does not account for any phase specific cell death. For further reading, see Walker (1954), Mitchison (1971), Watson (1991), Bagwell (1993).

Extension of the Logic

The illustrated logic of Fig. 1 was first enunciated in 1954 (Walker 1954) and later developed within the context of data from flow cytometry by several groups. The logic can be extended to multiple parameters. Figure 2 shows simulated plots for DNA and cyclin A2 (Fig. 2f) compared to real data (Fig. 1a) from which the fraction profiles (Fig. 2b, c) were derived and noise added (Fig. 2d, e). By reversing the logic shown in Fig. 1, the programmed cell cycle expression (Fig. 2g, h) can be derived for any parameter, provided a set of continuous measurements can be made covering the cell cycle in a nonredundant manner. It is relatively easy to see that the logic can be extended to an unlimited number of parameters and that from primary cytometric data, expression profiles for any measurable biochemical feature can be derived. This has distinct implications for mathematical models of cell cycle regulation (e.g., Csikasz-Nagy et al. 2006), since the expression profiles are not different in a relative sense from model outputs of state variable levels over time.

DNA Content Analysis for Simple Proliferating Populations

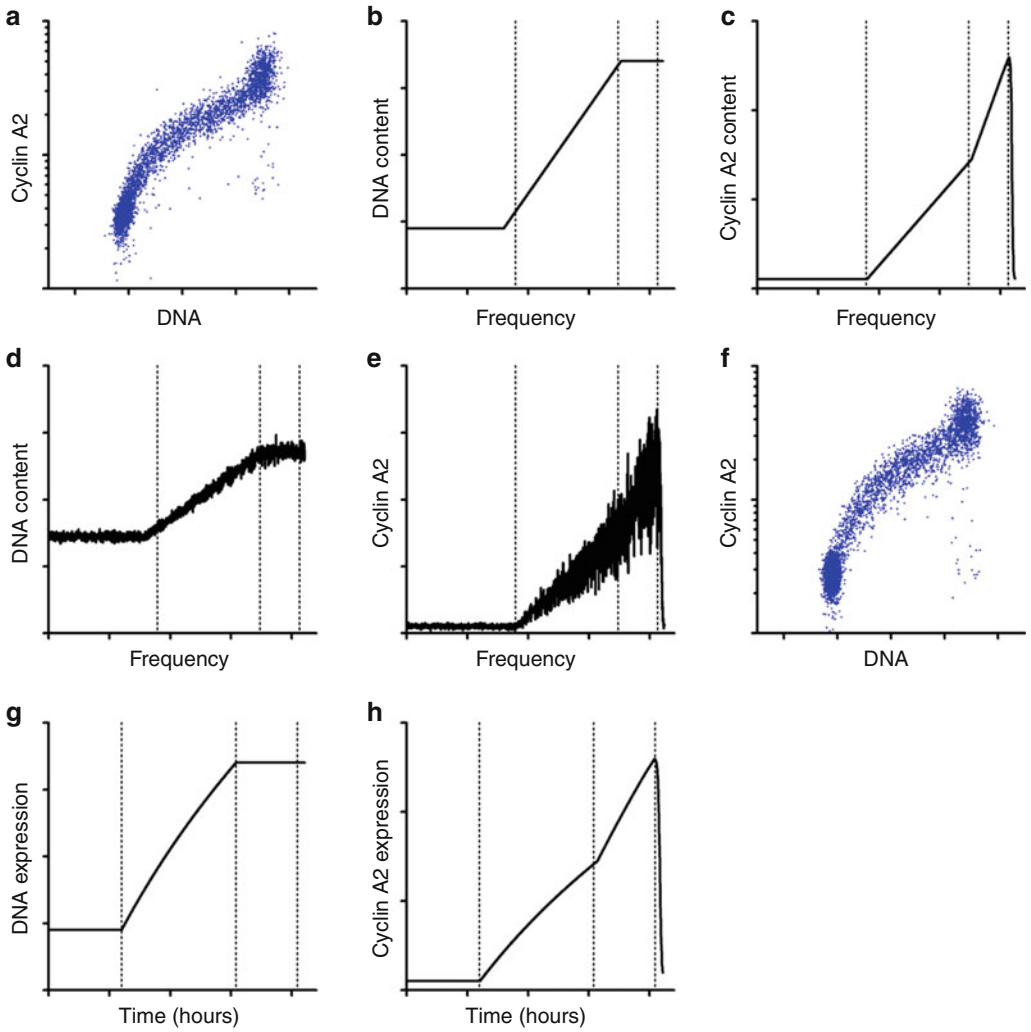
Practical use of DNA content measurements is to calculate the fractions of G0/G1, S, and G2/M cells. Because the error in the measurements are normally distributed (given that cell fixation and staining, and instrument operation are within standard practice), a typical analysis is performed by fitting a complex function composed of summed normal distributions centered on the primary and secondary peaks (G0/G1 and G2/M) and an S phase component that is either (1) a series of Gaussian functions, (2) a series of trapezoids with Gaussian "broadened" ends, or (3) a polynomial with "broadened" ends. Generally, the terms (mean and standard deviation) of the Gaussian functions are defined by the G0/G1 peak and all subsequent Gaussian standard deviation terms are calculated from that using the coefficient of variation (standard deviation/mean) multiplied by the mean of the subsequent Gaussian. The same logic is used to "broaden" the ends of either a trapezoid series or a single polynomial function. Broadening either starts at the center of each peak or some fractional value toward the S phase side of each peak (see Fig. 3b). The purpose is to account for statistical overlap of each component (G0/G1, S, and G2/M). Although the Gaussian series is the most appealing from a statistical principles point of view, in practice the approach is the least robust and is generally used for perturbed populations. The polynomial model works well for asynchronous, unperturbed populations, since it robustly accounts for the theoretical, expected shape of S phase. The trapezoid model is a good compromise between the other two, able to robustly fit S phases of almost any shape. Most commercial software contains some form of DNA content distribution fitting. Figure 3 shows a typical analysis using a polynomial fit to S phase.

Limitations

Cell populations like liver parenchyme cells in vivo, most mammalian cell cultures, and mammalian tumor cells often undergo endoreduplication (failure to undergo mitosis and cytokinesis during a mitotic cycle). Tumor cells and mammalian cell cultures also fail to undergo binary division (cytokinesis) and thus create cycling cell populations with two or more nuclei that cycle together. Additionally, as tumor cells and tumorigenically transformed mammalian cell cultures evolve, multiple stable stem lines with fractional genomes (aneuploidy) are present in the populations. Most of these features can be dealt with using single parameter DNA content analysis with sound logic, a theoretical underpinning, and robustness obtained by years of practice and widespread use. At some point, these analyses become tortured. ModFit LT (Verity House, Topsham, ME) and MultiCycle AV (Phoenix Flow Systems, San Diego, CA) are two software packages that support complex models.

Further Reading

There are many good reviews and chapters describing DNA content analysis (Dean 1987; Gray et al. 1990; Watson 1991; Bagwell 1993; Rabinovitch 1993).



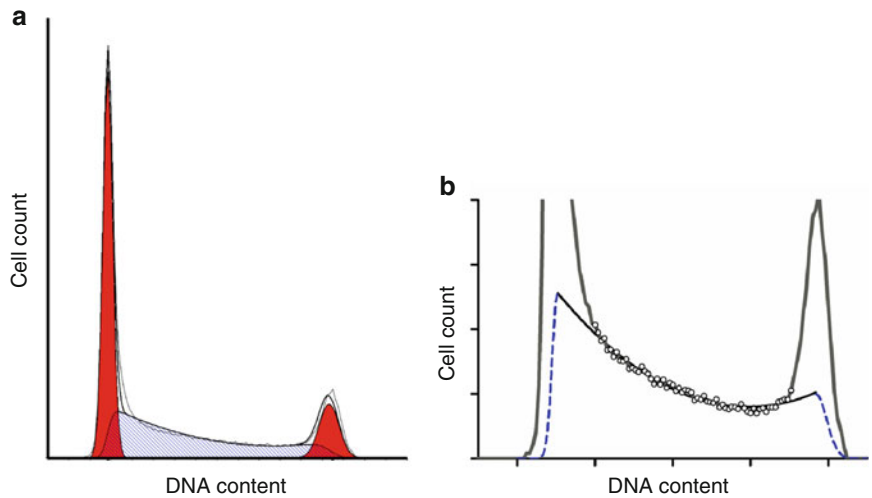
Cell Cycle Analysis, Flow Cytometry, Fig. 2 Relationship between cell cycle-related expression and cytometry of two parameters. **Panel (a)** shows data from exponentially growing RKO cells that were stained for cyclins A2, B1, DNA content, and phospho-S10-histone H3 (Yan et al. 2004). The cell population frequency profiles for DNA content and cyclin A2 calculated from the data shown in Panel (a) are shown in **Panels (b)** and **(c)**. Calculation of these profiles was performed generally as described in Frisa and Jacobberger (2009). A lengthy, more detailed description is described in Jacobberger et al. (2011b). **Panels (g)** and **(h)** are the calculated cell cycle–related expression of DNA and cyclin A2, calculated by reversing the logic in Fig. 1. **Panel (f)** is a bivariate plot of the data in **Panels (d)** and **(e)**. Note the similarity of the simulated data (**f**) and the real data (**a**)

Two Parameter Cell Cycle Analysis

Analysis of the cell cycle in more than one parameter is almost always aimed at increasing the number of cell cycle compartments that can be monitored. The following entries are a sampling of the major approaches for extending cell cycle analysis by adding another parameter to DNA content.

Double-Stranded and Denatured DNA

There are many nucleic acid-binding dyes that, in conjunction with RNase treatment, produce results like those shown in Fig. 3. In addition, there are dyes like Hoechst 33,258 and DAPI that bind nucleic acids but only fluoresce significantly when bound to DNA. Metachromatic dyes like acridine orange fluoresce at



Cell Cycle Analysis, Flow Cytometry, Fig. 3 Common DNA content analysis. The data are from a human T lymphoma cell line (Molt4). **(a):** The model is composed of three components, two Gaussian functions (*red*) and a "broadened" polynomial (blue, hatched) that are used to fit the areas under the data that can be attributed to cells with 2C and 4C DNA (G1 and G2 + M) and for S phase (G1< S<G2). **(b):** Illustrates the fitting of the central S phase data (*circles*) to a second-order polynomial function (black line), extending that line to some predefined point under the Gaussian distributed data in either direction (here, the means of each Gaussian component), then "broadening" or "tailing" the curve with partial Gaussian functions using the coefficients of variation from Gaussian components (*red* components in a) and the amplitudes at the ends of the polynomial. (ModFit LT was used to fit the distribution in a; Graphpad Prism was used to generate b)

one wavelength when intercalated into double-stranded nucleic acids and at another wavelength when bound to single-stranded nucleic acids. This characteristic can be exploited by treating fixed cells with RNase to remove RNA and acid to partially denature residual DNA. The level of denaturability of DNA changes within the cell cycle, and by measuring fluorescence at the two wavelengths, then plotting total fluorescence (adding the two measurements) versus the fluorescence for single-stranded DNA, one can identify five cell cycle compartments – G1 can be subdivided in two, G1A or G1B (early and late), S, G2, and M. For further reading, see Darzynkiewicz et al. (1980), Darzynkiewicz and Traganos (1990).

DNA and RNA Content

Early protocols used metachromatic dyes like acridine orange which fluoresces green when intercalated into double-stranded nucleic acids and red when bound to single-stranded nucleic acids or other polyanions. Because DNA is double stranded and cellular RNA is largely single stranded, acridine orange can be used to quantify DNA and RNA. Later efforts rely on DNA-specific dyes (DAPI, Hoechst 33,258) coupled with the RNA binding dye, Pyronin Y. These approaches can distinguish deep G0 (non-cycling) cells from cycling cells in addition to the typical G1, S, and G2/M analysis of DNA content. The acridine approach requires some skill and knowledge of chemistry. Ribosomal RNA is essentially what is measured. For further reading, see Darzynkiewicz et al. (1980), Darzynkiewicz and Traganos (1990).

DNA and Protein Content

Using dyes that covalently couple with primary amines or dyes that preferentially bind protein, protein content measurements can be coupled with DNA content. In general, these approaches are not often used. Protein content correlates with RNA content. In principal, the approach should be useful for studying imbalanced growth (an imbalance between the mitotic cycle and the process of synthesizing approximately twice the cell's mass). For further information, see Crissman and Steinkamp (1987), Darzynkiewicz and Traganos (1990).

DNA Content and BrdU

The fluorescence from the DNA binding dyes, Hoechst 33,342 or 33,258, is quenched when the dye is bound to ▶ 5-Bromo-2-deoxyuridine (BrdU) substituted DNA. By coupling this dye with a dye that does not quench and fluoresces at different wavelengths, this feature has been exploited elegantly by two laboratories to measure the phase fractions of successive cell cycles. For further reading, see Poot et al. (1994).

DNA Content and Immunofluorescence

Immunofluorescence coupled with DNA content cytometry opens up a very broad ability to probe biology related to the cell cycle. Immunofluorescence probes can home in on a population of interest in complex mixtures, e.g., markers for keratin can isolate the epithelial component in a solid tumor (Glogovac et al. 1996). Other cell processes can be queried with respect to the cell cycle, e.g., activation of DNA repair pathways using the γH2Ax epitope (Tanaka et al. 2007). Below, two common immunofluorescence approaches are described.

DNA Content and BrdU BrdU-substituted DNA can be detected by antibodies. This is commonly used to either prove that the S phase cells detected by DNA content analysis were synthesizing DNA at the time of sampling (other possibilities are that the cells were dead, dying, or quiescent) or to determine the average phase times by pulse-chase or continuous labeling experiments. For further reading, see Gray et al. (1990), Gray et al. (1987), Dean (1987), Terry and White (2001).

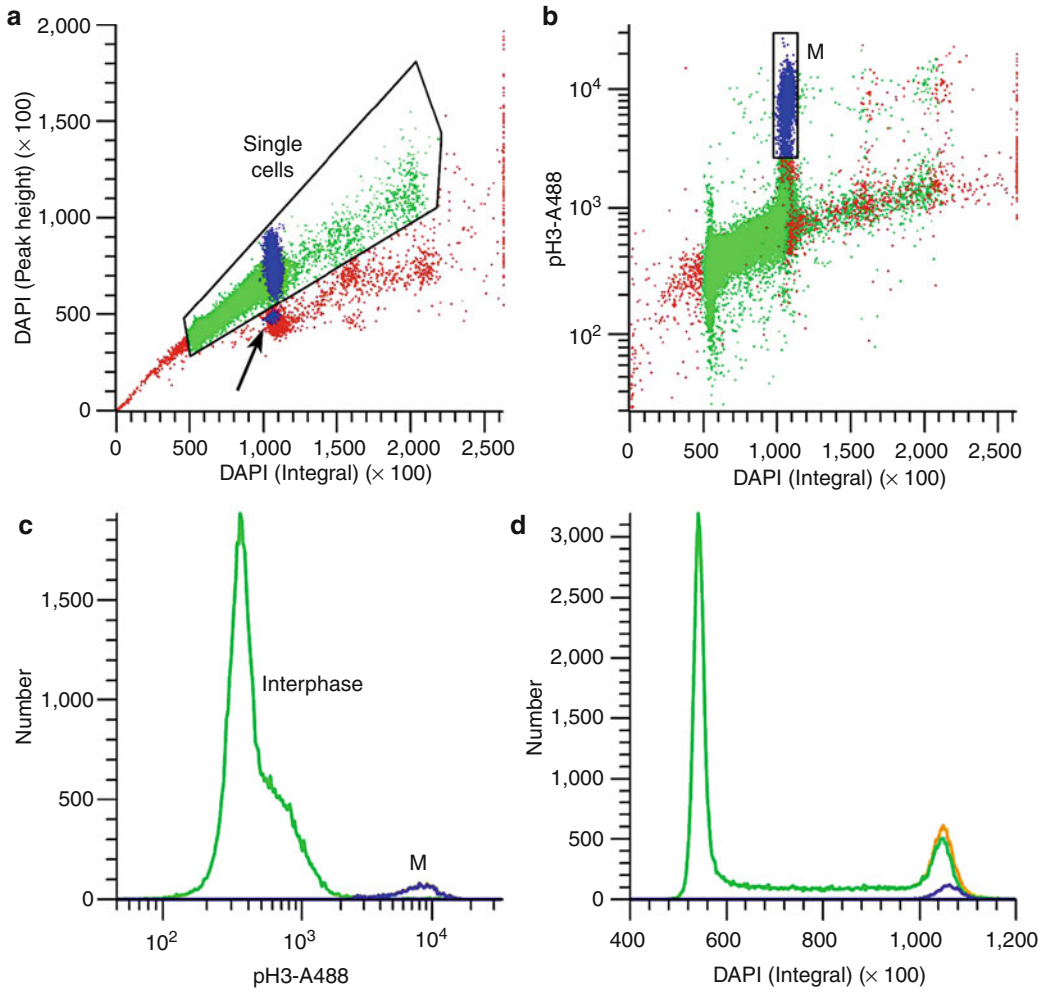
DNA and a Mitotic Marker Many proteins are specifically phosphorylated or phosphorylated more often when cells enter mitosis. Phospho-specific antibodies can be raised to these epitopes. These antibodies couple well with DNA content to provide G0/G1, S, G2, and M analysis. Figure 4 shows typical data. For a short but comprehensive review, see Darzynkiewicz (2008). For methods, see Juan et al. (2001).

Multiparametric Cell Cycle Analysis

Almost all cytometry assays of the cell cycle are multiparametric. Commonly, even for DNA content alone, measuring the pulse height and the integrated pulse (the single cell pulse as viewed on an oscilloscope) provides an ability to eliminate aggregates from the assay, and measurement of excitation light scatter provides a discriminator based on cell size and granularity or surface geometric complexity. However, these are largely utilitarian. Conversely, measuring epitopes, the expression profiles of proteins or protein modifications that oscillate within the cell cycle and that oscillate out of phase with each other provides an analytical paradigm for creating cell states that can be used to parse, in a very fine manner, the effects of various treatments on the cell cycle, e.g., drug effects.

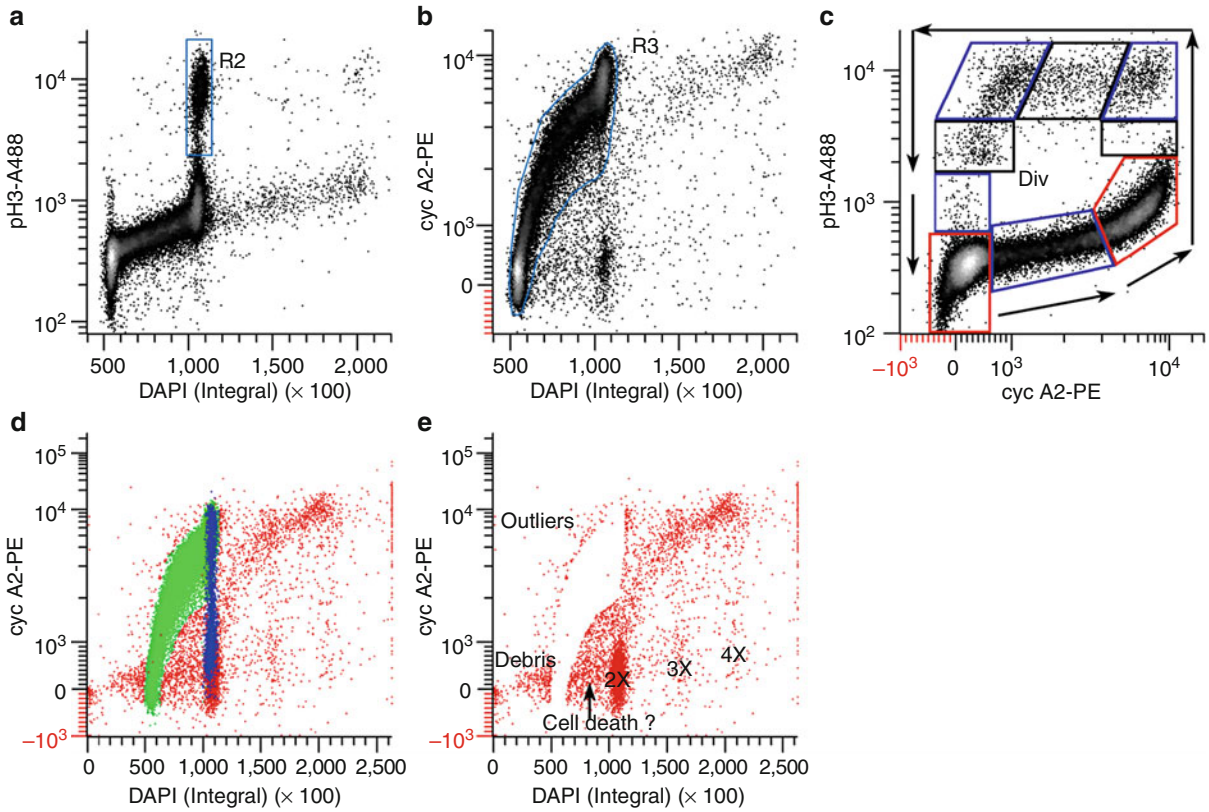
DNA, Cyclin A2, and Phospho-S10-Histone H3 (pH3)

Many epitopes that report the levels of proteins or other parent molecules can be coupled to provide additional sub-compartmentalization of the cell cycle or other information. The triad of DNA, cyclin A2, and pH3 deserves special mention because their joining into one assay provides the ability to deal with endoreduplication/binucleate cycles (see Limitations, above). Figures 4 and 5 show the principal features of this analysis. Figure 5 shows the capture of the primary (stem line) cycle (2C → 4C). Figure 5a shows a "region," R2, set around mitotic cells. Figure 5b is "NOT gated" on R2 and a region was set around the 2C → 4C interphase cells (R3). The data of both 5A and 5B have been gated to include singlet cells and exclude aggregates as in Fig. 4. Figure 5c shows all the events including singlet interphase cells (*green*), singlet mitotic cells (*blue*), and debris/dead cells/outliers and aggregates (*red*), which are not used in the analysis. Figure 5e is a look at the discarded data. The labels 2X, 3X, and 4X point to G1 aggregates (containing 2, 3, or 4 cells) together with endoreduplicated/multinucleate cycling cells for which each event is a single cell. Finally, Fig. 5c shows cyclin A2 versus pH3 for the 2C → 4C stem line. Regions are contiguously set around clusters defined either by changes in cell frequency or by changes in the "direction" of the data (*arrows*). The arrows show "movement" of an ideal cell through the data space. Div is the interface where cell division occurs. Figure 5e is particularly relevant to Systems Biology, since it is easy to see that the expression pattern of pH3 and cyclin A2 form a unidirectional closed loop, and this renders the ability to extract programmed expression profiles of these and other



Cell Cycle Analysis, Flow Cytometry, Fig. 4 DNA content analysis that differentiates mitotic cells. Exponentially growing HeLa cells were fixed and stained for DNA content, cyclin A2, and histone H3 phosphorylated at serine 10 (pH3). **Panel (a)** is a plot of the DNA signal trace* peak height versus the integrated area under the peak. This plot is used to reduce the fraction of aggregate cells in the analysis. See Rabinovitch (1993) for a discussion of the logic, theory, and practical considerations supporting this method. Singlet events, falling within the region labeled "Single cells," are colored green. Debris (events with less than a G1 level of fluorescence) and aggregate events are colored red. Mitotic cells are selected in **Panel (b)** by selecting the cells within the cluster with 4 C DNA content and high expression of pH3 (rectangle label "M"). The mitotic events are colored blue. **Panel (c)** shows the distribution of pH3 for the 2 C → 4 C interphase (green line) and M cells (blue line). Quantification of the fractions of 2 C → 4 C interphase and M cells can be done from this plot. **Panel (d)** shows the DNA content for all of the 2 C → 4 C cells (yellow line), the 2 C → 4 C interphase cells (green line), and the 2 C → 4 C M cells (blue line). Both plots in panels (c) and (d) were obtained by using the gate: [(Single cells AND R3) OR M]. (R3 is a region on cyclin A2 versus DNA, not shown in this figure, but shown in Fig. 5). Quantification of the fractions of 2 C → 4 C G1, S, and G2 are performed on a plot of the green line in Panel (d), using the methods illustrated in Fig. 3.*the digitized electrical measurement as a function of time as a cell or particle moves through the laser beam

state variables. By calculating the mean levels of these or other parameters and the frequency of cells within each region, the programmed expression profile can be obtained by the principles outlined in sections Basis for Cell Cycle Analysis and Extension of the Logic. For further reading, see Jacobberger et al. (2008), Soni et al. (2008), Avva et al. (2011), Jacobberger et al. (2011a, b).



Cell Cycle Analysis, Flow Cytometry, Fig. 5 Complex cell cycle analysis. The data file is the same as in Fig. 4. **Panel (a)** represents the same view as Panel (b) in Fig. 4, except that these data have been gated on the "Single cells" region. **Panel (b)** shows cyclin A2 versus DNA content, Boolean gated on "Single cells" AND NOT "M," which equals 2 C → 4 C and 4 C → 8 C interphases. The gate, ("Single cells" AND R3) OR R2, recombines interphase with mitosis, including mitotic cells that are excluded in the singlet region (arrow, Fig. 4a). This gate has been applied to **Panel (c)**, which provides a bivariate view of the entire 2 C → 4 C cycle starting from a state that is equivalent to "birth/early G1" and one equivalent to the end of the cell cycle. These states are separated by "Div." The arrows show the unidirectional movement of cells through the state space defined by DNA content, cyclin A2, and pH3. **Panels (d)** and **(e)** are identical plots of cyclin A2 versus DNA content, showing the 2 C → 4 C interphase population (*green dots*) and mitotic cells (*blue dots*) in Panel (d) and the events that are not used with these data removed from the plot (**Panel e**)

References

Avva J, Weis MC, Soebiyanto RP, Jacobberger JW, Sreenath SN (2011) CytoSys: a tool for extracting cell-cycle-related expression dynamics from static data. Methods Mol Biol 717:171–193

Bagwell CB (1993) Theoretical aspects of flow cytometry data analysis. In: Bauer KD, Duque RE, Shankey TV (eds) Clinical flow cytometry: principles and application. Williams & Wilkins, Baltimore, p 635

Crissman HA, Steinkamp JA (1987) Multivariate cell analysis (Techniques for correlated measurements of DNA and other cellular constituents. In: Gray JW, Darzynkiewicz Z (eds) Techniques in cell cycle analysis. Humana Press, Clifton, p 407

Csikasz-Nagy A, Battogtokh D et al (2006) Analysis of a generic model of eukaryotic cell-cycle regulation. Biophys J 90(12):4361–4379

Darzynkiewicz Z (2008) There's more than one way to skin a cat: yet another way to assess mitotic index by cytometry. Cytom A 73(5):386–387

Darzynkiewicz Z, Traganos F (1990) Multiparameter flow cytometry in studies of the cell cycle. In: Melamed MR, Lindmo T, Mendelsohn ML (eds) Flow cytometry and cell sorting. Wiley-Liss, New York, p 824

Darzynkiewicz Z, Traganos F et al (1980) New cell cycle compartments identified by multiparameter flow cytometry. Cytometry 1(2):98–108

Dean PN (1987) Data analysis in cell kinetics research. In: Gray JW, Darzynkiewicz Z (eds) Techniques in cell cycle analysis. Humana Press, Clifton, p 407
DiSalvo CV, Zhang D et al (1995) Regulation of NIH-3 T3 cell G1 phase transit by serum during exponential growth. Cell Prolif 28(9):511–524
Frisa PS, Jacobberger JW (2009) Cell cycle-related cyclin b1 quantification. PLoS One 4:e7064
Glogovac JK, Porter PL et al (1996) Cytokeratin labeling of breast cancer cells extracted from paraffin-embedded tissue for bivariate flow cytometric analysis. Cytometry 24(3): 260–267
Gray JW, Dolbeare F et al (1987) Flow cytokinetics. In: Gray JW, Darzynkiewicz Z (eds) Techniques in cell cycle analysis. Humana Press, Clifton, p 407
Gray JW, Dolbeare F et al (1990) Quantitative cell-cycle analysis. In: Melamed MR, Lindmo T, Mendelsohn ML (eds) Flow cytometry and cell sorting. Wiley-Liss, New York, p 824
Jacobberger JW, Frisa PS et al (2008) A new biomarker for mitotic cells. Cytom A 73(1):5–15
Jacobberger JW, Sramkoski RM et al (2011a) Multiparameter cell cycle analysis. Methods Mol Biol 699:229–249
Jacobberger JW, Sramkoski RM et al (2011b) Analysis of cell cycle cytometry data
Juan G, Traganos F, Darzynkiewicz Z (2001) Methods to identify mitotic cells by flow cytometry. Methods Cell Biol 63:343–354
Mitchison JM (1971) The biology of the cell cycle. Cambridge University Press, Cambridge
Poot M, Hoehn H et al (1994) Cell-cycle analysis using continuous bromodeoxyuridine labeling and Hoechst 33358-ethidium bromide bivariate flow cytometry. Methods Cell Biol 41:327–340
Rabinovitch PS (1993) Practical considerations for DNA content and cell cycle analysis. In: Bauer KD, Duque RE, Shankey TV (eds) Clinical flow cytometry: principles and application. Williams & Wilkins, Baltimore, pp 117–142
Soni DV, Sramkoski RM et al (2008) Cyclin B1 is rate limiting but not essential for mitotic entry and progression in mammalian somatic cells. Cell Cycle 7(9):1285–1300
Tanaka T, Huang X et al (2007) Cytometry of ATM activation and histone H2AX phosphorylation to estimate extent of DNA damage induced by exogenous agents. Cytom A 71(9):648–661
Terry NH, White RA (2001) Cell cycle kinetics estimated by analysis of bromodeoxyuridine incorporation. Methods Cell Biol 63:355–374
Walker PMB (1954) The mitotic index and interphase processes. J Exp Biol 31(1):8–15
Watson JV (1991) Introduction to flow cytometry. Cambridge University Press, Cambridge
Yan T, Desai AB et al (2004) CHK1 and CHK2 are differentially involved in mismatch repair-mediated 6-thioguanine-induced cell cycle checkpoint responses. Mol Cancer Ther 3(9):1147–1157

Cell Cycle Analysis, Live-Cell Imaging

Andreas Doncic and Jan M. Skotheim
Department of Biology, Stanford University School of Medicine, Stanford, CA, USA

Synonyms

Single-cell time-lapse microscopy; Time-lapse microscopy

Definition

Time-lapse microscopy is a form of live-cell imaging (▶ Cell Cycle Analysis, Live-Cell Imaging) where living cells are grown and monitored over time. Single-cell analysis of the cell cycle has been used to study the genetic networks regulating ▶ cell cycle transitions. This technique is able to leverage variability in individual cells to uncover regulatory features of the underlying molecular network. In addition, the long durations of time-lapse experiments can be used to minimize the influence of stress from sample preparation. However, care must be taken regarding illumination levels and environmental conditions so that photodamage or a variable environment does not perturb the cell cycle. Environmental conditions may be accurately controlled using microfluidic devices permitting imaging. In summary, the confluence of improved fluorescent proteins and ever-increasing computational speed and memory has led to new age in live-cell imaging.

Characteristics

Requirements: Stable Conditions

Live-cell microscopy requires cells to be kept alive during the course of the experiment. At a minimum, this necessitates reliable temperature control and consistent nutrient supply. For metazoan cells, such as mammalian tissue culture cell lines frequently used in biomedical research, care must also be taken to control the CO_2-concentration. In addition, growth

and proliferation of mammalian cells depends on their cell density so studies of cell cycle kinetics are necessarily affected. Thus, cell cycle researchers using mammalian systems should carefully control for cell density.

Requirements: Conditions for Optimal Imaging

It is helpful for cells to remain in place for imaging. Steady temperature control will remove a certain amount of drift in the z-direction. Mammalian cells will adhere to the hard surfaces at the bottom of tissue culture dishes, some of which have glass slide bottoms for imaging purposes. In addition, hard dishes or soft hydrogels can be coated with fibronectin or collagen to promote adherence and proliferation. For many mammalian cell applications, it is preferable to use relatively low magnification (4×, 10× or 20×) because the larger depth of focus of low magnification dry lenses makes it easier to keep the cells in focus over long times. Unless the dynamics of small cellular structures need to be resolved, it is recommended to use low magnification, which can be easily used to quantify fluorescence in either the whole cell or the nucleus.

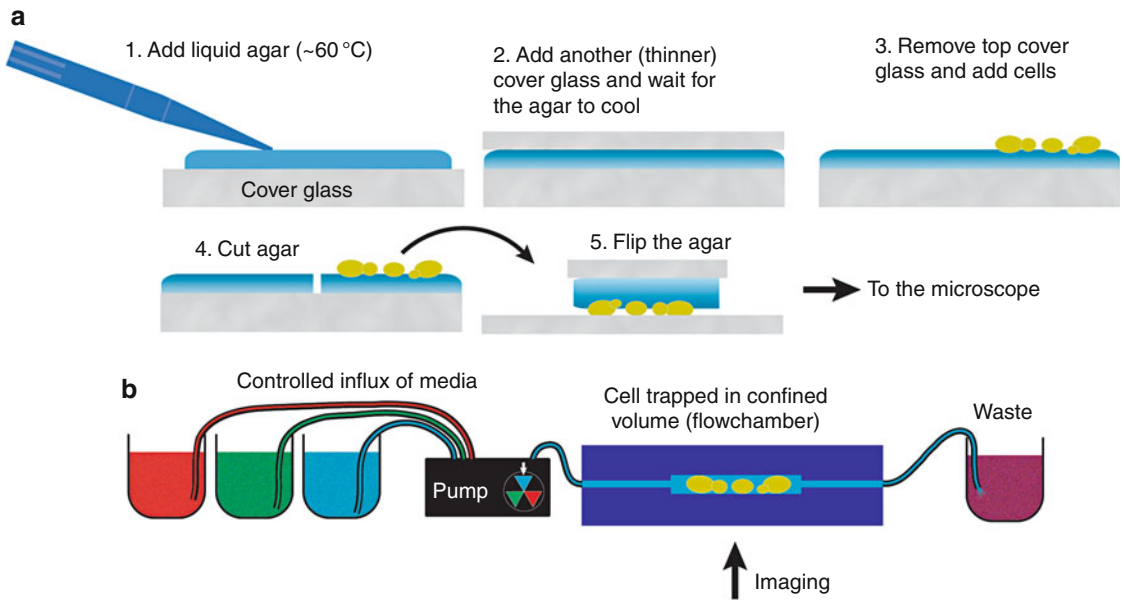
To constrain yeast cells, they are normally either sandwiched between two surfaces in the imaging chamber (agar/flowcell) or physically attached to the cover slip. Attaching yeast with the lectin protein concanavalin A (conA) might work for some applications but it is not recommended for long-time imaging since new buds either disappear or grow out of the imaging plane. If only one condition/medium is going to be used, yeast can be forced to grow in 2D by sandwiching the cells between the cover glass and a slab of ~1.5% low-melt agar mixed with the growth medium (Fig. 1a). For yeast imaging, we recommend 63× high NA oil-immersion lenses. These lenses balance the need for magnification with the need to collect as much light as possible and 100× lenses will produce substantially darker images, which is a drawback for photosensitive yeasts. A drawback of high magnification oil-immersion lenses is the thin focal plane, which requires careful focusing in the z-direction through the course of the experiment. In particular, the focusing problem is exacerbated when multiple x-y positions are used because moving the lens will produce a high pressure in the lubricating oil film tending to push the slide out of focus. Thus, either software or hardware-based auto-focusing routines must be used to keep the sample in focus through the experiment. We recommend hardware-based focusing (e.g., Zeiss Definite Focus; Nikon Perfect Focus), which is rapid and accurate.

Requirements: Environmental Control

The integration of live-cell imaging with microfluidic devices (flowcell) capable of accurately controlling and rapidly changing the extracellular environment will prove increasingly useful. Rapid change in the extracellular media can be used to analyze responses to environmental change (e.g., adding mating pheromone to budding yeast as in Fig. 2) and to exogenously control expression of synthetic constructs. For example, in budding yeast, the *MET3* promoter can be placed ahead of any gene of interest so that expression occurs only in media lacking methionine. It takes 5–10 min to activate gene expression upon methionine removal, while the addition of methionine to the medium shuts off expression in approximately 5 min (Charvin et al. 2008).

Even though a variety of microfluidic devices exist, handling different types of analysis (large/small-scale) or different cell types (yeast, mammalian, etc.), the basic principle behind them remains the same (Fig. 1b): Cells are trapped in a chamber where the influx of media and temperature is precisely controlled. The chamber is located on top of the microscope objective for imaging. The commercially available yeast flowcell from Cellasic is easy to use and functions well for low throughput experiments (www.cellasic.com). The high throughput system from the Hansen group is currently the state of the art for yeast work (Taylor et al. 2009). A similar platform for mammalian cell imaging from the Quake lab is also recommended (Gomez-Sjoberg et al. 2007).

Regardless of the type of experimental setting it is important not to overexpose the cells while using ▶ fluorescence microscopy. Overexposure may lead to cell cycle arrest and/or cell death. It is recommended to control for overexposure by measuring cell cycle kinetics (rate of cell division) with and without fluorescence. Particular care must be taken when multiple images of a z-section are desired. Budding and fission yeasts are more sensitive to fluorescence imaging than human cell lines.



Cell Cycle Analysis, Live-Cell Imaging, Fig. 1 Two ways to prepare yeast cells for imaging: **(a)**. Add 3–4 ml of 1.5% low-melt agar, heated to ~60°C, to a cover glass. Leave it to cool a few minutes and then sandwich the agar with a cover slip and wait until the agar cools completely. Once the agar is cool, remove the top cover glass and add the cells on top of the agar. Cut out the region of agar containing the cells and flip it 180° onto the imaging cover glass/dish used for microscope imaging. **(b)**. Schematic of a microfluidic device. Cells are grown in a flowcell where media influx and temperature are controlled externally

Data Acquisition

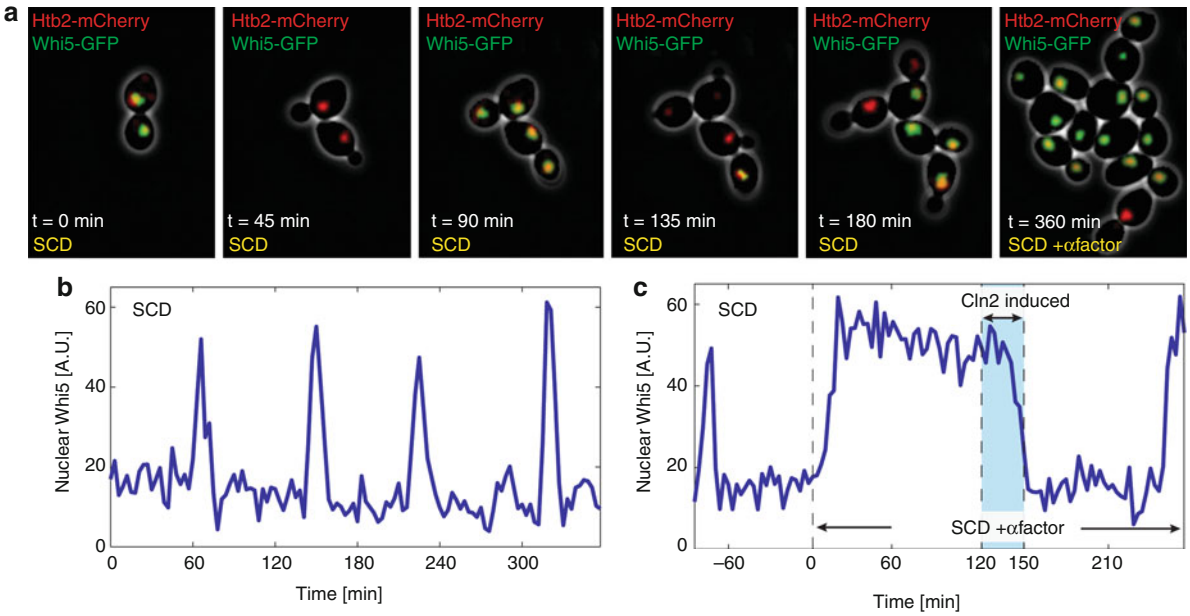
There are two general ways to monitor cells through time: Either track the individual cells through time or sample large numbers of cells without replacement (population-based measurements) after cell cycle synchronization. On one hand, single-cell analysis does not require perturbative synchronization methods, reveals sharper transitions, and can be used to measure cell-to-cell variation due to stochastic fluctuations. On the other hand, population-based methods (e.g., flow cytometry, coulter counter) can be used to rapidly acquire large numbers of measurements, and can be used in conjunction with standard biochemical techniques.

Image acquisition is normally done using ▶ fluorescence microscopy in combination with an additional imaging technique (e.g., bright field, phase contrast, or differential interference contrast) to determine the location of the cells. Depending on the cellular process studied, a variety of ▶ fluorescent markers may be used. Recent technical advances by the Tsien group have greatly expanded the quality and color spectrum of available fluorophores (Shaner et al. 2005). The best ▶ fluorescent markers used for cell cycle analysis have the following qualities:

- Bright enough to yield a good signal at low intensity illumination that does not induce photodamage. Enhanced fast folding GFP, or Venus (yellow) are both excellent; mCherry (red) is much dimmer and has slower maturation kinetics.
- Clearly mark one/several cell cycle events. In general, measurements of changes in protein localization or degradation are preferable to measurements of changing concentration because even fast maturing GFP proteins fold in ~10–15 min.
- Do not interfere with the natural function of the fluorescently labeled protein.

Commonly used markers for some cell cycle transitions can be found in Table 1:

A novel technique for live-cell imaging developed by the Pines group relies on a FRET (fluorescence resonance energy transfer) sensor to measure CDK-Cyclin B activity in single live cells (Gavet and Pines 2010) Similarly, the Kapoor lab used FRET probes to measure spatial and temporal gradient of aurora B activity during mitosis in live cells



Cell Cycle Analysis, Live-Cell Imaging, Fig. 2 (**a**). Composite phase and fluorescence images of budding yeast cells grown in a flowcell with synthetic complete glucose medium (SCD). This strain contains integrated Htb2-mCherry (a nuclear marker) and Whi5-GFP fusion proteins (marking early G1, see table1). Note that the cells arrest after addition of α-factor (mating pheromone). (**b**). Example nuclear Whi5-GFP time series of a cell grown in SCD. Whi5-GFP enters the nucleus at anaphase and exits the nucleus at the point of commitment to cell division in mid-G1. (**c**). Example nuclear Whi5-GFP time course of a cell arrested in α-factor (added at time = 0) and then forced out of arrest using the inducible *MET3*-promoter to express the G1-cyclin *CLN2* (see text)

(Fuller et al. 2008). The ability to measure protein activity, rather than protein concentration, in individual cells is a great advantage because of the various inhibiting and activating post-translational modifications affecting function. Thus, we expect FRET-based techniques to become increasingly important.

Analysis

Image Analysis

To obtain single-cell data it is necessary to segment (finding the cell outline) and track individual cells over time. Segmentation is normally done on the entire cell or on some subcellular compartment of special interest like the nucleus (Sigal et al. 2006). In general, segmentation starts by manipulating the original image to enhance the difference between "cell" and "non-cell" regions. The exact nature of this step depends on the imaging and the markers used. Next, a threshold is applied to distinguish cells from non-cells. After the cells are identified, parameters useful for cell cycle data analysis such as size, average fluorescence, and timing of signal appearance/disappearance can be extracted.

Although standard algorithms and commercial products exist for image analysis, each application tends to be unique enough to require customization. As the image processing needs of the scientific community increase in the coming decade, we expect commercial products to improve and be more widely applied.

An example of Cell Cycle Analysis Using Live-Cell Imaging

A striking example of the potential in single-cell cell cycle analysis can be seen in the work by Di Talia et al. to measure size control in budding yeast (Di Talia et al. 2007). Yeast exhibit size homeostasis: Small cells must grow more than larger cells through their cell cycle (▶ Cell Cycle, Cell Size Regulation). To examine this outstanding problem, Di Talia and colleagues used yeast labeled with a Myo1-GFP fusion protein (used to divide the cell cycle into G1 and S/G2/M intervals) and *ACT1pr-DsRed*, a red fluorescent protein under the control of the strong and constitutive *ACT1*-promoter (used to measure cell size).

Cell Cycle Analysis, Live-Cell Imaging, Table 1 Some commonly used cell cycle markers

Cell cycle event/phase	Cellular process/localization	Gene(s)	Organism/ cell type	Ref
Cytokinesis/G1 initiation	Disappearance of the septin ring	Cdc10, Myo1	Budding yeast	(Di Talia et al. 2007)
Start transition (mid G1)	Whi5 export from the nucleus	Whi5	Budding yeast	(Di Talia et al. 2007)
S-phase initiation	Reappearance of the septin ring	Cdc10, Myo1	Budding yeast	(Di Talia et al. 2007)
Metaphase to anaphase transition	Nuclear separation (also gives lineage information)	Htb2 or any other nuclear marker	Budding yeast	Many
Cell cycle transitions, mitotic exit	Clb2 degradation/Cdc14 localization	Clb2, Cdc14	Budding yeast	(Drapkin et al. 2009)
Mitosis	Anaphase spindle status	Tub1	Budding yeast	(Drapkin et al. 2009)
G1 & G2	Sensor nuclear	Proliferating cell nuclear antigen (PCNA)-based biosensor	Human cell line	(Hahn et al. 2009)
S	Sensor nuclear at replication foci	PCNA-based biosensor	Human cell line	(Hahn et al. 2009)
M	Sensor located throughout the cell	PCNA-based biosensor	Human cell line	(Hahn et al. 2009)
End of M-phase to G1/S	Sensor nuclear	Truncated human DNA helicase B (HDHB)-based biosensor	Human cell line	(Hahn et al. 2009)
S & G2, cell cycle transitions, G2/M	Sensor cytoplasmic	HDHB-based biosensor	Human cell line	(Hahn et al. 2009)
M	Sensor located throughout the cell	HDHB-based biosensor	Human cell line	(Hahn et al. 2009)
G1 to S & G1	Licensing	Fragment of Cdt1 bound to mKO (red/orange)	Metazoan cells	(Sakaue-Sawano et al. 2008)
S, G2 & M	Licensing inhibition	Fragment of geminin bound to mAG (green)	Metazoan cells	(Sakaue-Sawano et al. 2008)

Using these markers, Di Talia and colleagues were able to measure both cell sizes and the duration of G1-periods for single cells. Because individual cells are of various sizes at birth and budding, correlation between cell size and interval duration is possible. Thus, single-cell variability is leveraged to infer regulatory properties. Di Talia et al. found that only smaller daughter cells exhibited size control (an inverse correlation between size at birth and G1 duration).

Cross-References

- ▶ Cell Cycle Data Analysis
- ▶ Cell Cycle Transitions, G2/M
- ▶ Cell Cycle Transitions, Mitotic Exit
- ▶ Cell Cycle, Cell Size Regulation
- ▶ Fluorescence Microscopy
- ▶ Fluorescent Markers
- ▶ Live Cell Imaging
- ▶ Mitosis

References

Charvin G, Cross FR, Siggia ED (2008) A microfluidic device for temporally controlled gene expression and long-term fluorescent imaging in unperturbed dividing yeast cells. PLoS ONE 3:e1468

Di Talia S, Skotheim JM, Bean JM, Siggia ED, Cross FR (2007) The effects of molecular noise and size control on variability in the budding yeast cell cycle. Nature 448:947–951

Drapkin BJ, Lu Y, Procko AL, Timney BL, Cross FR (2009) Analysis of the mitotic exit control system using locked levels of stable mitotic cyclin. Mol Syst Biol 5:328

Fuller BG, Lampson MA, Foley EA, Rosasco-Nitcher S, Le KV, Tobelmann P, Brautigan DL, Stukenberg PT, Kapoor TM (2008) Midzone activation of aurora B in anaphase

produces an intracellular phosphorylation gradient. Nature 453:1132–1136
Gavet O, Pines J (2010) Progressive activation of CyclinB1-Cdk1 coordinates entry to mitosis. Dev Cell 18:533–543
Gomez-Sjoberg R, Leyrat AA, Pirone DM, Chen CS, Quake SR (2007) Versatile, fully automated, microfluidic cell culture system. Anal Chem 79:8557–8563
Hahn AT, Jones JT, Meyer T (2009) Quantitative analysis of cell cycle phase durations and PC12 differentiation using fluorescent biosensors. Cell Cycle 8:1044–1052
Sakaue-Sawano A, Kurokawa H, Morimura T, Hanyu A, Hama H, Osawa H, Kashiwagi S, Fukami K, Miyata T, Miyoshi H, Imamura T, Ogawa M, Masai H, Miyawaki A (2008) Visualizing spatiotemporal dynamics of multicellular cell-cycle progression. Cell 132:487–498
Shaner NC, Steinbach PA, Tsien RY (2005) A guide to choosing fluorescent proteins. Nat Meth 2:905–909
Sigal A, Milo R, Cohen A, Geva-Zatorsky N, Klein Y, Alaluf I, Swerdlin N, Perzov N, Danon T, Liron Y, Raveh T, Carpenter AE, Lahav G, Alon U (2006) Dynamic proteomics in individual human cells uncovers widespread cell-cycle dependence of nuclear proteins. Nat Meth 3:525–531
Taylor RJ, Falconnet D, Niemisto A, Ramsey SA, Prinz S, Shmulevich I, Galitski T, Hansen CL (2009) Dynamic analysis of MAPK signaling using a high-throughput microfluidic single-cell imaging platform. Proc Natl Acad Sci USA 106:3758–3763

Cell Cycle Analysis, Phase Fraction Analysis

► Cell cycle analysis, flow cytometry

Cell Cycle Analysis, Systematic Gene Overexpression

Hisao Moriya
Okayama University Research Core for Interdisciplinary Sciences, Kita-ku, Okayama, Japan

Definition

The ► cell cycle is achieved by the regulators that are expressed in appropriate amounts, in appropriate cell cycle periods. To elucidate the regulatory mechanism of the cell cycle, experiments in which the regulators are artificially overexpressed and the effects are observed have often been performed. By performing this gene overexpression experiment (► Cell Cycle Analysis, Systematic Gene Overexpression) systematically, novel cell cycle regulators that could not be isolated by loss-of-function experiments can be identified. There is also an approach to analyze ► robustness of the cell cycle by measuring limits of gene overexpression.

Characteristics

Introduction

Overexpression is artificially performed by ► promoter-swapping or copy number increase of the target gene. In model organisms in which these genetic manipulations are available, the cell cycle analyses by the systematic gene overexpression have been performed. Genes produce cell cycle defects upon overexpression have been isolated by the phenotype of transformed cells or individual organisms with cDNAs or genomic fragments expressed under the control of strong inducible promoters. In the post-genomic era, DNA libraries containing each of every protein-coding sequence on the genome have been constructed. Using those libraries, genuine whole genome overexpression-based screenings have been performed. Currently the budding yeast is the only model organism in which these analyses have been done.

In addition to the screening, there is an approach to study the robustness of the cell cycle by the systematic measurement of limits of overexpression of the cell cycle–related genes. The limit data was also used to evaluate and refine an integrative cell cycle model. The concrete examples of the cell cycle analysis of individual model organisms using systematic gene overexpression are introduced below.

Budding Yeast (*Saccharomyces cerevisiae*)

Several extensive screenings to identify genes that cause the cell cycle defect upon overexpression have been performed in the budding yeast (Stevenson et al. 2001; Sopko et al. 2006; Niu et al. 2008). In all these studies, *GAL1* promoter by which the target gene is highly expressed in galactose medium was used, and ► flow cytometry was used to detect the cell cycle defects. Through these analyses, about 250 genes were identified. Because the set of genes includes cyclin genes that play central roles in the cell cycle, the set should include the genes directly involved in the

cell cycle regulation. On the other hand, many genes those are known to be involved in other processes than the cell cycle were also identified, though for all genetic experiments it is difficult to circumvent obtaining indirect factors.

Along with the extension of our understanding of the cell cycle mechanisms, an integrative mathematical model of the cell cycle has been constructed (Chen et al. 2004). Robustness against parameter fluctuations is useful to evaluate the plausibility of biological models (Morohashi et al. 2002). By measuring limits of gene overexpression systematically, promoter of the cell cycle of real cell against (gene expression) parameter fluctuations can be measured. genetic Tug-Of-War (gTOW) method is used for this purpose. In gTOW, instead of promoter swapping, the target gene with its native promoter is cloned into a plasmid of which the copy number increases in the leucine-limiting condition. If overexpression of the target gene causes the cell cycle defect, the copy number of the plasmid (and the gene) per cell must be less than the overexpression limit. Using gTOW, limits of overexpression of about 30 cell cycle regulator genes were measured, and the data was used to evaluate and refine the integrative mathematical model (Moriya et al. 2006; Kaizu et al. 2010).

Fission Yeast (*Schizosaccharomyces pombe*)

The fission yeast is another model organism in which the regulatory system of the cell cycle has been extensively studied. In this yeast, the target gene can be artificially overexpressed using *nmt1* promoter that is induced in thiamine-limiting condition. Although the induction of *nmt1* promoter is relatively slower (∼18 h) than the budding yeast *GAL1* promoter (∼1 h), defects of the fission yeast cell cycle can be easily observed with the elongated morphology (*cdc* phenotype) under the microscope. Until now, 21 genes were isolated by the systematic screening with *cdc* phenotype of the transformants of a cDNA library (Tallada et al. 2002). Technically, genuine exhaustive genomic screening and gTOW can be performed in the fission yeast as well.

Higher Eukaryotes

In higher eukaryotes (multicellular organisms), in addition to the difficulty in genetics manipulations, it is a matter how defects in the cell cycle are observed. There are thus few examples of the systematic screenings of the genes that cause the cell cycle defects upon overexpression. In the fruit fly, there is a system (*GAL4* driving systems) by which the target gene can be overexpressed in organ specific manner. Using 2,300 fly lines in which each target locus (gene) is overexpressed, 32 genes were isolated by the "small eye phenotype" that is an indicator of the cell cycle defects (Tseng and Hariharan 2002). It is estimated that only less than 10% genes are studied in this analysis.

The Future of the Cell Cycle Analysis Using Systematic Gene Overexpression

Genetic screening using overexpression enables us to isolate genes that do not show any cell cycle defect upon gene disruption. For example, cyclin genes that make gene family in the budding yeast does not show significant phenotype upon disruption, but they show strong cell cycle defect upon gene overexpression. However, while the gene knockout experiment gives the unified result because the target gene is completely inactivated, gene overexpression experiments usually do not give unified results because there are many ways to overexpress the target gene (i.e., promoter swapping using different promoters, copy number increase), which lead to the different expression levels. In fact, there are surprisingly few overlaps among the gene sets isolated in the three budding yeast genome-wide screening described above (Niu et al. 2008).

To obtain the unified result in the overexpression of every gene, we need to systematically measure the overexpression limit that causes certain cell cycle defect. An experiment to continuously observe the cell cycle defect upon gradual increase of the target gene expression will be a one suitable way. Such experiment can be done in the model organisms such as the budding yeast (e.g., Charvin et al. 2008). In higher eukaryotes, the similar technology needs to be developed using cultured cell first.

Cross-References

- ▶ Cell Cycle
- ▶ Cell Cycle Analysis, Flow Cytometry
- ▶ Promoter
- ▶ Robustness

References

Charvin G, Cross FR, Siggia ED (2008) A microfluidic device for temporally controlled gene expression and long-term fluorescent imaging in unperturbed dividing yeast cells. PLoS ONE 3(1):e1468

Chen KC, Calzone L, Csikasz-Nagy A, Cross FR, Novak B, Tyson JJ (2004) Integrative analysis of cell cycle control in budding yeast. Mol Biol Cell 15(8):3841–3862

Kaizu K, Moriya H, Kitano H (2010) Fragilities caused by dosage imbalance in regulation of the budding yeast cell cycle. PLoS Genet 6(4):e1000919

Moriya H, Shimizu-Yoshida Y, Kitano H (2006) In vivo robustness analysis of cell division cycle genes in *Saccharomyces cerevisiae*. PLoS Genet 2(7):e111

Morohashi M, Winn AE, Borisuk MT, Bolouri H, Doyle J, Kitano H (2002) Robustness as a measure of plausibility in models of biochemical networks. J Theor Biol 216(1):19–30

Niu W, Li Z, Zhan W, Iyer VR, Marcotte EM (2008) Mechanisms of cell cycle control revealed by a systematic and quantitative overexpression screen in *S. cerevisiae*. PLoS Genet 4(7):e1000120

Sopko R, Huang D, Preston N, Chua G, Papp B, Kafadar K, Snyder M, Oliver SG, Cyert M, Hughes TR, Boone C, Andrews B (2006) Mapping pathways and phenotypes by systematic gene overexpression. Mol Cell 21(3):319–330

Stevenson LF, Kennedy BK, Harlow E (2001) A large-scale overexpression screen in *Saccharomyces cerevisiae* identifies previously uncharacterized cell cycle genes. Proc Natl Acad Sci USA 98(7):3946–3951

Tallada VA, Daga RR, Palomeque C, Garzón A, Jimenez J (2002) Genome-wide search of *Schizosaccharomyces pombe* genes causing overexpression-mediated cell cycle defects. Yeast 19(13):1139–1151

Tseng AS, Hariharan IK (2002) An overexpression screen in *Drosophila* for genes that restrict growth or cell-cycle progression in the developing eye. Genetics 162(1):229–243

Cell Cycle Arrest After DNA Damage

Jared Toettcher
University of California, San Francisco, CA, USA

Synonyms

Cell cycle arrest mechanisms; DNA damage response; Extrinsic DNA damage checkpoints; p53 and the cell cycle

Definition

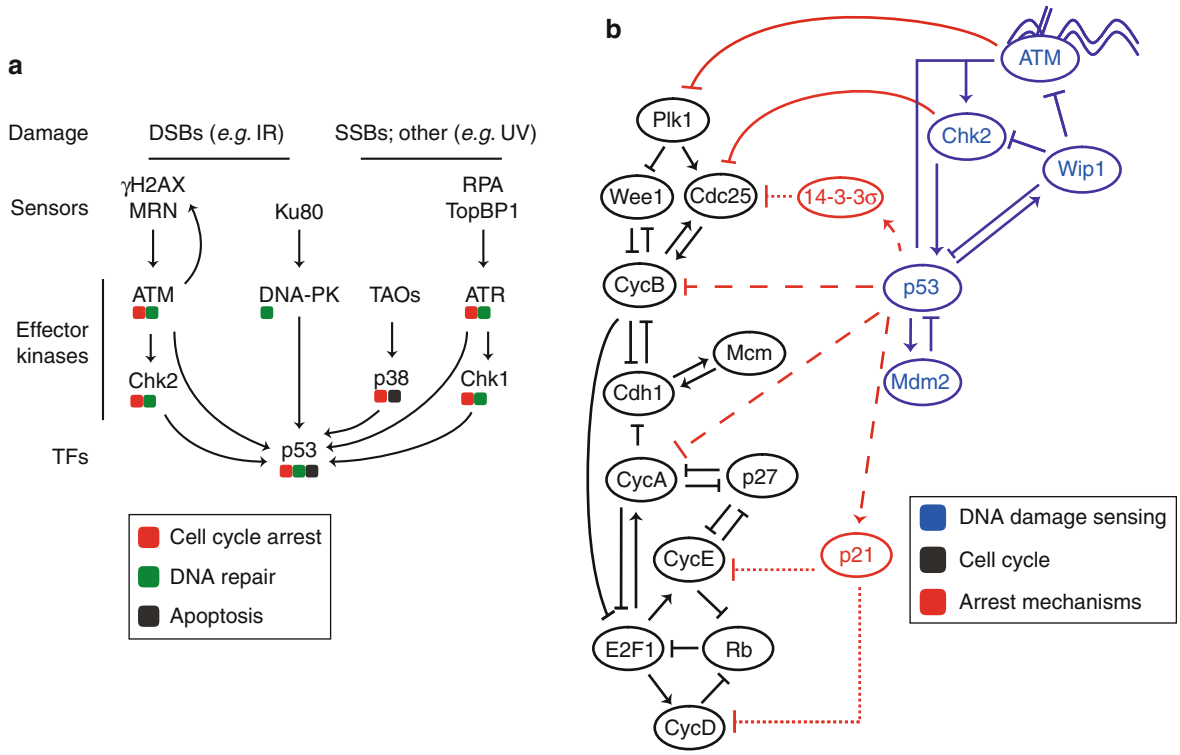
Cell cycle arrest after DNA damage describes the interconnection between two complex signaling processes – DNA damage sensing and the cell cycle – by a variety of biochemical interactions. Damage may arise from various sources, including radiation, chemical agents, or errors during DNA synthesis or cell division. The resulting damage is sensed by a signaling network that halts the cell cycle by modulating cyclin/Cdk activity. Cell cycle arrest can be transient to allow repair of DNA damage, or can persist indefinitely as a senescence-like state. This essay describes mechanisms of DNA damage-induced cell cycle arrest, their dynamics, and their effect on eventual cell fate. It also discusses mathematical modeling approaches used to gain insight into these processes.

Characteristics

Arrest Mechanisms: Connecting DNA Damage Signaling to the Cell Cycle

Progression through the cell cycle can be viewed as having two fundamental goals: the duplication of genetic information during the synthesis phase (S phase), and the division of this material among daughter cells during mitosis (M phase) (▶ Cell Cycle of Mammalian Cells). Damage to the cell's DNA interferes with both goals, leading to the possibility of erroneous genetic material being replicated or partitioned among daughters. Such errors can drive mutagenesis, and play a central role in cancer progression.

The first line of defense against such errors is to stop the cell cycle when damage is sensed, preventing the initiation of DNA replication or cell division. This process, known as cell cycle arrest, must satisfy three criteria: it must be activated quickly after damage is delivered, to prevent cell cycle progression in the presence of damage; it must remain active as long as damage is present; and finally, if cells resume cycling after arrest, they must reenter the cell cycle in the same phase from which they exited. Violation of any of these conditions could lead to mutagenic errors. A group of molecular interactions that transmit the presence of damage to halt the cell cycle is known as a cell cycle



Cell Cycle Arrest After DNA Damage, Fig. 1 DNA damage signaling and its connection to the cell cycle. (**a**) DNA damage signaling connects damage sensors to cell fate through multiple levels of effectors. The presence of damage leads to the local activation of sensor proteins at the damage sites. These sensors in turn activate kinase cascades leading to growth arrest (indicated by *red boxes*), DNA repair (*green boxes*) or apoptosis (*black boxes*). (**b**) Multiple mechanisms of arrest connect DNA damage to the cell cycle. Pictured here are arrest mechanisms implicated in the response to DNA double-stranded breaks. Post-translational catalytic steps (*solid lines*), stoichiometric inhibitor binding (*dotted lines*) and transcriptional activation or repression (*dashed lines*) for each arrest mechanism are indicated

arrest mechanism. Many arrest mechanisms are known, and can act to halt the cell cycle in G1, S, G2, or M phase.

The DNA Damage Response: Multiple Branches and Timescales

Damage to the cell's DNA is sensed by proteins of three different classes: sensors that interact directly with damaged DNA; effectors that undergo post-translational modification, allowing them to mediate a downstream signaling response; and in some cases transcription factors that mediate longer-term transcriptional responses (Fig. 1a).

The first two classes – sensors and effectors – act quickly to halt the cell cycle, and are often considered together as the mediators of extrinsic cell cycle checkpoints (Reinhardt and Yaffe 2009). They are discussed in more detail in the corresponding Essay (▶ Cell Cycle Checkpoints). DNA damage leading primarily to double-stranded breaks (DSBs) is sensed by the Mre11-Nbs1-Rad50 (MRN) protein complex binding at the site of damage. The MRN complex activates an effector kinase cascade including the phosphoinositide 3-kinase (PI3K) ataxia telangiectasia mutated kinase (ATM) and checkpoint kinase 2 (Chk2). A positive feedback loop whereby ATM phosphorylates histone H2AX at the site of damage leading to additional ATM activation ensures a strong damage response even to a single DSB. A second PI3K, DNA-PK, is also activated at the site of DSBs.

Other forms of damage, including DNA adducts and single-stranded breaks (SSBs), are sensed by a homologous, parallel kinase cascade. Again, DNA binding proteins such as RPA localize at damage sites, and proceed to activate kinases such as ATM-related kinase (ATR) and thousand-and-one amino acid

kinase (TAO). Subsequent kinase cascades activate checkpoint kinase 1 (Chk1) and the mitogen-activated protein kinase p38. This is by no means an exhaustive list of components involved in sensing and transmitting damage signals, and many additional regulators affect the activity of the components described here. In addition, specific programs of DNA repair induced by each damage type depend highly on the subset of damage sensors and transducers activated as well as the current cell cycle phase (▶ DNA Repair).

In contrast to the sensor-effector machinery, much of which is broadly conserved from yeast to mammals, the slower transcriptional arm of DNA damage-induced cell cycle arrest appears to be restricted to multicellular organisms. It plays a similar role in species ranging from *Drosophila melanogaster* to humans. Its central player is the stress-induced transcription factor p53. In mammals, the DNA damage kinases p38, ATM, ATR, DNA-PK, Chk1, and Chk2 can all phosphorylate p53, protecting it from homeostatic degradation by ubiquitin ligases such as Mdm2. A variety of other post-translational modifications can tune p53 stability and transcriptional activity (acetylation, methylation, SUMOylation, etc.). Elevated levels of modified p53 lead to both the induction and repression of various genes involved in the cell cycle, apoptosis, and DNA repair.

p53 activation is itself regulated by multiple positive and negative feedback loops that serve to integrate information between cell survival and stress pathways or maintain low levels of p53 under unstressed conditions (Harris and Levine 2005). Moreover, this feedback regulation has consequences for p53's temporal activation. DSB induction across a broad range of mammalian contexts – from cultured fibroblasts to live mice – leads to a series of pulses of p53 activation on a timescale of hours (Batchelor et al. 2009). It has been shown that oscillation depends on two negative feedback connections: p53 transcriptionally induces its negative regulator Mdm2 as well as the phosphatase Wip1, which can inactivate p38, ATM, and Chk2 (Batchelor et al. 2008).

The Molecular Logic of DNA Damage Signaling to the Cell Cycle

Many connections between DNA damage signaling and the cell cycle have been identified. These connections have largely been uncovered through experiments identifying interacting partners of key cell cycle regulators (e.g., cyclins, Cdc25 phosphatases, Plk1), and overexpressing or disrupting them individually to induce or abrogate arrest. They act on both G1 and G2 cell cycle components by a variety of biochemical mechanisms, including stoichiometric inhibition, enzymatic modification, and transcriptional induction/repression of cell cycle proteins.

Kinase, phosphatase and protease activity provide the fastest mode of regulation on the cell cycle (Fig. 1b). These connections proceed from DNA damage effector kinases such as ATM, Chk1, and Chk2 to phosphorylate various cell cycle targets. For instance, to signal G1 arrest, ATM can phosphorylate cyclin D to target it for degradation. ATM can also act on G2 cells by phosphorylating Plk1, preventing cyclin B/Cdk1 activation. The dual specificity phosphatases Cdc25A, Cdc25B, and Cdc25C are key substrates of the DNA damage effector kinases. Chk1 phosphorylation of Cdc25A leads to its degradation and G1 arrest, while phosphorylation by Chk1/2 of Cdc25B/C leads to their inhibition by 14-3-3 proteins and G2 arrest.

In addition, stoichiometric inhibitors can play a crucial role in cell cycle regulation. The transcriptional induction of these stoichiometric cell cycle inhibitors, as well as direct transcriptional repression of cyclins, constitutes a slower but potentially longer-lived response to DNA damage (Fig. 1b). For instance, p53 can induce expression of the CDK inhibitor p21 and the Cdc25C inhibitor 14-3-3σ (▶ CDK Inhibitors). p53 also plays a role in the transcriptional repression of cyclins A and B.

It should be noted that not all of these mechanisms described in this section have been shown to be active in all cell types, or in response to all stimuli. We still lack a comprehensive picture of which combinations of arrest mechanisms control an individual cell's response to damage. However, from those few studies addressing combinations of arrest mechanisms in the same cells, a key motif emerges: that of fast (post-translational) p53-independent arrest initiation acting in parallel with slower (transcriptional) p53-dependent arrest maintenance. For instance, in G1 arrest induced by ionizing radiation, cyclin D is promptly ubiquitinated in a p53-independent fashion, followed by the slower induction of the p53 transcriptional target p21 (Agami and Bernards 2000). Similarly, G2 arrest is induced quickly by Cdc25C phosphorylation by Chk2, followed by p53-dependent downregulation of cyclins A and B (Toettcher et al. 2009).

Understanding Cell Cycle Arrest via Computation

Differential equation models of the cell cycle typically undergo autonomous oscillation under normal conditions (▶ Cell Cycle Modeling, Differential Equation). Cell cycle arrest arises naturally from such a model, when some parameter or modeled species is modified such that oscillation ceases and a stable steady state is reached. In contrast to the checkpoint-centered view in which the cell cycle "checks" the status of damage signaling before committing to S or M phase, this dynamical systems-centered view holds that damage signaling modifies the cell cycle oscillator such that it settles to a stable, arrested state. Upon removal of the arrest stimulus, the arrested state may either persist or oscillation may resume. Notably, there is no guarantee that cycling continues from the same phase at which it arrested.

Modeling approaches have the advantage that cell cycle arrest mechanisms can be tested individually – or in combination – for their ability to promptly elicit arrest, maintain arrest, and resume cycling after damage repair. An attempt to characterize all 24 possible arrest mechanisms in a recent model of the eukaryotic cell cycle (Csikász-Nagy et al. 2006) identified five classes of arrest states corresponding to G1, S, G2, and M phase arrest, as well as a G1-like state with G2 DNA content (Toettcher et al. 2009) (Table 1). Such models can also be fit to experimental data to determine which classes of arrest mechanisms govern the response in vivo. These experimentally driven approaches are complementary to theoretical investigations of models' oscillation and steady state behavior using techniques such as bifurcation analysis (▶ Cell Cycle Model Analysis, Bifurcation Theory).

The modeler faces two challenges in developing quantitative models of DNA damage and the cell cycle, discussed in this and the following paragraph. First, it is crucial to inform a model by fitting its parameters to account for a number of experimental observations. Typical data might include the amount time spent by freely cycling cells in each cell cycle phase, whether certain cell cycle mutants are able to cycle, and how protein levels change during arrest. Fitting this data accurately and efficiently requires computing the sensitivity of protein levels and the times of cell cycle transitions from a model (▶ Cell Cycle Models, Sensitivity Analysis). While it is straightforward to compute the sensitivity of protein concentrations at fixed timepoints, it is more challenging to compute the sensitivity of the *times* at which cell cycle transitions occur. If the time of cell cycle transition can be associated with a local maximum or minimum of the ith species concentration (e.g., cyclin E/Cdk2 to mark the G1/S transition, or cyclin A/Cdk1 to mark S/G2), the sensitivity of this time t^* to model parameters $\mathbf{p}$ can be represented as

$$\frac{dt^*}{d\mathbf{p}} = \frac{\nabla_{\mathbf{x}}\mathbf{f_i}}{(\nabla_{\mathbf{x}}\mathbf{f_i})\cdot\mathbf{f}}\frac{\partial \mathbf{x}}{\partial \mathbf{p}} - \frac{\nabla_{\mathbf{p}}\mathbf{f_i}}{(\nabla_{\mathbf{x}}\mathbf{f_i})\cdot\mathbf{f}}$$

where $\frac{\partial \mathbf{x}}{\partial \mathbf{p}}$ is the standard concentration sensitivity of state variables $\mathbf{x}$ with respect to model parameters (Rand et al. 2006). This quantitative approach enables the generation of models that match a variety of experimental observations.

Second, the modeler must be able to generate experimentally relevant predictions. Many experimental studies are based on measurements from large numbers of cells. These might be either population-averaged techniques such as Western blots, or population distribution measurements such as flow cytometry (▶ Cell Cycle Analysis, Flow Cytometry). To afford comparison to data, populations of models must be simulated representing collections of cells. However, there is a subtlety associated with properly sampling populations of dividing cells: since a freely cycling cell population contains twice as many newborn cells as cells that are just about to divide, the distribution of cell ages in the population is skewed toward younger cells. Generally, the probability of finding a cell at age t since its last division can be represented by the relation

$$p(t) = (2\ln 2)2^{-t/T}$$

where T is the cell cycle length (Mitchison 1971). Thus, the modeler must sample the model's initial states according to this distribution. A corollary to this observation is that the fraction of freely cycling cells observed in G1, S, and G2 (e.g., by flow cytometry) *is not equal* to the amount of time spent by a cell in those phases, but must be corrected accordingly.

Consequences of Cell Cycle Arrest

Cell cycle arrest is typically considered to be a temporary and reversible state, persisting until damage is repaired, after which cycling is resumed. This is

Cell Cycle Arrest After DNA Damage, Table 1 Computational analysis of individual arrest mechanisms affecting eight key cell cycle species by stoichiometric inhibition, enzymatic inactivation, or transcriptional repression. Entries are grouped first by the arrest state induced by the each mechanism, and second by the species inhibited

Arrest type	Inhibited species	Stoichiometric	Enzymatic	Transcriptional
G1	Cyclin D	✓	✓	✓
	Cyclin E	✓	✓	✓
S	Cdc25	✓	✓	✓
	Cyclin B	✓	✓	
G2	E2F	✓	✓	✓
	Cyclin A			✓
M	APC^{Cdh1}	✓	✓	✓
	Cdc20	✓	✓	✓
G1 cyclins; 8N DNA	Cyclin A	✓	✓	
	Cyclin B			✓

✓ Predicted to cause cell cycle arrest

indeed in many contexts, especially for low levels of damage where repair is concluded quickly. However, two alternatives to this course of action can be observed: adaptation, in which cells resume cycling even in the presence of damage; and senescence, where cells remain arrested permanently, even after damage is repaired. Adaptation plays a role in bacteria and yeast, but is largely detrimental to multicellular organisms due to the risk of mutagenesis it carries. In contrast, senescence is the safest course of action, provided nearby undamaged cells can compensate for the loss of the damaged cell's replicative capacity.

A growing body of work suggests a connection between senescence and DNA damage (Toettcher et al. 2009; Bartkova et al. 2006; Di Micco et al. 2006). A senescence-like arrest state can result from DSB-inducing irradiation (Toettcher et al. 2009), and the DSB response is required for oncogene-induced senescence (▶ Cell Cycle, Cancer Cell Cycle and Oncogene Addiction) (Bartkova et al. 2006; Di Micco et al. 2006). During this response, even G2-arrested cells appear to enter a G1-like arrest state, with high levels of cyclin E and p21, and low levels of cyclins A and B. When such an arrest state is disrupted (e.g., by deletion of p21) cells can undergo a second round of DNA replication, a phenomenon known as endoreduplication (▶ Endoreplication).

Cross-References

- ▶ Cell Cycle Checkpoints
- ▶ Cell Cycle Model Analysis, Bifurcation Theory
- ▶ Cell Cycle Modeling, Differential Equation
- ▶ Cell Cycle Models, Sensitivity Analysis
- ▶ Cell Cycle of Mammalian Cells
- ▶ Cell Cycle, Cancer Cell Cycle and Oncogene Addiction
- ▶ DNA Repair
- ▶ Endoreplication

References

Agami R, Bernards R (2000) Distinct initiation and maintenance mechanisms cooperate to induce G1 cell cycle arrest in response to DNA damage. Cell 102:55–66

Bartkova J, Rezaei N, Liontos M, Karakaidos P, Kletsas D, Issaeva N, Vassiliou LV, Kolettas E, Niforou K, Zoumpourlis VC, Takaoka M, Nakagawa H, Tort F, Fugger K, Johansson F, Sehested M, Andersen CL, Dyrskjot L, Orntoft T, Lukas J, Kittas C, Helleday T, Halazonetis TD, Bartek J, Gorgoulis VG (2006) Oncogene-induced senescence is part of the tumorigenesis barrier imposed by DNA damage checkpoints. Nature 444:633–637

Batchelor E, Mock CS, Bhan I, Loewer A, Lahav G (2008) Recurrent initiation: a mechanism for triggering p53 pulses in response to DNA damage. Mol Cell 30:277–289

Batchelor E, Loewer A, Lahav G (2009) The ups and downs of p53: understanding protein dynamics in single cells. Nat Rev Cancer 9:371–377

Csikász-Nagy A, Battogtokh D, Chen KC, Novák B, Tyson JJ (2006) Analysis of a generic model of eukaryotic cell-cycle regulation. Biophys J 90:4361–4379

Di Micco R, Fumagalli M, Cicalese A, Piccinin S, Gasparini P, Luise C, Schurra C, Garre M, Nuciforo PG, Bensimon A, Maestro R, Pelicci PG, d'Adda di Fagagna F (2006) Oncogene-induced senescence is a DNA damage response triggered by DNA hyperreplication. Nature 444:638–642

Harris SL, Levine AJ (2005) The p53 pathway: positive and negative feedback loops. Oncogene 24:2899–2908

Loewer A, Batchelor E, Gaglia G, Lahav G (2010) Basal dynamics of p53 reveal transcriptionally attenuated pulses in cycling cells. Cell 142:89–100
Mitchison JM (1971) The biology of the cell cycle. Cambridge University Press, Cambridge
Rand DA, Shulgin BV, Salazar JD, Millar AJ (2006) Uncovering the design principles of circadian clocks: mathematical analysis of flexibility and evolutionary goals. J Theor Biol 238:616–635
Reinhardt HC, Yaffe MB (2009) Kinases that control the cell cycle in response to DNA damage: Chk1, Chk2, and MK2. Curr Opin Cell Biol 21:245–255
Toettcher JE, Loewer A, Ostheimer GJ, Yaffe MB, Tidor B, Lahav G (2009) Distinct mechanisms act in concert to mediate cell cycle arrest. Proc Natl Acad Sci USA 106:785–790

Cell Cycle Arrest Mechanisms

▶ Cell Cycle Arrest After DNA Damage

Cell Cycle Checkpoints

Flavio Amara, Marco Muzi-Falconi and Paolo Plevani
Dipartimento di Scienze Biomolecolari
e Biotecnologie, Università di Milano, Milan, Italy

Synonyms

Adaptation; Cell cycle; DNA damage checkpoint; Recovery; Spindle checkpoint.

Definition

In this entry we describe the molecular mechanisms of the cell cycle checkpoints. We focused our attention on the DNA damage checkpoint and on the spindle assembly and spindle positioning checkpoints, controlling chromosome segregation in mitosis.

Characteristics

Intrinsic and Extrinsic Cell Cycle Checkpoints

The checkpoints are evolutionarily conserved surveillance mechanisms controlling the order and timing of cell cycle transitions. They are organized as signal transduction cascades blocking or slowing down cell cycle progression at specific stages. Checkpoints are triggered by sensor proteins detecting, directly or indirectly, cell cycle perturbations and transmitting the signal, through the action of protein kinases, to effector proteins that stop cell cycle progression until the signal activating the checkpoint has been turned off. These mechanisms have been highly conserved during evolution, and checkpoint defects result in genome instability, which is frequently associated to tumor development. The checkpoint controls are elicited through molecular events regulating their activation, maintenance, and inactivation resulting, respectively, in cell cycle arrest, maintenance of the arrest for a certain time and recovery of cell cycle progression.

These surveillance mechanisms can be divided into intrinsic regulatory pathways, ensuring the orderly progression of cell cycle events under physiological conditions, and extrinsic pathways that are activated in response to specific clues, such as damage to DNA or cellular structures. The intrinsic checkpoints act by controlling the activity of cell cycle dependent kinases (CDKs) mainly at the G1/S boundary and at the metaphase to anaphase transition in mitosis; such mechanisms are described in other entries of the encyclopedia.

DNA Damage Checkpoints

The DNA damage checkpoint is required for the efficient response to genotoxic stress. The checkpoint is activated when lesions in the DNA are detected and the mechanisms involved differ slightly at various cell cycle phases. DNA damage during the G1 phase activates the G1/S checkpoint preventing entry into S phase. The presence of DNA lesions while cells replicate their genome slows down the kinetics of DNA replication (intra-S checkpoint), and if the chromosomes are damaged in G2, the activation of the G2/M checkpoint avoids chromosome segregation before repair.

Precise and complete DNA replication in every cell cycle and repair of DNA lesions are critical for the maintenance of genetic stability; failures in these processes reduce cell survival and lead to cancer susceptibility. Cell cycle arrest is not the only final outcome of the DNA damage checkpoint response; indeed, it has been demonstrated that checkpoint activation regulates the choice of recombination pathways, influences transcription of DNA repair genes, stabilizes

Cell Cycle Checkpoints, Table 1 Major orthologous checkpoint proteins in budding yeast and mammalian cells

S.cerevisiae	Mammal	Function
RFA 1, 2, 3	RPA 1, 2, 3	ssDNA binding protein
MEC1/DDC2	ATR/ATRIP	Sensor (PI3K kinase)
TEL1	ATM	Sensor (PI3K kinase)
DDC1-RAD17-MEC3	RAD9-RAD1-HUS1	Sensor (9-1-1 PCNA-like clamp)
DPB11	TOPBP1	Relevant for MEC1/ATR activation
RAD9	BRCA1/53BP1	Adaptor
MRE11-RAD50-XRS2 (MRX)	MRE11-RAD50-NBS1 (MRN)	Lesion processing
SAE2	CtIP	Lesion processing
SGS1	BLM, WRN	Lesion processing
EXO1	hEXO1	Lesion processing
CHK1	CHK1	Effector kinase (S/T kinase)
RAD53	CHK2	Effector kinase(S/T kinase)

stalled replication forks and, in multicellular eukaryotes, it may promote apoptosis when the damage is irreparable.

MEC1/ATR DNA Damage Checkpoint Signaling

In *Saccharomyces cerevisiae*, the two main players of the DNA damage checkpoint activated in response to many DNA helix-distorting lesions are two kinases encoded by the *MEC1* and *RAD53* genes. They are the orthologues of ATR and CHK2 in human cells (Table 1).

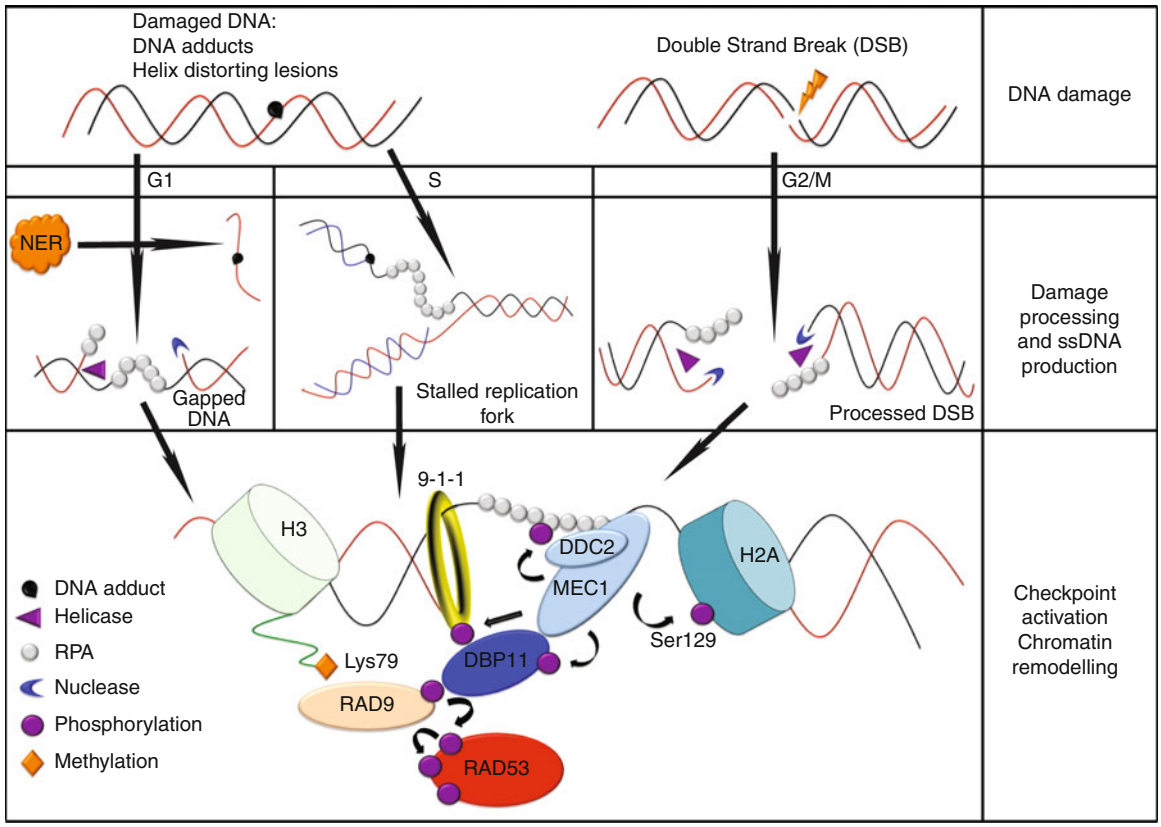
In budding yeast, MEC1 acts as the apical kinase in the checkpoint cascade and likely senses abnormal amounts of single-stranded (ss) DNA stretches covered by the replication protein A (RPA), which are generated by processing of various DNA lesions (Fig. 1).

RPA-coated ssDNA is thought to recruit the apical checkpoint kinase (MEC1-DDC2 or TEL1, in budding yeast; ATR-ATRIP or ATM in mammalian cells), and the 9-1-1 complex (RAD17-MEC3-DDC1 in budding yeast) triggering the signal transduction cascade (Zou and Elledge 2003). As a consequence of this activation, MEC1 phosphorylates, directly or indirectly, a number of factors (e.g. DDC2, H2A, DDC1, RAD9, DPB11/TOPB1, RAD53/CHK2, CHK1, etc.) and these subsequent phosphorylation events are used to follow the signal through the cascade (Fig. 1). Activated RAD53/CHK2 amplifies the signal through autophosphorylation events and inhibits the cell cycle machinery by phosphorylating various targets, leading to cell cycle arrest (Harrison and Haber 2006). RAD53/CHK2 and CHK1 activation influence various DNA metabolic pathways, like homologous recombination, origin firing during S phase, nuclease activity, and gene expression.

Another essential player of checkpoint activation is represented by the 9-1-1 complex. This heterotrimeric complex show limited sequence homology to the PCNA homotrimeric clamp and it is therefore often referred to as the checkpoint clamp (Table 1 and Fig. 1). The 9-1-1 complex is loaded onto DNA by a checkpoint clamp loader, a form of replication factor-C (RFC), in which the RFC1 subunit is substituted by RAD24 (*S. cerevisiae*) or RAD17 (mammalian cells) (Majka and Burgers 2003). The 9-1-1 clamp promotes checkpoint activation *in vivo* by influencing MEC1/ATR recruitment and its substrate specificity; the 9-1-1 clamp has a role in modulating chromatin binding of RAD9/53BP1 and subsequent RAD53/CHK2 phosphorylation events.

Checkpoint Signaling in Response to Double Strand Breaks (DSBs)

When the DNA integrity is challenged by discontinuities in the helix backbone, as those generated by double strand breaks, the checkpoint human kinase ATM is loaded near the DSBs and, together with ATR, activate the checkpoint response. ATM activation requires the presence of the MRN complex, which has a DNA tethering capacity and possesses both endo- and exo-nucleolytic activities. Although the nuclease activity of the MRX/MRN complex is not critical for checkpoint activation, the presence of a physically



Cell Cycle Checkpoints, Fig. 1 DNA damage checkpoint activation

assembled complex and the action of other nucleases and helicases (such as EXO1 and SGS1) are important.

In *S. cerevisiae* TEL1, the ATM orthologue, participates only marginally in the checkpoint response induced by a single DSB, but its role becomes relevant in the presence of multiple DSBs and/or when the initiation of DSB ends processing is defective. In this organism, a single irreparable DSB triggers the *MEC1* pathway, which is activated by extensive resection of the DSB DNA ends (Lazzaro et al. 2009). If the break is not rapidly repaired, nucleolytic activities produce long ssDNA regions that, via the ssDNA binding protein RPA, recruit MEC1 activating the checkpoint cascade (Fig. 1).

Chromatin Remodeling and Checkpoint Maintenance

Chromatin modifications contribute to checkpoint function. In fact, histone proteins are also modified in response to DNA damage. In *S. cerevisiae*, the Ser129 residue of H2A is phosphorylated in a MEC1-dependent manner in response to a variety of genotoxic treatments and the modified γH2A molecules localize near a DSB (van Attikum and Gasser 2009); the same modification occurs on the Ser139 conserved residue of the mammalian H2AX histone variant. It is assumed that RPA-coated ssDNA is required to activate the checkpoint, but the MEC1/ATR and TEL1/ATM-dependent phosphorylation of the H2A/H2AX histones (Fig. 1) is relevant for checkpoint maintenance, since in its absence the checkpoint signal is turned off prematurely. In yeast, one of the major roles of H2A phosphorylation is the recruitment of chromatin remodeling complexes near the DNA lesions. Another histone modification important for the activation of the DNA damage checkpoint is methylation of Lys79 of histone H3 (H3-Lys79Me) by the specific methyl transferase DOT1. In fact, the main mediator proteins in the checkpoint cascade (RAD9 in *S. cerevisiae* and 53BP1 in mammals) can be recruited near a chromatin damaged site through two parallel pathways. The first one is dependent upon H3-Lys79 methylation, while the second pathway acts through

the function of the DPB11/TOPB1 adaptor proteins (Fig. 1).

Turning Off the DNA Damage Checkpoint Signal

In cells with repairable DNA damage, the checkpoint arrest is maintained until repair is completed. Then, cells can resume the cell cycle turning off the checkpoint signal in a process known as recovery. Instead, if an irreparable lesion is present, cells eventually override the cell cycle arrest in a process called adaptation. In mammals, it has been proposed that adaptation to damage is linked to cancerogenesis and, likely, this event is normally prevented by inducing the apoptotic pathway. Inactivation of the RAD53 kinase is required to recover from checkpoint-mediated cell cycle arrest in *S. cerevisiae*. Various protein phosphatases, including PTC2, PTC3, PPH3, and GLC7, have been found to be important for checkpoint inactivation. Interestingly, their relative contribution to the recovery process seems to be influenced by the type of genotoxic stress causing checkpoint activation (DSBs, alkylating agents, hydroxyurea). The same phosphatases act on phosphorylated γH2A and, not surprisingly, homologous phosphatases influence checkpoint inactivation also in human cells.

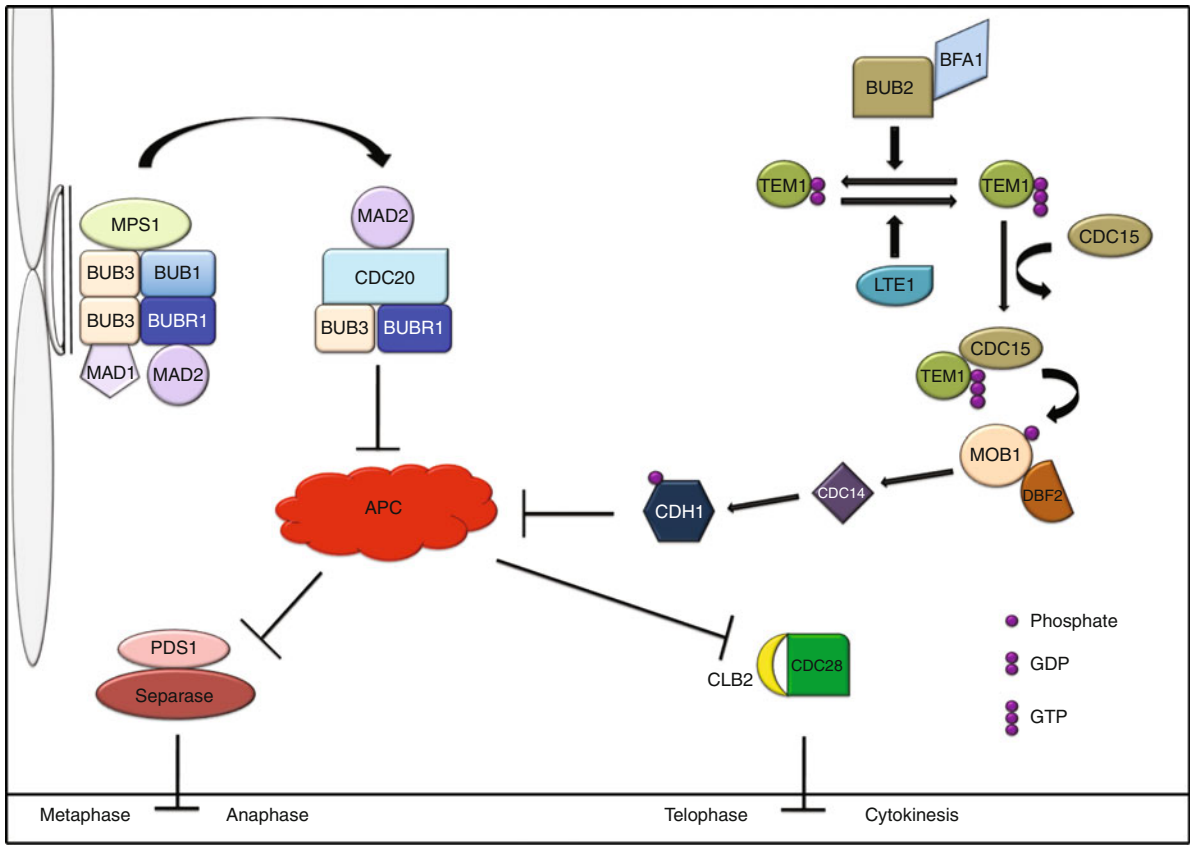
While it was somehow expected that reversal of checkpoint signaling would involve the action of phosphatases acting on the master effector kinases of the cascade, it was interesting to learn that POLO-like kinases (such as PLK1 and CDC5) also play an important role in switching off the checkpoint. It is likely that recovery and adaptation, being two sides of the same coin, are going to share common players and mechanisms. However, genetic screenings in budding yeast revealed that some mutations allow to distinguish between the two processes. For example, the *cdc5-ad* allele and deletions of certain recombination genes (such as *KU70/80, TID1* and *RAD51*) prevent the adaptation process without affecting recovery.

S. cerevisiae Spindle Assembly and Spindle Position Checkpoints

Chromosome segregation at mitosis is controlled by two surveillance mechanisms: the spindle assembly and the spindle positioning checkpoints. Accurate segregation requires bipolar attachment of sister chromatids to the mitotic spindle, which is mediated by a proper connection between kinetochores and spindle microtubules. Kinetochore capture and microtubules bi-orientation are stochastic processes taking a variable amount of time to complete. During that time individual chromosomes may be detached from the microtubules or be connected only to one spindle pole. The spindle assembly checkpoint (SAC) delays the metaphase to anaphase transition until the sister chromatids are properly attached to the spindle in a bipolar orientation. In budding yeast, cell cycle arrest at the G2/M transition is mediated by inhibition of the CDC20-anaphase promoting complex (APC) ubiquitin ligase, thus preventing proteolysis of the securin PDS1 until complete bi-orientation is achieved (Fig. 2).

Checkpoint proteins (MAD1, MAD2, BUB1, BUBR1, BUB3, and MPS1) all accumulate at unattached kinetochores and form various complexes, many of which can inhibit the APC (Fang et al. 1998) (Fig. 2). APC is a multiprotein complex that targets several proteins for degradation during mitosis through the associated specificity factor CDC20. Securin and cyclins are ubiquitylated by CDC20-APC; therefore, to delay anaphase onset in the presence of spindle defects, the checkpoint must block CDC20-APC mediated PDS1 degradation. Experimental evidence suggests that in response to spindle defects, MAD2 exchanges from a MAD1/MAD2 complex to a CDC20/MAD2 complex sequestering CDC20 away from the APC and blocking PDS1 degradation (Fig. 2).

In *S. cerevisiae*, spindle misorientation is detected by the spindle positioning checkpoint (SPOC) which prevents mitotic exit. The target of this control is the mitosis exit network (MEN), and more specifically the activation of the TEM1 GTPase (Adames et al. 2001). TEM1 cycles between GDP- and GTP-bound states, regulated by the putative guanine nucleotide exchange factor (GEF) LTE1 and the two-component GTPase activating protein (GAP) BFA1/BUB2. The last one recruits TEM1 to the bud-directed spindle pole, where TEM1 is kept inactive until the pole crosses the neck into the bud. GTP-TEM1 then binds to the protein kinase CDC15, which phosphorylates and activates the protein kinase DBF2. MOB1 binds to DBF2 and, in a poorly understood manner, MOB1/DBF2 stimulates the release of the CDC14 phosphatase from the nucleolus and contributes to cytokinesis. CDC14 dephosphorylates CDH1, leading to the activation of CDH1-APC complex, which triggers cyclin degradation and exit from mitosis. A possible crosstalk between the SAC and SPOC is likely, but not yet fully demonstrated (Lew and Burke 2003).



Cell Cycle Checkpoints, Fig. 2 The spindle assembly and spindle positioning checkpoints

Summary

The maintenance of genome stability is an essential process, which needs a careful control. Indeed, all eukaryotic cells evolved surveillance mechanisms, called checkpoints, sensing the presence of DNA damage in the genome or alterations in cellular structures controlling chromosome segregation in mitosis. DNA integrity can be challenged by lesions caused by a variety of chimico-physical agents, or by replication stress caused by special DNA structures or by a limited supply of deoxyribonucleotides. The DNA integrity checkpoint is a signal transduction cascade conserved from yeast to man and the apical factors in the pathway are protein kinases. The spindle assembly checkpoint controls the status of kinetochore/microtubule attachment delaying exit from mitosis until all kinetochores have formed correct bipolar connections with the spindle. In budding yeast, and likely in all eukaryotes, the alignment of the mitotic spindle with the axis of cell polarity is controlled by another surveillance mechanism, called the spindle positioning checkpoint.

Cross-References

- DNA Damage
- DNA Repair
- Histones
- Kinetochores
- Mitosis
- Protein Kinases
- Phosphatases

References

Adames NR, Oberle JR, Cooper JA (2001) The surveillance mechanism of the spindle position checkpoint in yeast. J Cell Biol 153(1):159–168

Fang G, Yu H, Kirschner MW (1998) Direct binding of CDC20 protein family members activates the anaphase-promoting complex in mitosis and G1. Mol Cell 2(2):163–171
Harrison JC, Haber JE (2006) Surviving the breakup: the DNA damage checkpoint. Annu Rev Genet 40:209–235
Lazzaro F, Giannattasio M, Puddu F, Granata M, Pellicioli A, Plevani P, Muzi-Falconi M (2009) Checkpoint mechanisms at the intersection between DNA damage and repair. DNA Repair 8:1055–1067
Lew DJ, Burke DJ (2003) The spindle assembly and spindle position checkpoints. Annu Rev Genet 37:251–282
Majka J, Burgers PM (2003) Yeast Rad17/Mec3/Ddc1: a sliding clamp for the DNA damage checkpoint. Proc Natl Acad Sci USA 100:2249–2254
van Attikum H, Gasser SM (2009) Crosstalk between histone modifications during the DNA damage response. Trends Cell Biol 19:207–217
Zou L, Elledge SJ (2003) Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes. Science 300:1542–1548

Cell Cycle Data Analysis

Lars Juhl Jensen
Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, Copenhagen, Denmark

Definition

Cell cycle data analysis is an umbrella term used to describe a variety of related methods for analyzing large datasets related to the cell cycle. These data typically come in the form of time courses a high-throughput assay, for example, DNA microarrays, are applied to samples from multiple time points during one or more cell cycles. Key aspects of cell-cycle analysis include data quality assessment, identification of cell-cycle-regulated genes and proteins, integrative analysis of different data types, and evolutionary comparisons across species.

Characteristics

Like in many other parts of molecular biology, the advent of high-throughput assays has had a major impact on cell-cycle research. ▶ DNA microarray technologies have been particularly influential, as they enabled systematic, genome-wide studies of ▶ gene expression through the cell cycle of a ▶ model organism (▶ Cell Cycle Analysis, Expression Profiling). Today several such studies have been performed, measuring the expression of every gene at many time points during the cell cycles of budding yeast (▶ Cell Cycle, Budding Yeast), fission yeast (▶ Cell Cycle, Fission Yeast), and cell lines of human (▶ Cell Cycle of Mammalian Cells) or plant origin. Other large-scale studies include mass spectrometry data on protein levels and phosphorylation states in different phases of the cell cycle and systematic screens for substrates of cyclin-dependent kinases (▶ Cyclins and Cyclin-dependent Kinases).

The sheer size of these types of datasets implies that computational methods are needed to analyze them. Because most of the datasets are expression time courses covering one more cell cycles, a central problem is to identify cell-cycle-regulated (cycling) genes or proteins based on expression profiles. This is a nontrivial problem because the data may contain considerable amounts of noise and because they may contain other biological signals than the one of interest (e.g., stress response caused by the synchronization of cell cultures). Over the past few years numerous methods for identification of cycling genes have been developed; a brief description and a ▶ receiver operating characteristic (ROC) curve can be found for most of them in Gauthier et al. (2010). Key features of all the best performing approaches are that they integrate all available data, that they search specifically for a ▶ periodic oscillation consistent with the length of the cell cycle, and that they take into account the ▶ oscillation amplitude (as opposed to only the period).

The availability of genome-wide expression data from multiple model organisms also allows for analyses of the evolution of cell-cycle regulation. Once a set of cycling genes or proteins has been identified for each of several organisms, cross-species comparisons can be performed (de Lichtenberg et al. 2008). Mapping the expression data onto ▶ functional modules and complexes can give a higher-level view of cell-cycle regulation that is not obvious from analyses performed at the level of individual genes or proteins (de Lichtenberg et al. 2008). Finally, correlations can be discovered between different layers of cell-cycle regulation – for example, transcriptional and post-translational regulation – by using ▶ Fisher's exact test to compare

the gene and protein lists obtained through analysis of different types of high-throughput data (de Lichtenberg et al. 2008).

Cross-References

- ▶ Cell Cycle Analysis, Expression Profiling
- ▶ Cell Cycle of Mammalian Cells
- ▶ Cell Cycle, Budding Yeast
- ▶ Cell Cycle, Fission Yeast
- ▶ Cyclins and Cyclin-dependent Kinases
- ▶ DNA Microarrays
- ▶ Fisher's Test
- ▶ Functional Modules and Complexes
- ▶ Gene Expression
- ▶ Model Organism
- ▶ Oscillation Amplitude
- ▶ Periodic Oscillation
- ▶ Receiver Operating Characteristic (ROC) Curve

References

de Lichtenberg U, Jensen TS, Brunak S, Bork P, Jensen LJ (2008) Evolution of cell cycle control: same molecular machines, different regulation. Cell Cycle 6:1819–1825

Gauthier NP, Jensen LJ, Wernersson R, Brunak S, Jensen TS (2010) Cyclebase.org: version 2.0, an updated comprehensive, multi-species repository of cell cycle experiments and derived analysis results. Nucleic Acids Res 38:D699–D702

Cell Cycle Database

Roberta Alfieri and Luciano Milanesi
Institute for Biomedical Technologies – CNR (Consiglio Nazionale delle Ricerche), Segrate, Milan, Italy

Synonyms

Cell division database

Definition

The Cell Cycle Database is a biological resource, which collects the most relevant information related to genes and proteins involved in human and yeast cell cycle (see also ▶ Cell Cycle, Biology) processes, one of the most studied complex system in biology (see also ▶ Complex System). The database, which is accessible at the web site http://www.itb.cnr.it/cellcycle, has been developed in a systems biology context, since it also stores the cell cycle mathematical models (see also ▶ Mathematical Model, Model Theory) published in the recent years, with the possibility to simulate them directly from the web interface. Cell cycle information from humans is primary considered since this resource aims to support biomedical studies in the context of cancer research. Then the database content is extended toward the budding yeast cell cycle because of the genomic similarity between human and budding yeast and the large number of models available for this organism. According to this choice, the data integration (see also ▶ Data Integration) concerns all genes and proteins involved in the cell cycle models of both the budding yeast *S. cerevisiae* and the *H. sapiens*. This information is taken from the most recent literature and plays a crucial role to contextualize behavior of each cell cycle component.

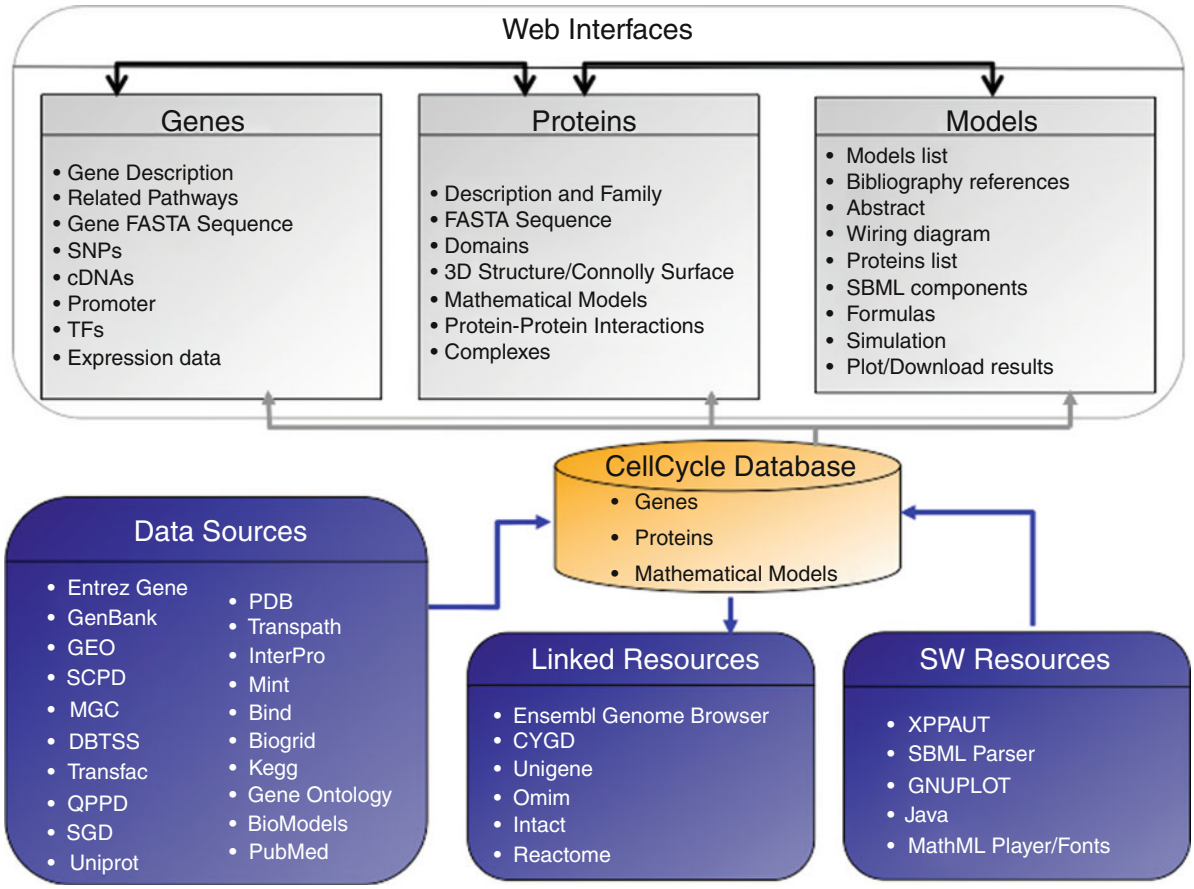
Characteristics

Motivation

The aim of our initiative is to give an exhaustive view of the cell cycle process starting from its building blocks, genes and proteins, arriving to the pathway they create, represented by the models, as summarized in Fig. 1.

The principal goal here is the integration of cell cycle information, which can be useful for researchers in the context of systems biology studies. From the user's point of view, this initiative presents two important features: the first is a data integration system for genes and proteins involved in yeast and human cell cycle processes; the second is a section dedicated to cell cycle models and their mathematical simulation.

The significant information related to cell cycle genes and proteins is a useful annotation of the models' components and facilitates the exploration of the relevant features of the whole network. The structure of this resource allows the storage of new data deriving from *cell cycle models* (▶ Cell Cycle Models, Sensitivity Analysis), due to its particular structure and the pipeline for the automatic data updating: the database



Cell Cycle Database, Fig. 1 Architecture of the Cell Cycle Database. The core of the resource is a data warehouse system, which collects data from the most relevant bioinformatics resources (Data Sources) and provides links to external resources (Linked Resources). The Cell Cycle Database comprises user-friendly interfaces which show information and, in general, provide the functions of the resource, even relying on third party free software

administrator can update the database content by gene or protein name through the web interface. When a new entry is inserted in the core table, all the external tables will be updated in cascade, while when a new entry is inserted in one of the external table no inward updating occurs. As a result all tables of the database are updated according to the infrastructure, which is designed for automated data integration.

This database is able to collect the most important information related to cell cycle genes and proteins, which are drawn from the analysis of the cell cycle information available in literature and the existing pathway databases. The database also contains protein-protein interactions taken from several resources making the information on the cell cycle interaction network as complete as possible.

Furthermore, a repository of the most recent published cell cycle models to allow the exploration of their mathematical structure through SBML (Systems Biology Markup Language) (Hucka et al. 2003) components and their mathematical simulation is available.

Functionality

The database stores 26 models of the cell cycle interaction networks and it allows the mathematical simulation over time of the quantitative behavior of each model component. To accomplish this task, a web interface for browsing information related to cell cycle genes, proteins, and mathematical models is available.

Models and proteins information are directly linked in the resource: for each protein we provide

information on the models in which it is involved. In the protein report a list of the published models is available with a direct link to the specific model report.

In this framework a pipeline, which allows users to deal with the mathematical part of the models, in order to solve, using different variables, the ordinary differential equation systems that describe the biological process, has been implemented. Up to now among the 26 models stored in the database, 13 models have the related SBML file and for 12 of them it is possible to run simulations.

The flexibility of this database allows the addition of mathematical data, which are used for simulating the behavior of the cell cycle components in the different models. The resource is useful both to retrieve information about cell cycle model components and to analyze their dynamical properties. The Cell Cycle Database can be used to find system-level properties, such as stable steady states and oscillations, by coupling structure and dynamical information about models.

Accessibility

The web interface allows to retrieve model-related information and a pipeline has been implemented to deal with the mathematical part of the models, in order to solve the ordinary differential equations systems that describe the biological processes. This pipeline essentially consists in the XHTML-MML-XSL style sheet implementation, as it will be discussed below. In fact, using this system it is possible to visualize the mathematical description of the model and to run simulations by introducing modifications of the initial conditions of state variables and parameters.

Each model is presented in a report structured in three sections: the publication data, the SBML data structure, and the numerical simulation part. The first section contains the detailed publication data, the diagram of the model, the related XML file, and the list of all the proteins involved in the model, which are linked to the related Cell Cycle Database protein report.

In the SBML data structure section, users can explore the SBML components of the selected model including its mathematical expressions. Mathematical formulas within the SBML models are expressed using the Mathematical Markup Language (MML). To visualize the expressions on the web our resource relies on a XHTML + MML page, following the w3c specifications. In this page MML is in-lined with HTML and at the beginning of the page an instruction calls a XSL stylesheet, which allows the formulas to be viewed correctly. This technology allows using the browser functions to find strings inside mathematical expressions and to change their size, operations which are usually not possible using images to represents formulas.

The simulation section allows users to simulate a model using the software XPPAUT (Ermentrout 2002) and to plot results on the fly in order to capture the dynamical properties of the biological process. In order to enable the simulation, this section lists the model species (state variables, typically protein concentrations), its parameters (such as kinetic constants), its algebraic rules, and XPPAUT internal options, using default values. Users can change the initial values in order to analyze the different natures of the system dynamics, performing sensitivity and bifurcations analysis. Once the computation is completed users can download XPPAUT input and output files and plot results. The web interface allows users to plot both time courses and phase diagrams for each model after the simulation step. Results are shown with images exported by GNUPLOT (Williams and Kelley 1998), the popular portable command-line function plotting software.

Model Annotation

As the primary data source we consider the BioModels Database (Le Novere et al. 2006) from which we collect the mathematical model specifically developed for yeast and mammalian cell cycles. We also integrate other published models, which are not stored in the BioModels Database, manually retrieving them from literature or from the CellML repository (Cuellar et al. 2003).

Each model is presented in a report, which is structured in three sections: the publication data, the SBML data structure, and the numerical simulation part. The first section contains the detailed publication data (such as the authors, PubMed ID, the abstract, and journal information), the diagram of the model and the related XML file, if available, and the list of all the proteins involved in the model, which are linked to the related Cell Cycle Database protein report. The protein identifiers are chosen according to the UniProt Database annotation.

More in detail, the Cell Cycle Database contains the literature information related to each model, the input

for the simulation software, and the XML file coded with SBML specifications (Hucka et al. 2003). We choose SBML since it is an internationally supported and widely used language for metabolic networks, cell-signaling pathways, regulatory networks, and many other biological pathways. However, there are published cell cycle models not yet implemented in SBML: for this reason, some SBML models included in the database are manually generated using the JigCell Model Builder software (Vass et al. 2004), a model editor which allows the construction of biochemical reaction networks in SBML format, and are validated using the Systems Biology Workbench SBML validator. Mathematical formulas within the SBML models are expressed using Mathematical Markup Language (MathML or MML) (▶ Systems Biology Markup Language (SBML); ▶ MathML) (Ausbrooks et al. 2003).

Model Curation

As discussed before, the Cell Cycle Database contains models from Biomodels, literature, and authors' websites, where available. Concerning the models from Biomodels, the model curation is ensured from the database reliability, which we trusted in. To ensure the model curation for the literature models, we tested and simulated them in order to reproduce the results shown in the related publication papers.

New Model Additions

The addition of new model in the Cell Cycle Database is manually curated. We periodically monitor both Biomodels and CellML repositories and select the new cell cycle models. Moreover we periodically analyze literature publications in order to discover new mathematical models about the cell cycle which are not stored in the model repositories. New models will undergo the annotation process described before and they will be included in the database.

Cross-References

- ▶ Cell Cycle, Biology
- ▶ Complex System
- ▶ Data Integration
- ▶ Mathematical Model, Model Theory

References

Ausbrooks R et al. W3C Recommendation. Mathematical Markup Language (MathML) version 2.0, 2 edn. [http://www.w3.org/Math/]. Accessed 21 October 2003

Cuellar AA, Lloyd CM, Nielsen PF, Bullivant DP, Nickerson DP, Hunter PJ (2003) An overview of cellML 1.1, a biological model description language. SIMULATION: Trans Soc Model Simul Int 79(12):740–747

Ermentrout B (2002) Simulating, analyzing, and animating dynamical systems: a guide to xppaut for researchers and students, SIAM 2002, Philadelphia

Hucka M et al (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. Bioinformatics 19(4):524–531

Le Novere N, Bornstein B, Broicher A, Courtot M, Donizelli M, Dharuri H, Li L, Sauro H, Schilstra M, Shapiro B, Snoep JL, Hucka M (2006) BioModels database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. Nucleic Acids Res 34: D689–D691

Vass M, Allen N, Shaffer CA, Ramakrishnan N, Watson LT, Tyson JJ (2004) The jigcell model builder and run manager. Bioinformatics 20(18):3680–3681

Williams T, Kelley C (1998) GNUPLOT: an interactive plotting program, version 3.7 organized by: David Denholm

Cell Cycle Dynamics, Bistability and Oscillations

John J. Tyson[1] and Béla Novák[2]
[1]Department of Biological Sciences, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA
[2]Department of Biochemistry, Oxford Centre for Integrative Systems Biology, University of Oxford, Oxford, UK

Synonyms

Multiple steady states; Toggle switch

Definition

A molecular regulatory system can maintain itself indefinitely at a steady state where, for every time-varying species, the total rate of formation of the

chemical is exactly balanced by its total rate of removal. In mathematical terms, we can write

$$\frac{dx_i}{dt} = f_i(x_1, x_2, ..., x_n) - r_i(x_1, x_2, ..., x_n), \quad i = 1, ..., n \tag{1}$$

where x_i is the concentration of species i, and $f_i(x_1, x_2, \ldots, x_n)$ and $r_i(x_1, x_2, \ldots, x_n)$ are its rates of formation and removal. At a steady state, $dx_i/dt = 0$ for all i,

$$f_i(x_1, x_2, ..., x_n) = r_i(x_1, x_2, ..., x_n), i = 1, ..., n. \tag{2}$$

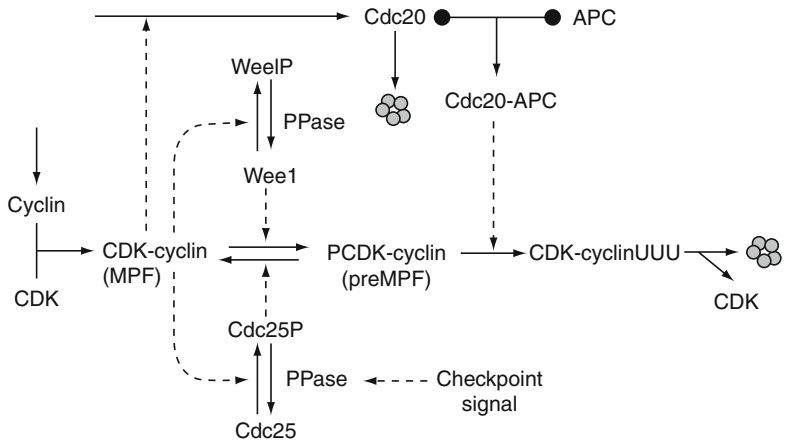
For any well-specified chemical reaction network, there always exists at least one steady state solution; call it $\mathbf{x}^0 = \left(x_1^0, x_2^0, ..., x_n^0\right)$. Such a steady state can be classified as either *stable* or *unstable* with respect to small perturbations away from the steady state. If all small perturbations (in any direction in the n-dimensional state space of the reaction network) eventually return to $\mathbf{x}^0$, then this steady state is said to be (locally asymptotically) stable. If there exist small perturbations (in some particular directions in state space) that depart from $\mathbf{x}^0$ and never return, then the steady state is said to be unstable.

Since the rates of formation and removal of any chemical species in a reaction network are algebraic functions of species concentrations, Eq. 2 is a system of n nonlinear algebraic equations, which in general may have multiple solutions in the positive orthant: $x_i \geq 0$ for all i. In this case, we can denote the different steady states as $\mathbf{x}^0{}_j$, where $j = 1, \ldots, m$. A typical situation is the case of three steady states ($m = 3$), where two of the steady states are stable and one is unstable. This case is called *bistability*. In general, there may be more than two stable steady states, in which case we refer to *tristability* or *multistability*. We may describe a system with $m > 1$ as having *multiple steady states* when we are not sure of the total number of stable steady states.

Bistability is a typical feature of reaction networks with positive feedback, although one must be aware that positive feedback is not always readily apparent in reaction networks as they are conventionally drawn. Double-negative feedback is a particularly common motif for bistability: X_1 inhibits X_2, and X_2 inhibits X_1. In this case, the two stable steady states are (X_1 active, X_2 inactive) and (X_1 inactive, X_2 active). The intermediate steady state (X_1 semi-active, X_2 semi-active) is unstable.

Multistability plays an important role in the theory of cell cycle progression. It is proposed that the characteristic states of cell cycle arrest (G1-arrest, G2-arrest, metaphase-arrest) correspond to alternative stable steady states of the underlying chemical reaction network controlling the activities of cyclin-dependent kinases (CDKs). In this view, cells progress through the cell cycle (G1 $\rightarrow$ S/G2 $\rightarrow$ M $\rightarrow$ G1 $\rightarrow$ ...) by irreversible transitions from the G1 steady state to the S/G2 steady state (the G1-S transition, also called "Start" or the "Restriction Point"), from the S/G2 steady state to the M steady state (the G2-M transition), and from the M steady state to the G1 steady state (the metaphase-anaphase transition, also called "exit from mitosis"). In this view, cell cycle checkpoints work by stabilizing one of these three steady states and preventing the transition to the next phase of the cell cycle.

Bistability is intimately connected to the existence of spontaneous limit cycle oscillations in regulatory networks with both positive feedback (to create alternative stable steady states) and negative feedback (to induce spontaneous switching between the two stable steady states). This combination of positive and negative feedbacks is precisely the case in the regulatory system governing the eukaryotic cell cycle (Tyson and Novak 2008). In somatic cells, checkpoint signals prevent spontaneous cycling, but in some circumstances these checkpoints are removed, and the cell cycle proceeds as a spontaneous, unfettered, limit cycle oscillation. For example, during early embryogenesis, the fertilized egg undergoes a series of rapid cell divisions without growth, until it reaches the "mid-blastula transition," when checkpoint proteins are expressed and the cell cycle regains the characteristic G1-S, G2-M, and M-G1 transitions of somatic cells. It is also possible to create mutant yeast cells that lack checkpoint controls; these cells divide faster than they grow, getting smaller and smaller each cycle until they die. These observations suggest that, under most circumstances, periodic cell divisions are governed not by spontaneous limit cycle oscillations but by multistability and irreversible transitions (Tyson and Novak 2008).



Cell Cycle Dynamics, Bistability and Oscillations, Fig. 1 CDK regulatory network for frog egg cell cycles. In this network diagram, a solid arrow represents a biochemical reaction, and a dashed arrow represents a catalytic effect on a reaction. A T-shaped arrow represents the association of two proteins to form a complex. Black balls on the crossbar indicate a reversible binding reaction. The small gray balls represent proteolytic fragments of a protein. Newly synthesized cyclin B combines with CDK subunits (in excess) to form active CDK-cyclin dimers (called MPF). The CDK subunit is phosphorylated by Wee1 to form a less active form of MPF, called preMPF. The inhibitory phosphate group can be removed by Cdc25. Wee1 and Cdc25 activities are also regulated by phosphorylation, catalyzed by MPF: Wee1P is the less active form, and Cdc25P is the more active form. Cyclin B is labeled for proteolysis by an E3 ubiquitin ligase, the APC, which works in collaboration with a targeting subunit, Cdc20. The synthesis of Cdc20 is activated by MPF

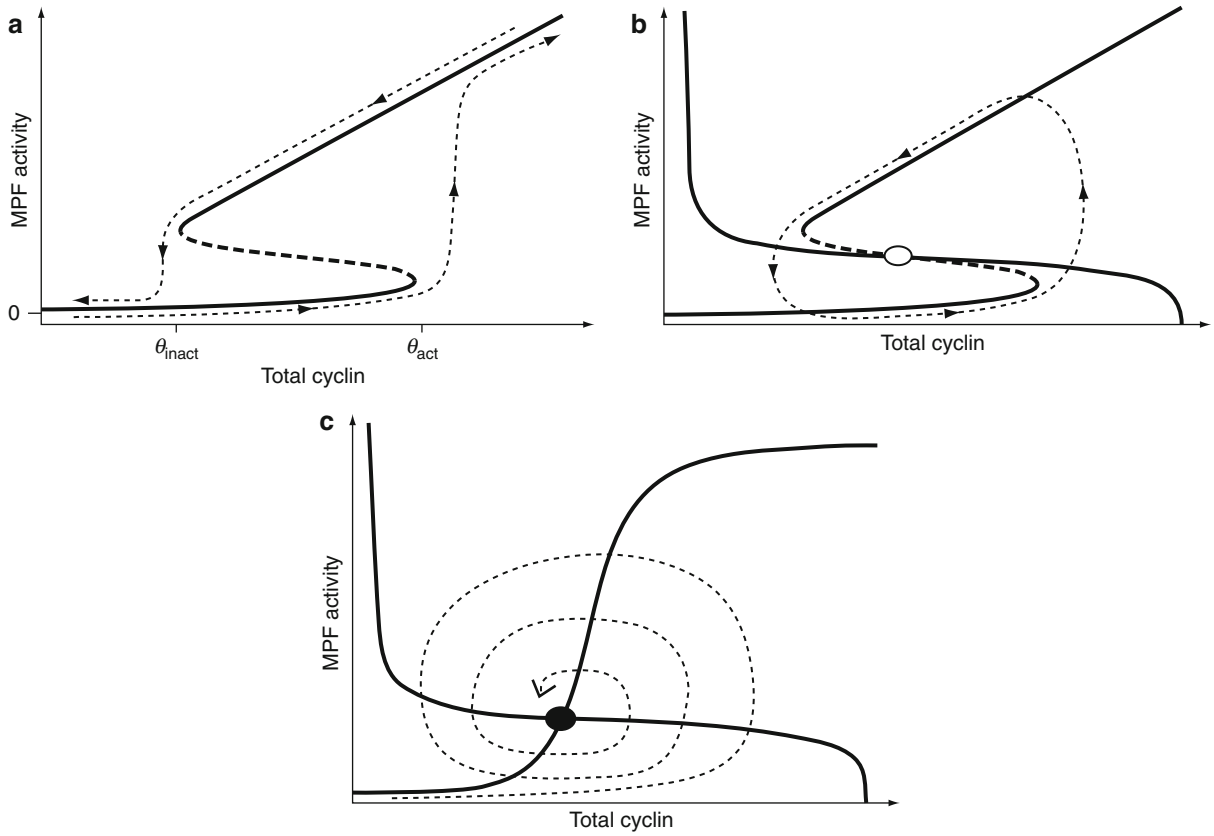
Characteristics

History

In an influential review of cell cycle regulation, Murray and Kirschner (1989) asked whether progression through the cell cycle is more like "dominoes" (a dependent sequence of events: one falling domino pushing over the next) or a "clock" (an autonomous oscillation, ticking along independent of the events being timed). There is good experimental evidence for both views. Early genetic analysis of the budding yeast cell cycle provided evidence for two parallel dependent sequences (the budding sequence and the DNA replication sequence) that diverged in G1 phase (at the Start transition) and reconnected at the end of the cycle (exit from mitosis). On the other hand, biochemical studies of frog egg extracts suggested an autonomous oscillation of mitosis-promoting factor (MPF) that drives periodic DNA replication and mitosis, but generates periodic bursts of MPF activity quite independently of chromosomes and nuclei. In the first view, cell cycle progression is orchestrated by a gene regulatory network flipping genes on and off in a strict sequence intimately connected to cell growth. In the second view, cell cycle progression is governed by a protein interaction network that drives MPF activity up and down in waves of synthesis and degradation of cyclin proteins.

Systems Biology

These two views were brought together by Novak and Tyson (1993) in an early example of "molecular systems biology." They used the wiring diagram in Fig. 1 to derive a mathematical model of spontaneous MPF oscillations in frog egg extracts and (in later papers) of cell cycle mutants in fission yeast. According to their theory, the regulation of CDK-cyclin activity by tyrosine phosphorylation and dephosphorylation of the CDK subunit, by Wee1 (the kinase) and Cdc25 (the phosphatase), creates a bistable switch that governs the transition from G2 phase into mitosis (a state of high CDK activity); see Fig. 2a. This switch has all the properties of a classic cell cycle checkpoint. To pass the checkpoint, a cell must fully replicate its DNA and grow to a sufficient size. Once past the checkpoint the transition is irreversible; the cell does not slip back into G2 phase and try to enter mitosis a second time. Rather, the cell must complete the stages of mitosis and activate cyclin degradation at the metaphase-anaphase transition. Cyclin degradation allows the



Cell Cycle Dynamics, Bistability and Oscillations, Fig. 2 Signal-response curves: MPF activity as a function of total cyclin. (**a**) Bistable switch. If cyclin synthesis and degradation are blocked, then the total cyclin concentration ([MPF] + [preMPF]) in a frog egg extract can be manipulated experimentally. For [total cyclin] between the two thresholds, θ_{inact} and θ_{act}, the control system has two stable steady states separated by an unstable steady state. The dotted line shows how MPF activity will respond to slowly increasing [total cyclin] and then slowly decreasing [total cyclin]. (**b**) Relaxation oscillations. If cyclin is steadily synthesized, then [total cyclin] will increase when MPF is in the low-activity state, because Cdc20 is unavailable, but once the system flips to the state of high MPF activity, then Cdc20 is produced and cyclin is degraded faster than it is synthesized. The bistable switch flips back to the low-activity state once [total cyclin] drops below the inactivation threshold. The white circle represents an unstable steady state. (**c**) Damped sinusoidal oscillations. If the CDK phosphorylation site (a tyrosine residue) is mutated to phenylalanine, then the positive feedback loops involving Wee1 and Cdc25 are disengaged and bistability is lost. The system now undergoes damped oscillations to a stable steady state (black circle). If there is enough time delay in the negative feedback loop through Cdc20-APC, then the control system might exhibit sustained oscillations, but they are quite distinct from the relaxation oscillations in panel B

bistable switch to be reset to the state of low CDK activity. Repeated cycles of DNA synthesis and mitosis correspond to repeated flipping of the switch from a stable steady state of low CDK activity (interphase) to a stable steady state of high CDK activity (M phase) and back again. (Physicists and engineers call this behavior a "hysteresis loop".)

Novak and Tyson showed, furthermore, that under the special conditions in an early frog embryo or in a frog egg extract, the CDK-cyclin control system can generate spontaneous oscillations of MPF activity, unconstrained by requirements of cell growth, DNA synthesis, DNA damage, or mitotic spindle functions. The spontaneous oscillations are driven by periodic bursts of cyclin degradation, which are in turn generated by the periodic activation of MPF by dephosphorylation of the CDK subunit. These spontaneous oscillations result from an interplay between the bistable switch and a negative feedback loop (MPF activates the cyclin degradation machinery which

destroys cyclin subunits, causing a loss of MPF activity); see Fig. 2b. (Physicists and engineers call this behavior a "relaxation oscillator.")

The mathematical model of Novak and Tyson not only gave a systematic account of a variety of experimental observations on fission yeast, budding yeast, and frog eggs, but also made a series of striking predictions:

1. The "cyclin threshold for MPF activation" that had been observed by Solomon et al. (1990) is only one-half of a hysteresis loop. There should be a separate, distinctly lower cyclin threshold for MPF inactivation (Fig. 2a).
2. For cyclin levels well above threshold, the time lag for MPF activation should be roughly constant (as observed by Solomon et al. (1990)), but as cyclin level approaches the activation threshold from above, the time lag for MPF activation should increase dramatically.
3. The cyclin level for MPF activation should be an increasing function of unreplicated DNA in an extract.
4. If the bistable switch is disabled by interfering with the inhibitory phosphorylation of CDK, then MPF oscillations may still be possible, but they will be faster, more smooth (not abrupt bursts of MPF activity), and probably damped (Fig. 2c).

Predictions 1 and 2 are generic properties of bistable systems with hysteresis. Prediction 3 was a novel proposal for the mode of action of a checkpoint signal that delays entry into mitosis. Prediction 4 relates to classic properties of oscillations driven by a "simple negative feedback loop" in comparison to relaxation oscillations in a "substrate-depletion relaxation oscillator."

Experimental Verification

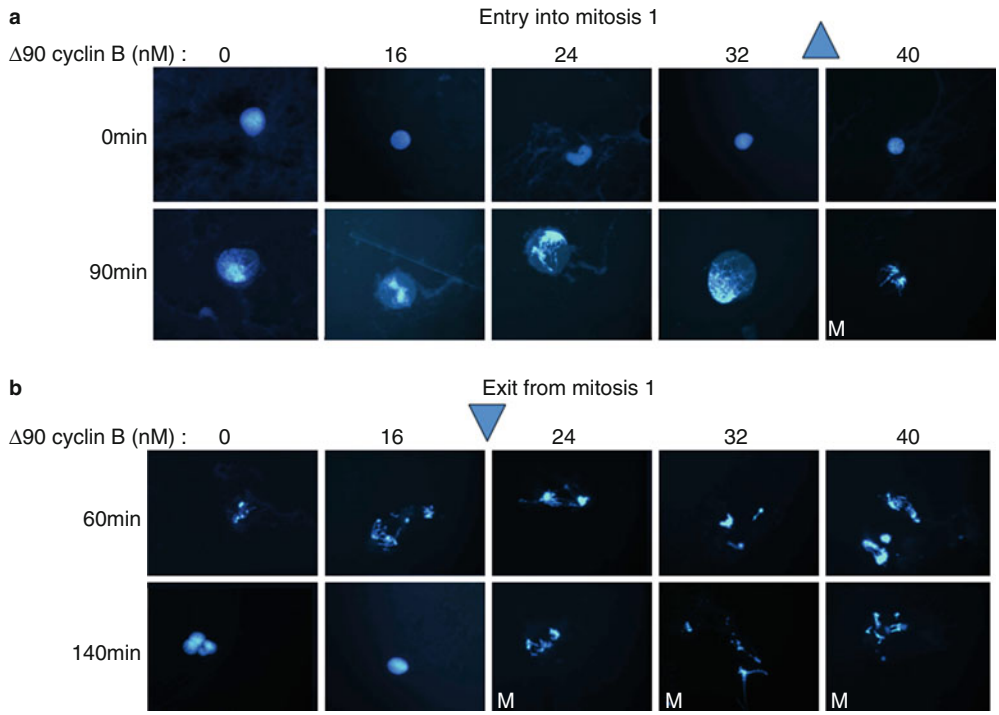
The first three predictions of the Novak-Tyson model were confirmed in frog egg extracts independently and simultaneously by Sha et al. (2003) and Pomerening et al. (2003); see Fig. 3. Slightly later, Pomerening et al. (2005) confirmed prediction 4 in a thorough and careful study of MPF oscillations in frog egg extracts; see Fig. 4.

In 1996, Kim Nasymth (1996) proposed that, at its heart, the budding yeast cell cycle is a repetitive alternation between two "self-maintaining" states (i.e., stable steady states): a G1 state with low CDK activity and an S-G2-M state with high CDK activity. This notion became the basis of a successful kinetic model of the budding yeast cell cycle by Chen et al. (2000). The model made many predictions that were tested in Fred Cross's laboratory. In particular, Cross et al. (2002) tested the idea that, under "neutral" conditions, budding yeast cells could persist indefinitely in either the G1 state, with low CDK activity, or the S-G2-M state, with high CDK activity, depending on the immediately prior history of how the cells are brought into the neutral conditions. This experiment is explained in the article on "Cell cycle, budding yeast."

Irreversibility

The concept of bistability provides an immediately obvious and intuitively satisfying explanation of the irreversibility of progression through the cell cycle. If cell cycle transitions are a result of passing from one stable steady state to another, then it is clear that the reverse transition cannot follow the same path. Once the transition is made, then a completely different set of circumstances must be brought into play to accomplish the reverse transition. For example, in Fig. 2a the activation of MPF is accomplished by increasing total cyclin concentration above the threshold, θ_{act}, where the unstable steady state merges with and annihilates the steady state of low MPF activity and forces the system to switch to the steady state of high MPF activity. If subsequently the cyclin level is caused to decrease, the control system does not jump back to the lower steady state at θ_{act}, which would be the case for a "reversible" transition. Rather, the cyclin level must decrease below a much smaller threshold, θ_{inact}, before the down-jump occurs.

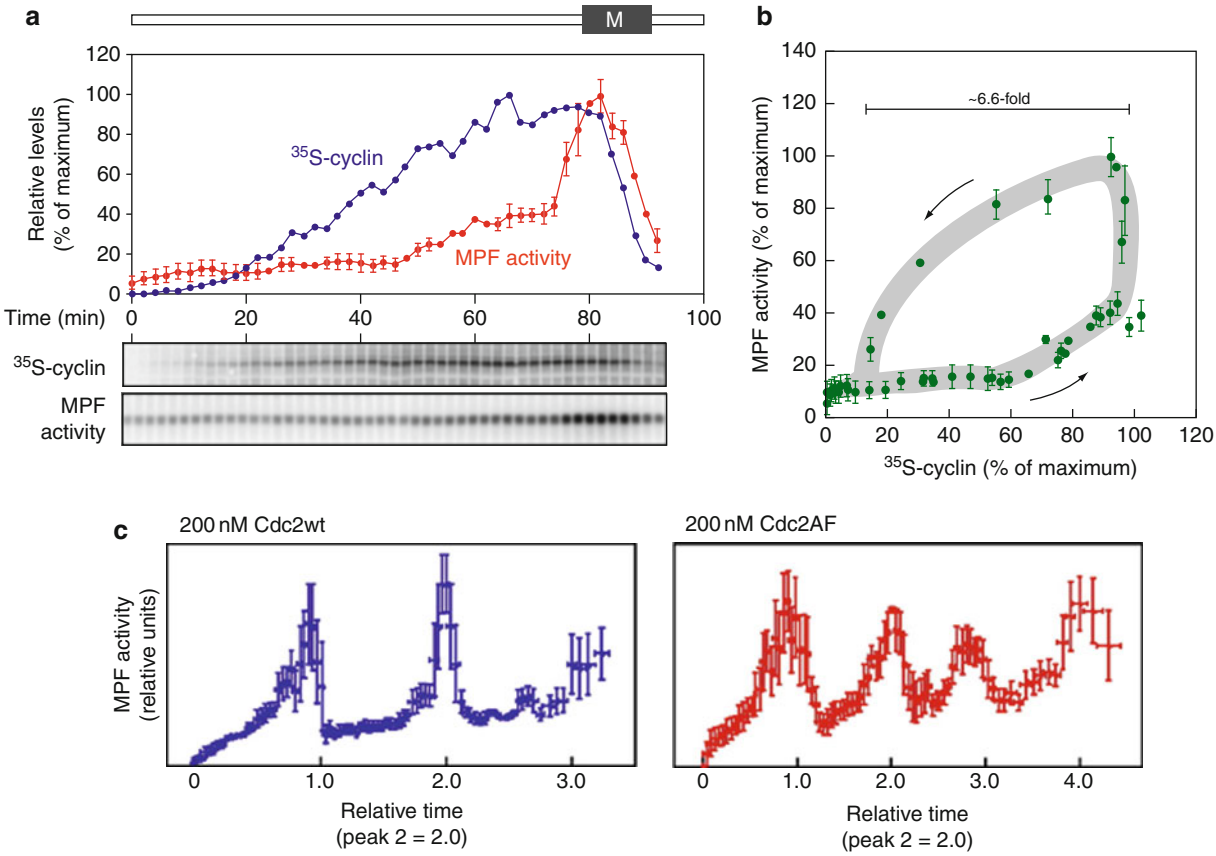
In the budding yeast case, the switch from the low-CDK state to the high-CDK state is driven by Cln-dependent kinase activity, but the switch back is driven by a completely different mechanism, dependent on the activities of Cdc20-APC (degradation of B-type cyclins) and Cdc14 (a CDK-counteracting phosphatase). In this case, the up-jump is a one-way switch: it is induced by Cln-dependent kinase, but after the switch is made, the Cln-kinase activity can drop to 0 and the CDK activity will remain high. Down-jump occurs by means of a different one-way switch. In the "neutral" condition (Cln = Cdc20 = Cdc14 = 0), the control system can persist indefinitely in either the low-CDK or the high-CDK state. For more details, see the articles on "Cell cycle, budding yeast" and "Cell cycle dynamics, irreversible transitions."



Cell Cycle Dynamics, Bistability and Oscillations, Fig. 3 Experimental confirmation of bistability in frog egg extracts. From Sha et al. (2003), used by permission. (**a**) A frog egg extract, containing all the proteins in Fig. 1 except for cyclin, is prepared in the presence of sperm nuclei (indicators of MPF activity) and cycloheximide (a drug that prevents the synthesis of cyclin protein from endogenous mRNA). In the absence of cyclin subunits, the extract is blocked in interphase ($t = 0$), as indicated by the compact nucleus with intact nuclear membrane and dispersed chromatin (stained blue). At $t = 0$ the extract is supplemented with a measured amount of nondegradable ($\Delta 90$) cyclin, and 90 min later the extract is observed to see if the nuclei are in interphase (dispersed chromatin, intact membrane, low MPF activity) on in mitosis (condensed chromatin, breakdown of nuclear membrane, high MPF activity). Cyclin concentrations less than ~35 nM are insufficient to induce entry into mitosis, but [total cyclin] = 40 nM is above the threshold (blue up-triangle) for mitotic entry. (**b**) In a second experiment, the extract is supplemented with variable amounts of nondegradable cyclin at $t = 0$, but cycloheximide is not added until $t = 60$ min. By $t = 60$ min, the nuclei in each sample have been driven into mitosis 1 by a combination of the nondegradable cyclin added at $t = 0$ and the degradable cyclin subunits synthesized from the extract's endogenous mRNA. As the extracts try to exit from mitosis 1, the endogenous cyclin subunits are degraded by Cdc20-APC, but the exogenous $\Delta 90$ cyclin subunits resist degradation. Cycloheximide prevents any further cyclin synthesis in the extracts. At $t = 140$ min the extracts are assayed for the cell cycle phase of the nuclei. Cyclin concentrations greater than ~20 nM are sufficient to maintain the nuclei in a mitotic state. [Total cyclin] = 16 nM is below the threshold (blue down-triangle) for inactivation of MPF and exit from mitosis. Cyclin concentrations of 24 and 32 nM are clearly in the bistable region: the nuclei can persist stably in interphase or in mitosis, depending on whether they are prepared initially in interphase or mitosis

Checkpoints

The concept of bistability also provides a natural framework for understanding the mechanisms of checkpoint controls. The function of a checkpoint is to block or delay progression of a damaged cell into the next phase of the cell cycle. For example, DNA damage should block cells from entering S phase, incomplete DNA ligation should block cells from entering M phase, and misaligned chromosomes should block cells from exiting mitosis. If these transitions are governed by saddle-node bifurcations of a bistable system (as in Fig. 2a), then the transition can be delayed or blocked completely by raising the threshold for the transition (θ_{act}). From the mathematical model of the transition it is immediately obvious which components control the location of the threshold. For example, in Fig. 1, it is the activity of the CDK-counteracting phosphatase (PPase) that is the vulnerable point.



Cell Cycle Dynamics, Bistability and Oscillations, Fig. 4 Experimental confirmation of relaxation oscillations in frog egg extracts. From Pomerening et al. (2005), used by permission. (**a**) In a "cycling" extract, cyclin subunits are continuously synthesized (blue curve), but MPF activity (red curve) is low until total cyclin exceeds the threshold for MPF activation. The abrupt activation of MPF drives nuclei into mitosis (M). As the extract exits from mitosis, cyclin is degraded by Cdc20-APC, and MPF activity falls as a result. (**b**) The data in panel A is projected onto state space (MPF activity versus total cyclin level), as in panel A of Fig. 2bB. (**c**) Confirmation of prediction 4. To a cycling extract, containing ~200 nM endogenous Cdc25wt (wild-type) protein, is added 200nM of either Cdc25wt protein or Cdc2AF protein, which cannot be phosphorylated and inhibited by Wee1. Hence, in the extract on the right there are roughly equal amounts of Cdc2wt and Cdc2AF subunits, compared to the "control" extract on the left which contains ~400 nM Cdc2wt subunits. The mutant kinase subunits compromise the positive feedback loops in the model (Fig. 1) and change the properties of MPF oscillations. Compared to the blue curve, the MPF oscillations in the red curve are faster, more sinusoidal and noticeably damped, exactly as predicted by the mathematical model

Cross-References

- Cell Cycle Dynamics, Irreversibility
- Cell Cycle Model Analysis, Bifurcation Theory
- Cell Cycle Modeling, Differential Equation
- Cell Cycle of Early Frog Embryos
- Cell Cycle, Budding Yeast
- Cell Cycle, Fission Yeast

References

Chen KC, Csikasz-Nagy A, Gyorffy B, Val J, Novak B, Tyson JJ (2000) Kinetic analysis of a molecular model of the budding yeast cell cycle. Mol Biol Cell 11:369–391

Cross FR, Archambault V, Miller M, Klovstad M (2002) Testing a mathematical model for the yeast cell cycle. Mol Biol Cell 13:52–70

Murray AW, Kirschner MW (1989) Dominoes and clocks: the union of two views of the cell cycle. Science 246:614–621
Nasmyth K (1996) At the heart of the budding yeast cell cycle. Trends Genet 12:405–412
Novak B, Tyson JJ (1993) Numerical analysis of a comprehensive model of M-phase control in Xenopus oocyte extracts and intact embryos. J Cell Sci 106:1153–1168
Pomerening JR, Sontag ED, Ferrell JE Jr (2003) Building a cell cycle oscillator: hysteresis and bistability in the activation of Cdc2. Nat Cell Biol 5:346–351
Pomerening JR, Kim SY, Ferrell JE Jr (2005) Systems-level dissection of the cell-cycle oscillator: bypassing positive feedback produces damped oscillations. Cell 122:565–578
Sha W, Moore J et al (2003) Hysteresis drives cell-cycle transitions in Xenopus laevis egg extracts. Proc Natl Acad Sci USA 100:975–980
Solomon MJ, Glotzer M et al (1990) Cyclin activation of p34cdc2. Cell 63:1013–1024
Tyson JJ, Novak B (2008) Temporal organization of the cell cycle. Curr Biol 18:R759–R768

Cell Cycle Dynamics, Irreversibility

John J. Tyson[1] and Béla Novák[2]
[1]Department of Biological Sciences, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA
[2]Department of Biochemistry, Oxford Centre for Integrative Systems Biology, University of Oxford, Oxford, UK

Synonyms

Bistability; Checkpoints; Hysteresis

Definition

The cell cycle is the sequence of events whereby a cell replicates its chromosomes and partitions the identical sister chromatids to two separate nuclear compartments (Morgan 2007). In eukaryotic cells, the phases of DNA synthesis (S phase) and mitosis (M phase) are temporally distinct and separated by gap phases (G1 – unreplicated chromosomes, and G2 – replicated chromosomes). To maintain the proper ploidy of a cell lineage, generation after generation, it is essential that S and M phases are strictly alternating. This alternation is enforced by the irreversibility of three crucial transitions in the cell cycle: the G1/S, G2/M, and M/G1 transitions (Novak et al. 2007). By "irreversibility" we understand that, once cells have committed to a new round of DNA replication (at the G1/S transition), they do not typically slip back into G1 phase and do a second round of DNA replication. Similarly, once cells have committed to mitosis (at the G2/M transition), they do not typically slip back into G2 phase and try for a second mitotic division.

Like most "rules" in biology, this one has important exceptions (Morgan 2007). Some cells carry out repeated rounds of DNA replication without intervening mitoses, becoming polyploid. During meiosis, a germ line cell undergoes two meiotic divisions without an intervening S phase. But these exceptions only reinforce the centrality of the general rule of somatic cell cycles, namely, irreversible progression through the cell cycle phases (G1, S, G2, M) in strict order.

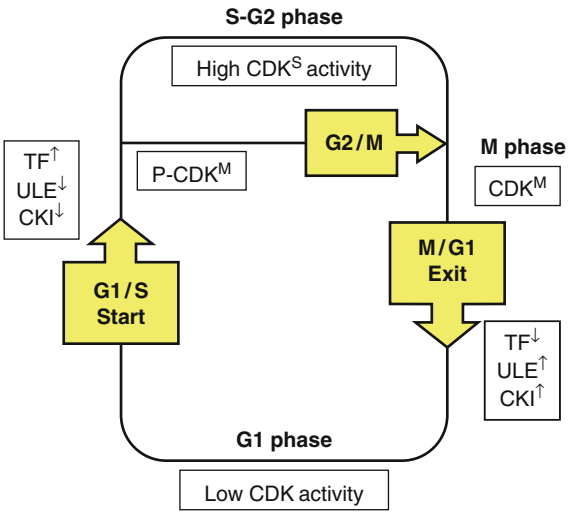
The irreversibility of the three transitions is intimately connected with cell cycle checkpoints, which halt further progression through the cell cycle whenever serious problems are detected (Murray 1992). For example, DNA damage incurred during G1 phase will block the G1/S transition until the damage is repaired. Failure to fully replicate and ligate DNA molecules during S phase will block the G2/M transition. Incomplete alignment of replicated chromosomes on the mitotic spindle will block the M/G1 transition. When a checkpoint is lifted, the cell proceeds irreversibly to the next phase of the cell cycle.

These irreversible transitions and checkpoints are controlled by complex molecular regulatory networks with both positive and negative feedback. Positive feedback creates a bistable switch, and negative feedback allows the switch to be flipped from one stable state to another (Tyson and Novak 2008). These systems-level properties of the regulatory network are crucial to irreversible progression through the cell cycle, and when they are disturbed by mutation, drugs, or disease, then cells make mistakes in DNA replication and partitioning, often with fatal consequences for the cell or for the multicellular organism harboring the rogue cells.

Characteristics

Physiology and Molecular Biology

Progression through the cell cycle is governed by a set of cyclin-dependent kinases (CDKs) that initiate DNA



Cell Cycle Dynamics, Irreversibility, Fig. 1 Three checkpoints/irreversible transitions in the eukaryotic cell cycle. G1 phase: low CDK activity. S-G2 phase: CDK-cyclin^S activity is high, CDK-cyclin^M activity is low. M phase: CDK-cyclin^M activity is high. At the G1/S transition, cyclin synthesis is turned on, cyclin degradation is turned off, and CDK inhibitors are destroyed. At the M/G1 transition, these switches are reversed. In S-G2 phase, CDK-cyclin^M activity is kept low by phosphorylation of the CDK subunit. At the G2/M transition, this inhibitory phosphorylation is removed

replication and mitosis and that inhibit the transition from metaphase to anaphase-telophase-cytokinesis ("exit from mitosis"). G1 phase cells are uncommitted to the cell cycle because they lack the necessary CDK activities. They are devoid of S- and M-phase cyclins because the transcription factors that would drive production of these cyclins are inactive, and the ubiquitin-ligating enzymes (ULE) that label these cyclins for proteolysis are active. In addition, G1 phase cells contain generous amounts of CDK inhibitors (CKIs) that bind to and inhibit CDK-cyclin dimers.

At the G1/S transition (called "Start" in budding yeast and the "Restriction Point" in mammalian cells), a cell must do three things: (1) turn on the synthesis of S- and M-phase cyclins, (2) turn off their degradation, and (3) get rid of the G1-phase CKIs (See Fig. 1). At the M/G1 transition ("Exit"), these three switches must be reversed. As cells proliferate, a lineage (mother-daughter-granddaughter) toggles back and forth between states of low CDK activity (G1) and high CDK activity (S-G2-M).

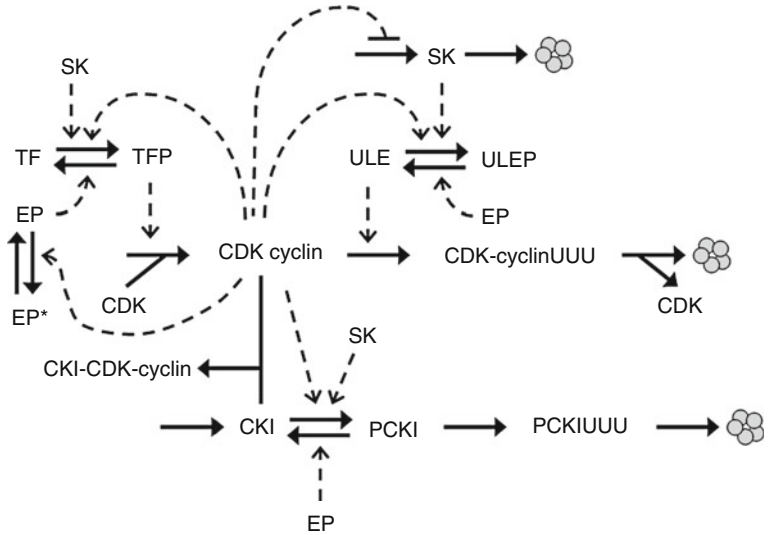
Embedded inside this fundamental toggle switch is a secondary toggle switch governing the transition into M phase (see Fig. 1). During S and G2 phases, the cell is producing M-phase cyclins, but CDK/cyclin-M activity is low because of inhibitory phosphorylation of the CDK subunit. At the G2/M transition, the kinases that impose this inhibition must be inactivated, and the phosphatases that remove this inhibition must be activated, thereby allowing the cell to transit abruptly into mitosis. At the M/G1 transition, these two switches must be reversed as well.

The scenario just described is a generic account of molecular events at the three basic transitions of the cell cycle. In any specific type of cell (e.g., budding yeast, fission yeast, mammalian cell) there may be variations on this general scheme, but most cell types studied in depth show evidence of CDK-governing toggle switches that are flipped on and off at alternating transitions of the cell cycle.

The greatest exception to this rule is the newly fertilized egg, a gigantic cell that undergoes rapid S-M cycles, lacking gap phases and checkpoints. It is arguable whether these cell cycles are governed by irreversible toggle switches or a simple periodic oscillator. Nonetheless, it is true that early embryonic cell cycles often make serious mistakes (producing polyploid or aneuploid cells) that lead ultimately to death of the embryo. These errors seem to be a price that organisms are willing to pay in order to hurry through the most vulnerable phase of their life cycle. In this context, these errors reinforce the notion that toggle switches and irreversible transitions are crucial to maintaining genomic integrity of proliferating eukaryotic cells.

Bistability and Hysteresis

From a theoretical perspective, irreversible transitions are intimately related to bistability of molecular regulatory networks. The article on "cell cycle dynamics, bistability and oscillations" discusses the molecular mechanisms underlying bistability and hysteresis at the G2/M transition. Here we describe the molecular basis of irreversibility at Start and Exit. Figure 2 illustrates the generic interactions among CDK-cyclin, its transcription factors, its ubiquitin-ligating enzymes, and its stoichiometric inhibitors. In each case, CDK is involved in a positive feedback interaction with a "friend" (TF) or a double-negative feedback interaction with an "enemy" (ULE or CKI). These interactions create the possibility of bistability: two stable steady states ("nodes") separated by an unstable steady



Cell Cycle Dynamics, Irreversibility, Fig. 2 CDK regulatory network for a generic eukaryotic cell. Cyclin synthesis is regulated by a transcription factor (TF), and cyclin degradation is regulated by a ubiquitin-ligating enzyme (ULE). CDK-cyclin dimers can also be inhibited by binding to an inhibitor protein (CKI). CKI stability is controlled by phosphorylation: PCKI is a good substrate for a different ubiquitin-ligating enzyme, which labels CKI for degradation by proteasomes. Because TF is activated by phosphorylation by CDK-cyclin, these two components are involved in a positive feedback loop. Because ULE is inactivated by phosphorylation by CDK-cyclin, these two components are involved in a double-negative feedback loop, as are CKI and CDK-cyclin. These interactions create a bistable control system, with a stable steady state of low CDK activity (G1 phase) and an alternative stable steady state of high CDK activity (S-G2-M phase); see Figs. 1 and 3. Starter kinase activity (SK) tips the control system toward the high CDK state, and exit phosphatase activity (EP) tips the system toward the low CDK state. SK and EP are related to CDK by negative feedback loops. SK upregulates CDK-cyclin by flipping the switch to the high CDK state, but high activity of CDK-cyclin downregulates the synthesis of SK. On the other hand, high activity of CDK-cyclin activates EP, which flips the switch to the low CDK state and consequently EP is inactivated

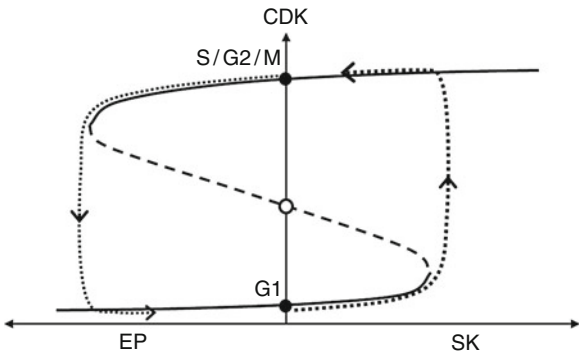
state (a "saddle point"). The stable steady states are characterized by either low or high CDK activity (Fig. 3). The CDK control system can persist in either of these stable steady states, corresponding to G1 or S-G2-M phases of the cell cycle, respectively.

For a G1 cell to commit to a new round of DNA replication and division, it must be promoted from the G1 steady state to the S-G2-M steady state. This is the job, generically speaking, of a "starter kinase" (SK), typically a special cyclin-CDK pair (Cln3-Cdc28 in budding yeast, cyclin D-Cdk4/6 in mammalian cells). Transient activation of SK can induce a transition from G1 to S-G2-M. Even if SK activity disappears after the transition, the control system will remain in the stable steady state of high CDK activity. In this sense, the G1/S transition is irreversible. To exit mitosis and return to G1, the cell must transit from the high- to the low CDK steady state. This is the job of generic "exit phosphatases" (EP) that are activated during anaphase and telophase of the cell cycle. These phosphatases push ULE and CKI toward their "on" states and TF toward its "off" state, allowing the bistable switch to flip off. The M/G1 transition is irreversible because, even though EP activity disappears in early G1, the control system remains in the stable G1 steady state (Tyson and Novak 2008).

Irreversibility and Proteolysis

For a time it was fashionable among molecular biologists to attribute irreversibility of cell cycle transitions to the proteolysis of key cell cycle regulators at each transition. Not only are cyclins dramatically degraded at the M/G1 transition (Minshull et al. 1989), but also CKIs are degraded at the G1/S transition (Schwob et al. 1994) and CDK-inactivating kinases are degraded at the G2/M transition (Michael and Newport 1998). As simple and appealing as this idea might be, it is clearly insufficient to explain the irreversibility/directionality of cell cycle progression.

First of all, proteolysis is not irreversible in a kinetic sense; its opposing process is protein synthesis. In general, the rates of protein synthesis and degradation



Cell Cycle Dynamics, Irreversibility, Fig. 3 Bistability, irreversibility, and hysteresis. The steady states of the control system in Fig. 2 are related to SK and EP activities by this bifurcation diagram. When SK = EP = 0, the control system has two stable steady states (black circles) separated by an unstable steady state (white circle). The cell can be in either G1 phase or S-G2-M phase, depending on its recent history. A newborn cell, by definition, is in G1 phase. For it to start a new round of DNA synthesis and degradation, a starter kinase (SK) must be activated. Activation of SK is controlled by checkpoint signals, such as growth factors, DNA damage sensors, size sensors, etc. Once SK flips the switch into the CDK-active state, it is no longer needed: SK activity can drop to 0, and the control system remains in the CDK-active state. For the cell to exit mitosis, undergo cytokinesis, and return to G1 phase, an exit phosphatase (EP) must be activated. EP activation depends on other checkpoint signals, such as complete replication of DNA and successful alignment of all replicated chromosomes on the mitotic spindle

will balance each other at a unique, stable steady state of protein abundance. To upset this balance, it is necessary to switch the rates of protein synthesis and degradation between low and high values, and such switching behavior is a property of feedback interactions in a regulatory network.

In addition, it is possible, by mutations and/or drug treatments, to eliminate proteolysis at any one of the three cell cycle transitions without compromising its irreversibility. For example, *ckiΔ* mutants in budding yeast proceed irreversibly through the G1/S transition without proteolysis of any of the remaining regulatory proteins (TF and ULE). Proteasome-inhibited mammalian cells, blocked in mitosis by nocodazole, can be induced to exit mitosis irreversibly by transient inhibition of CDK activity by flavopiridol (Potapova et al. 2006). Frog egg extracts, without either cyclin synthesis or degradation, can be induced to enter or exit mitosis simply by manipulating the activities of the CDK-inhibiting kinase and/or the CDK-activating phosphatase.

In conclusion, the irreversibility of cell cycle transitions, which is crucial to maintaining the genomic integrity of proliferating cells, is a systems-level property of the molecular networks that regulate CDK activity in eukaryotic cells. Positive feedback loops in the control system generate multiple stable steady states (G1, S-G2 and M states), and abrupt, irreversible transitions between these stable steady states are induced by pro-proliferative signals (growth factors, size increase, successful completion of prior cell cycle events) and inhibited by checkpoint signals (indicators of potentially fatal mistakes in the genome replication process). Irreversibility is not a property of any single gene, protein, or reaction but rather an emergent property of systems-level interactions.

Cross-References

- ▶ Cell Cycle Dynamics, Bistability and Oscillations
- ▶ Cell Cycle Model Analysis, Bifurcation Theory
- ▶ Cell Cycle Modeling, Differential Equation
- ▶ Cell Cycle of Early Frog Embryos
- ▶ Cell Cycle, Budding Yeast
- ▶ Cell Cycle, Fission Yeast

References

Michael WM, Newport J (1998) Coupling of mitosis to the completion of S phase through Cdc34-mediated degradation of Wee1. Science 282:1886–1889

Minshull J, Pines J, Golsteyn R, Standart N, Mackie S, Colman A, Blow J, Ruderman JV, Wu M, Hunt T (1989) The role of cyclin synthesis, modification and destruction in the control of cell division. J Cell Sci Suppl 12:77–97

Morgan DO (2007) The cell cycle: Principles of control. New Science, London

Murray AW (1992) Creative blocks: cell-cycle checkpoints and feedback controls. Nature 359:599–604

Novak B, Tyson JJ, Gyorffy B, Csikasz-Nagy A (2007) Irreversible cell-cycle transitions are due to systems-level feedback. Nat Cell Biol 9:724–728

Potapova TA, Daum JR, Pittman BD, Hudson JR, Jones TN, Satinover DL, Stukenberg PT, Gorbsky GJ (2006) The reversibility of mitotic exit in vertebrate cells. Nature 440:954–958

Schwob E, Bohm T, Mendenhall MD, Nasmyth K (1994) The B-type cyclin kinase inhibitor p40SIC1 controls the G1 to S transition in S. cerevisiae. Cell 79:233–244

Tyson JJ, Novak B (2008) Temporal organization of the cell cycle. Curr Biol 18:R759–R768

Cell Cycle Kinases

▶ Mitotic Kinases

Cell Cycle Model Analysis, Bifurcation Theory

John J. Tyson
Department of Biological Sciences, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA

Synonyms

Nonlinear dynamical systems theory

Definition

Bifurcation theory provides a classification of the expected ways in which the number and/or stability of invariant solutions ("attractors" and "repellers") of nonlinear ordinary differential equations may change as parameter values are changed. The most common qualitative changes are "saddle-node" bifurcations, "Hopf" bifurcations, and "SNIPER" bifurcations. At a saddle-node bifurcation, a pair of steady states, usually a stable node and an unstable saddle point, coalesce and disappear. At a Hopf bifurcation, a stable focus changes to an unstable focus and makes way for a small amplitude periodic solution ("limit cycle"). At a SNIPER bifurcation, the coalescence of a saddle point and a stable node creates an infinite-period limit-cycle solution. These bifurcations have clear physiological correlates in the regulation of DNA replication, mitosis, and cell division. Saddle-node bifurcations are related to checkpoints in the cell cycle: The establishment and removal of checkpoints correspond to the creation and annihilation of stable steady states at saddle-node bifurcations. The repetitive nature of the cell cycle (G1-S-G2-M-G1-etc.) is related to limit-cycle solutions of the underlying kinetic equations: The ability to oscillate spontaneously arises at either a Hopf or a SNIPER bifurcation.

Characteristics

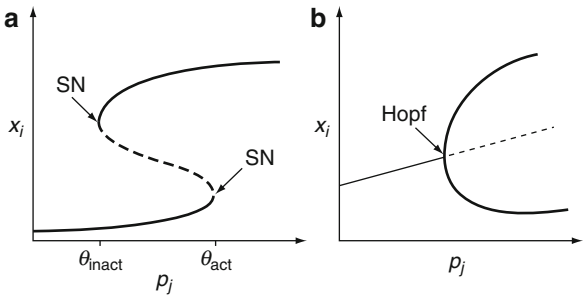
Historical Background

Since the days of Isaac Newton, ordinary differential equations (ODEs) have been used throughout the physical and life sciences to describe the temporal development of dynamical systems: from the solar system, to the clock radio, to the regulation of DNA replication and cell division. Initially the focus was on ODEs that could be solved exactly in terms of "elementary" functions of high school algebra and trigonometry or "special" functions of mathematical physics. But in the 1890s Poincaré (1899) introduced the "qualitative" theory of dynamical systems, i.e., systems of n nonlinear ODEs

$$\begin{aligned} &\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x};\mathbf{p}),\ \mathbf{x}(t) = \{x_1, ..., x_n\} = \text{Variables}, \\ &\mathbf{p} = \{p_1, ..., p_m\} = \text{Parameters} \end{aligned} \tag{1}$$

Poincare proposed to interpret these equations as a vector field in n-dimensional state space, $\mathbf{x}$, and to characterize this vector field by its invariant solutions, which can be either attractors or repellers. The crucial question for Poincare was not "what is the exact solution of the ODE?" but "how do the qualitative features of the attractors and repellers depend on the values of the parameters?" This latter question is the subject of bifurcation theory (Odell 1980; Strogatz 1994), which was developed in the mid-twentieth century by Andronov's school of Russian physicists and engineers for $n = 2$ (Andronov et al. 1966), and later by a host of mathematicians for the general case (Kuznetsov 2004).

A one-parameter bifurcation diagram begins with a plot of the steady-state value of a chosen dynamical variable, x_i, as a function of a chosen parameter, p_j, the "bifurcation parameter." In Fig. 1a, we plot a typical bifurcation diagram for a bistable system. Between the two thresholds, $\theta_{\text{inact}} < p_j < \theta_{\text{act}}$, the system can persist in either of two stable steady states (x_i small or x_i large). Precisely at the thresholds, $p_j = \theta_{\text{inact}}$ and $p_j = \theta_{\text{act}}$, the dynamical system undergoes a bifurcation from one type of behavior (a single stable steady state) to a qualitatively different type of behavior (bistability). This type of bifurcation is called a "saddle-node" or "fold." In Fig. 1b, we illustrate a "Hopf" bifurcation, in which a stable steady state loses stability and gives rise to stable limit-cycle



Cell Cycle Model Analysis, Bifurcation Theory, Fig. 1 One-parameter bifurcation diagrams. (**a**) Saddle-node bifurcation. *Solid line*: stable steady state; *dashed line*: unstable steady state. (**b**) Hopf bifurcation. *Thin solid line*: stable steady state; *thin dashed line*: unstable steady state; *thick solid line*: maximum and minimum values attained by a stable limit-cycle oscillation

oscillations. The limit cycles are born with small amplitude and grow in size as the parameter value pulls away from the bifurcation point. We shall meet some other types of bifurcations shortly.

Of special interest to systems biologists are the musings of Rene Thom (1989) on "structural stability and morphogenesis." Thom was highly regarded among mathematicians for his study of gradient dynamical systems,

$$\frac{d\mathbf{x}}{dt} = \nabla U(x_1, ..., x_n) \tag{2}$$

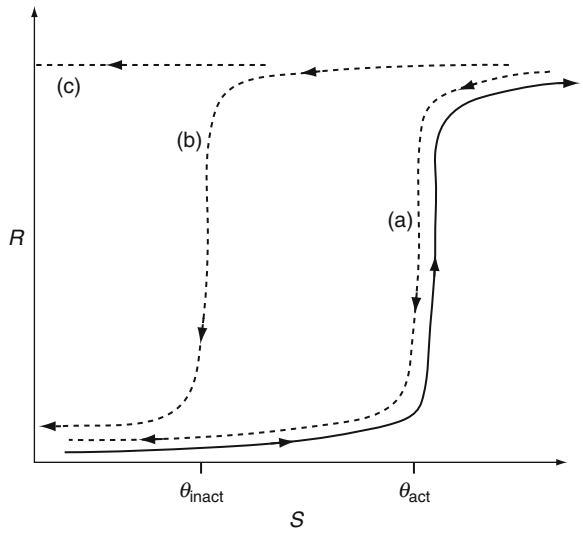
where $U(\mathbf{x})$ is a scalar function of the variables (think of it as the "potential energy" of the system) and $\nabla = (\partial/\partial x_1, ..., \partial/\partial x_n)$ is the gradient operator. The steady-state solutions of Eq. 2 are places where $\nabla U = 0$, i.e., "singularities" of the potential function. Thom's great contribution was to classify the topologically distinct types of singularities of potential functions in n dimensions. Next, Thom took the unusual step – unusual for a famous mathematician – to speculate that the bifurcations he had characterized might underlie the "unfolding" of a fertilized egg into a larva. Following Waddington's hypothesis that embryonic development is the evolution of a dynamical system on an "energy landscape," Thom pointed out that his complete classification of the qualitative changes of behavior that could be observed under this type of gradient dynamic must provide the key to understanding morphogenetic transitions.

Thom's ideas were bitterly opposed by both theoretical and experimental biologists of his generation, and (perhaps) for good reasons. First of all, morphogenesis is governed, we know, by the interactions of genes and proteins (i.e., a biochemical interaction network), which is *not* a gradient dynamical system. Hence, the bifurcations of relevance to molecular cell biologists are not the singularities of potential functions (Thom's case) but rather the generic bifurcations of nonlinear ODEs (the case of Andronov et al.). But Thom's more fundamental idea (stripped of its unfortunate alliance to Waddington's energy landscape) – that qualitative changes in cell physiology should be correlated with qualitative changes in the attractors and repellers of a vector field (a system of nonlinear ODEs) – is absolutely correct. It is the basis of the application of bifurcation theory to problems in molecular cell biology.

One-Parameter Bifurcation Diagrams and Signal-Response Curves

The connection between bifurcation theory and cell physiology is the signal-response curve. In a typical experiment, a molecular cell biologist might challenge cells with increasing amounts of an extracellular signal molecule and measure whether certain downstream genes are expressed or not. And a typical result is that, for low signal levels there is no expression, but for signal levels above a certain threshold there is strong expression of the gene (Fig. 2). In this circumstance, it is natural to ask what happens if the signal level is steadily decreased in cells that are expressing protein R? Do they turn off at the same signal strength where they turned on? Or at a much lower signal strength? Or not at all?

In the first case, the signal-response curve is perfectly smooth and reversible; there are no qualitative changes in the behavior of the control system as the signal varies up and down. In the second case, there is a region of bistability between the two thresholds, and the behavior of the control system is qualitatively different over three ranges of signal strength: for $S < \theta_{inact}$ there is a single stable steady state with R small; for $\theta_{inact} < S < \theta_{act}$ the control system can persist in either of two attractors (R small or R large) that are separated by an unstable steady state; and for $S > \theta_{act}$ there is a single stable steady state with R large. This is exactly the case of a one-parameter bifurcation diagram with saddle-node bifurcations bounding a zone of bistability (Fig. 1a).



Cell Cycle Model Analysis, Bifurcation Theory, Fig. 2 Signal-response curve. The experimentalist can vary the signal strength S (say, the concentration of an extracellular ligand) and observe the response R (say, the expression level of a gene induced by S). As S is slowly increased, the expression of R turns on abruptly; a typical threshold-type response. What happens as S is slowly decreased? There are three possibilities. (**a**) The gene expression turns on and off at the same threshold signal strength: the signal-response curve is smooth and reversible; e.g., a Hill function. (**b**) The threshold for gene inactivation (θ_{inact}) is lower than the threshold for gene activation (θ_{act}): The signal-response curve has a region of bistability and functions like a toggle switch. (**c**) The gene cannot be inactivated by lowering the signal strength even to zero: The control system functions as a one-way switch

Cell Cycle Model Analysis, Bifurcation Theory, Table 1 Generic bifurcations of dynamical systems

Name	Characteristics	Cell cycle correlate
Saddle-node	Creation and annihilation of pairs of steady states	Irreversible transitions; checkpoints
Hopf, supercritical	Birth of stable limit cycles of small amplitude and finite frequency	Spontaneous MPF oscillations in embryos
Hopf, subcritical	Birth of unstable limit cycles of small amplitude and finite frequency	Subcritical Hopf and cyclic fold bifurcations occur in pairs and may correlate with embryonic MPF oscillations
Cyclic fold	Creation and annihilation of pairs of limit cycles	
Saddle-loop	Annihilation of a limit cycle by a homoclinic orbit at a saddle point; finite amplitude and small frequency	SL and SNIPER bifurcations are closely related; they are involved in irreversible transitions in the yeast cell cycle (budding yeast and fission yeast)
SNIPER[1] or SNIC[2] (synonomous)	Annihilation of a limit cycle by a homoclinic orbit at a saddle-node; finite amplitude and small frequency	

[1]SNIPER = saddle-node infinite-period
[2]SNIC = saddle-node invariant-circle

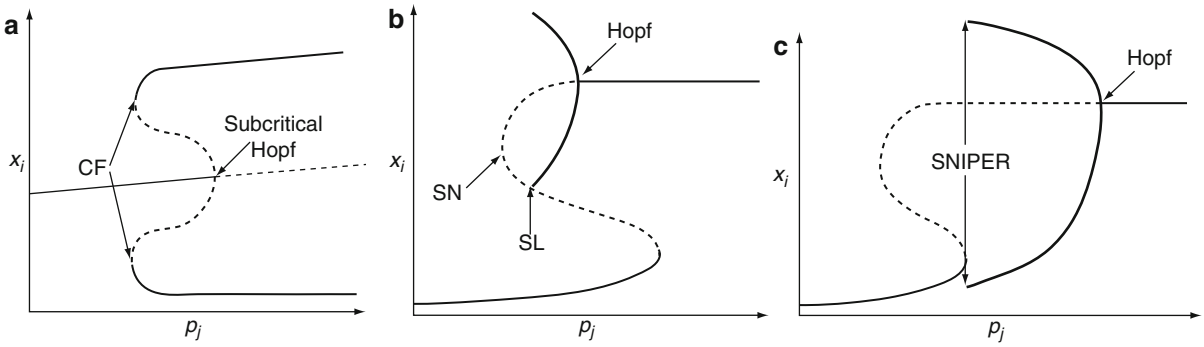
It is possible that $\theta_{inact} < 0$, in which case the signal, $S = [S]$ (a positive number), cannot be made small enough to flip the switch off. In this case, the control system is said to be a one-way switch. By increasing S, the switch can be turned on, but it cannot be turned off by decreasing S.

There are many convincing examples of toggle switches in molecular and cell biology generally (Tyson et al. 2003) and in cell cycle regulation particularly (▶ Cell Cycle Dynamics, Bistability and Oscillations).

In another common physiological situation, a cellular process begins to oscillate when a stimulating signal gets large enough (Goldbeter 1996). In this case, the signal-response curve exhibits a Hopf bifurcation, as in Fig. 1b.

These examples suggest that the signal-response curves often measured by cell physiologists are none other than one-parameter bifurcation diagrams in the parlance of applied mathematicians. If we may associate abrupt, qualitative changes in signal-response characteristics of living cells with bifurcations in vector fields of nonlinear dynamical systems, then it is natural to ask how many different types of generic bifurcations are exhibited by dynamical systems and what do they look like? Are there hundreds of different types of bifurcations to match the seemingly boundless variety of cellular behaviors? Or are all the peculiarities of cellular signal processing simply variations on a few common themes? (Table1)

The answer is the latter. In addition to the saddle-node and Hopf bifurcations illustrated in Fig. 1, there are only a few other common, generic, one-parameter bifurcations: subcritical Hopf, cyclic fold, saddle-loop, and SNIPER bifurcations (Fig. 3).



Cell Cycle Model Analysis, Bifurcation Theory, Fig. 3 The other common types of bifurcation points. (**a**) Subcritical Hopf bifurcation and cyclic fold (*CF*) bifurcation. (**b**) Saddle-loop (*SL*) bifurcation. (**c**) Saddle-node infinite-period (*SNIPER*) bifurcation

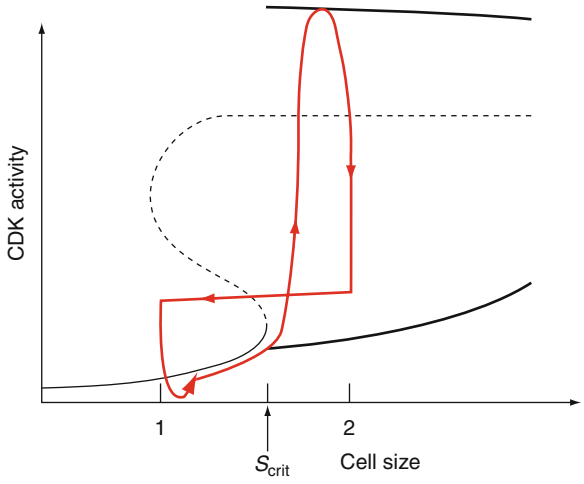
Of this we can be certain: Cell physiology is governed by underlying regulatory networks that consist of biochemical reactions among genes, RNAs, and proteins. These networks are dynamical systems; their dynamics are governed by nonlinear ODEs (biochemical kinetic equations). The solutions of these equations determine the time-dependent behavior of the cell, and the nature of these solutions are determined by the attractors and repellers of the dynamical vector field in state space. Qualitative changes in the behavior of cells must be reflections of qualitative changes in the nature of these attractors and repellers, i.e., on the generic bifurcations of nonlinear vector fields. Hence, the six types of bifurcations we have introduced must be the basic building blocks of all cellular signal-response curves. It is this connection between signal-response curves of living cells and one-parameter bifurcation diagrams of dynamical systems that is the heritage of Rene Thom's proposal.

An important caveat to this interpretation of bifurcation theory is the fact that single cells are very small, with limited numbers of molecules (10s, 100s, 1,000s) of each of the interacting species. Hence, continuous ODEs are only a first approximation to the dynamics of intracellular molecular control systems. The effects of stochastic variations of small numbers of molecules can have significant effects on the qualitative features of dynamical systems. Stochastic effects must be given due consideration, but subsequent to a thorough study of the system by bifurcation theory.

Relation to Cell Cycle Regulation

Bifurcation theory has been used to study the molecular basis of cell cycle regulation (e.g., Borisuk and Tyson 1998; Tyson et al. 2003; Csikász-Nagy et al. 2006). The basic idea behind these papers is that a eukaryotic cell progresses through the DNA replication–division cycle by a series of transitions (G1/S, G2/M, M/G1) that correspond to bifurcations of the underlying molecular control system. Before each transition, the cell is arrested in a stable steady state of the dynamical system that corresponds to a particular physiological state: G1-arrest, G2-arrest, or metaphase-arrest. To pass to the next stage of the cell cycle, the stable arrested state must be lifted, either by annihilation (at a SN or SNIPER bifurcation) or by losing stability (at a Hopf bifurcation). To prevent a transition, e.g., if there is some damage to the DNA or some problem in aligning chromosomes on the metaphase plate, then a "checkpoint" mechanism stabilizes the arrested state by moving the bifurcation point to some higher value of the progression signal(s). For example, a schematic diagram of the fission yeast cell cycle is provided in Fig. 4.

During early embryogenesis, from fertilization to the mid-blastula transition, mitotic cycles proceed rapidly and synchronously, without checkpoint controls. In this case, the DNA replication–division cycles seem to be driven by spontaneous limit-cycle oscillations. For more details, see the entry ▶ Cell Cycle Dynamics, Bistability and Oscillations.



Cell Cycle Model Analysis, Bifurcation Theory, Fig. 4 A schematic diagram of the fission yeast cell cycle. The bifurcation diagram in Fig. 3c is interpreted here as a signal-response curve relating cyclin-dependent kinase (CDK) activity to cell growth. CDK is a protein kinase that governs progression through the cell cycle. In fission yeast, low CDK activity corresponds to a G1/S/G2 state and high activity to mitosis. Cell size can be thought of as a bifurcation parameter: Cell size increases slowly as the cell grows, and the CDK control adapts quickly to an attractor of the vector field at the current size of the cell. The red curve is a "cell cycle trajectory." At the size of a newborn cell (size = 1), the only attractor is a stable steady state of low CDK activity. After a brief G1 period, the cell replicates its DNA and then pauses in G2 phase until it grows large enough to surpass the SNIPER bifurcation. The bifurcation point is the "critical size" for the G2/M transition in fission yeast. The dynamical system is attracted to a large amplitude limit cycle, which carries the cell into mitosis (CDK increasing). The cell exits mitosis when CDK activity is destroyed, and this is the signal for the cell to divide. Cell size is abruptly halved, and the newborn cells (each of size = 1) are attracted to the stable G1/S/G2 steady state. Notice that the cell cycle time (the time required to progress around the red loop) is identical to the mass-doubling time (the time necessary to grow from birth size = 1 to division size = 2)

Cross-References

- ▶ Cell Cycle Dynamics, Bistability and Oscillations
- ▶ Cell Cycle Dynamics, Irreversibility
- ▶ Cell Cycle Modeling, Differential Equation
- ▶ Cell Cycle Models, Sensitivity Analysis
- ▶ Cell Cycle of Early Frog Embryos
- ▶ Cell Cycle, Budding Yeast
- ▶ Cell Cycle, Fission Yeast

References

Andronov AA, Vitt AA, Khaikin SE (1966) Theory of oscillators. Pergamon, London (first published in Russian in1937)
Borisuk MT, Tyson JJ (1998) Bifurcation analysis of a model of mitotic control in frog eggs. J Theor Biol 195:69–85
Csikász-Nagy A, Battogtokh D, Chen KC, Novák B, Tyson JJ (2006) Analysis of a generic model of eukaryotic cell-cycle regulation. Biophys J 90:4361–4379
Goldbeter A (1996) Biochemical oscillations and cellular rhythms. Cambridge University Press, Cambridge
Kuznetsov YA (2004) Elements of applied bifurcation theory, Thirdth edn. Springer, New York
Odell GM (1980) Qualitative theory of systems of ordinary differential equations, including phase plane analysis and the use of the Hopf bifurcation theorem. In: Segel LA (ed) Mathematical models in molecular and cell biology. Cambridge University Press, Cambridge
Poincaré HJ (1899) Les méthodes nouvelles de la mécanique céleste, vol 1–3. Gauthiers-Villars, Paris (English translation by D. Goroff, 1993, American Institute of Physics, New York)
Strogatz SH (1994) Nonlinear dynamics and chaos. Addison-Wesley, Reading
Thom R (1989) Structural stability and morphogenesis. Addison-Wesley, Reading (first published in French in1972)
Tyson JJ, Chen KC, Novak B (2003) Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. Curr Opin Cell Biol 15:221–231

Cell Cycle Modeling Using Logical Rules

Adrien Fauré[1] and Denis Thieffry[2]
[1]Graduate School of Science and Engineering, Yamaguchi Univesity, Yamaguchi, Japan
[2]IBENS - UMR ENS - CNRS 8197 - INSERM 1024, Paris, France

Synonyms

Boolean modeling of cell cycle control; Logical modeling of cell cycle control

Definition

This essay provides a short overview of recent applications of Boolean and multilevel logical approaches to cell cycle modeling.

Characteristics

Introduction

Extensively studied, the intricate regulatory circuits controlling cell cycle (▶ Cell Cycle) in Eukaryotes constitute a daunting challenge and a recurrent reference in systems biology. Although quantitative data may be sometimes more easily obtained at the system level (as it is the case with cell cycle time length or mass at division), cell cycle progression is essentially described as a succession of qualitatively different phases. The phases G1, S, G2, M, and G0 (and subphases) can be characterized by the qualitative level of key markers, most particularly the different cyclins (▶ Cyclins and Cyclin-Dependent Kinases). Furthermore, mutant phenotypes observations are often limited to qualitative assessment of viability or arrest in a specific phase.

Consequently, the *logical formalism* (▶ Logical Model) has been recurrently used to build qualitative dynamical models of the regulatory networks controlling cell cycle in different organisms. Notably, as budding yeast (▶ Cell Cycle, Budding Yeast) and fission yeast (▶ Cell Cycle, Fission Yeast) have served as model organisms for pioneering experimental studies of cell cycle control, they constitute reference systems for many modeling studies.

Phenomenological models can already bring interesting insight into biological mechanisms. For example, the Boolean model published by Bähler and Svetina (2005) focuses on the different growth modes displayed by the fission yeast. Mutants are then defined as deletions of specific arrows in the model and compared to known mutations that affect growth modes. Such phenomenological understanding hints toward the underlying molecular mechanism.

In the recent years, several groups have focused on the development of predictive logical models for the molecular regulatory networks controlling cell cycle in budding yeast and fission yeast (Li et al. 2004; Bornholdt 2008; Irons 2009; Fauré et al. 2009). Several of these models have been used to assess novel simulation methods or model-building strategies, or yet to address theoretical questions regarding *modularity* (Irons 2009; Fauré et al. 2009) or dynamical *robustness* (Mangla et al. 2010). Globally, the success of logical approaches at reproducing complex wild-type and mutant phenotypes demonstrates that the dynamics of the cell cycle can be captured independently of quantitative assumptions on kinetic parameters (Bornholdt 2008; Fauré and Thieffry 2009).

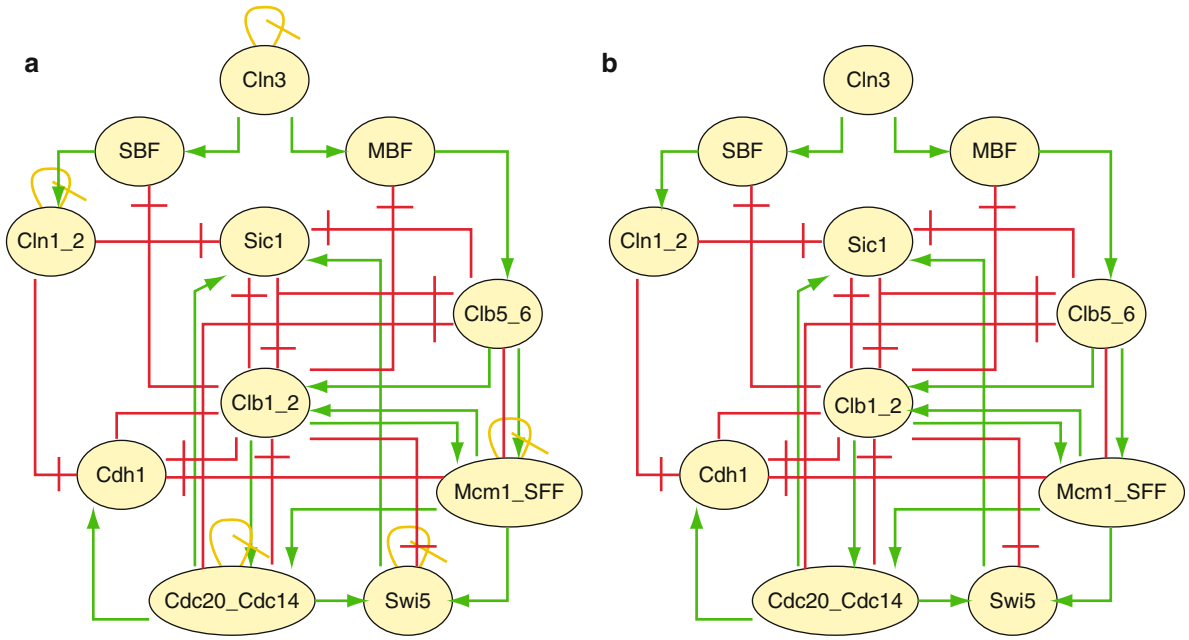
In what follows, we start by illustrating the logical approach through its application to the definition and the analysis of a simplified model of the regulatory network controlling the cell cycle in budding yeast. This reference model is used to emphasise variations in the definition of logical rules or yet regarding the use of different updating assumptions. The following section introduces more recent, prospective studies devoted to mammalian cell cycle control. The essay ends by a brief outlook section.

Boolean Modeling of the Molecular Network Controlling Budding Yeast Cell Cycle

The focus of published studies ranges from molecular regulatory networks to phenomenological models, the different scales being sometimes integrated in the same network. For example, some models explicitly incorporate variables representing cell cycle phases (Irons 2009), while others implicitly deduce them from the level of key variables (Li et al. 2004; Fauré et al. 2009).

Accordingly, as discussed by Bähler and Svetina (2005), there is no direct relationship between regulatory circuits and the molecular processes they represent. The same type of regulatory arc may be used to represent different processes, such as transcriptional activation of a gene, or post-translational modification of a protein. Conversely, similar processes may be represented differently, in particular those involving the representation of mass flow, such as multiproteic complex formation (Fauré and Thieffry 2009).

Regarding the definition of the logical rules associated with each regulatory component, two main approaches are used. A first approach consists in simply summing the positive and negative influences on each node (sometimes considering different weights for different edges of the network); depending on whether this sum lies below, above, or at a given threshold, the value of the variable will tend to decrease, increase, or remain unchanged, respectively. This formalism was used in the seminal Boolean model shown in Fig. 1, left. Using a generic summation rule with a threshold set to zero, the

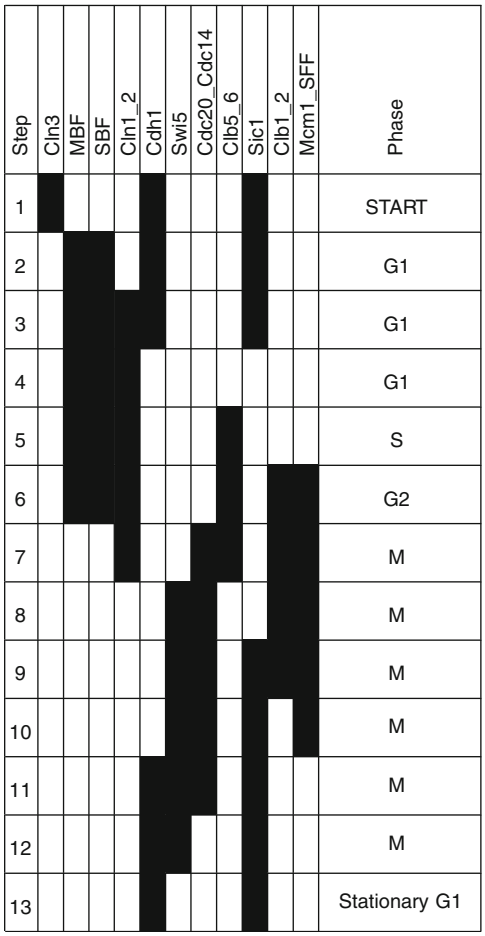


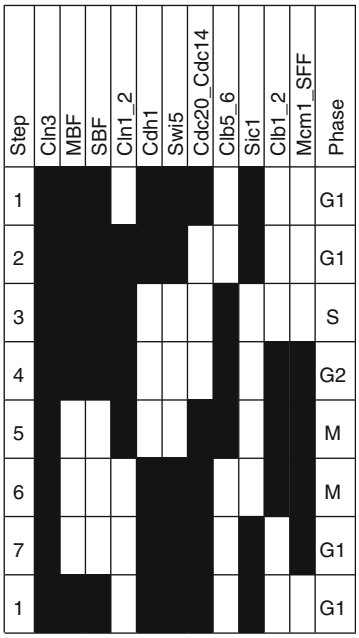
Cell Cycle Modeling Using Logical Rules, Fig. 1 Boolean models of the Budding yeast cell cycle. *Left*: Original Boolean model presented in (Li et al. 2004). This model was simulated synchronously (see Fig. 2) using a generic summation rule. In this framework, self-inhibiting arrows had to be introduced to account for degradation of components that only have positive regulators. *Right*: Boolean model adapted from the previous one using specific logical formulae (see Table 1). Both graphs have been drawn using GINsim (Naldi et al. 2009). Green normal arrows stand for activation, red T-arrows stand for inhibition, yellow T-arrows represent the "self-degradation" loops introduced in (Li et al. 2004)

authors had to include "self-degradation" loops (in yellow in Fig. 1, left), which do not represent true auto-inhibitory processes. Using a synchronous (deterministic) updating strategy, they identified seven stable states as the sole attractors of the system and further established that the large majority of Boolean states lead to the same stable state, which corresponds to the G1 phase. Furthermore, starting with initial conditions matching the entry into cell cycle, as shown in Fig. 2 (left), the system follows a trajectory consistent with the succession of molecular events and phases along budding yeast cell cycle.

A second approach consists in defining a specific logical rule for each node in the network. Fig. 1 (right) shows a transposition of the model by Li et al. (2004) into a logical regulatory graph, while Table 1 lists the logical rules associated with each component of the network. Notice that the network is somewhat simplified (elimination of the auto-inhibitory loops), while we have now different types of regulatory rules combining NOT, AND, and OR logical operators. In the absence of Cln3, this model has a unique stable state corresponding to the G0 phase. Starting with initial conditions enabling the initiation of the cell cycle (with Cln3 ON), this simplified model yields a cyclic attractor, which recapitulates the sequence of events along the cell cycle (see Fig. 2, right).

Using this second approach, several groups have designed more sophisticate models encompassing additional regulatory components (e.g., checkpoints (▶ Cell Cycle Checkpoints)) and recapitulating the behavior of the systems following various kinds of perturbations (e.g., single or multiple loss-of-function mutations) (Irons 2009; Fauré et al. 2009). Interestingly, these studies show that various kinetic assumptions lead to similar consistent dynamical behaviors, indicating that the control network is sufficiently robust to cope with stochastic fluctuations of synthesis/degradation rates. The multi-level model presented by Fauré et al. is arguably the most comprehensive logical cell cycle model to date.





Cell Cycle Modeling Using Logical Rules, Fig. 2 *Left*: simulation of the original model by Li et al. (2004) (cf. Figure 1, *left*). Starting from an initial state corresponding to G1, simulation yields a sequence of states that follows the normal course of the cell cycle through the different phases G1, S, G2, and M, until reaching a stable G1 state. The initial state corresponds to an "excited" G1 stable state, with the Cln3 variable set to 1. *Right*: simulation of the revised model, with the input Cln3 always ON (cf. Figure 1, right). Black cells denote activity of the corresponding components, white cells denote inactivity

Cell Cycle Modeling Using Logical Rules, Table 1 Specific logical rules associated with the revised version of the model by Li et al. (2004) presented in Figure 1 (*right*). "&", "|", and "!" stand for the Boolean operators AND, OR, and NOT, respectively

Components	Logical rules
MBF	Cln3 & !Clb1_2
SBF	Cln3 & !Clb1_2
Cln1_2	SBF
Cdh1	!(Cln1_2 | Clb5_6 | Clb1_2) | Cdc20_Cdc14
Swi5	((Mcm1_SFF | Cdc20_Cdc14) & !Clb1_2) | (Mcm1_SFF & Cdc20_Cdc14 & Clb1_2)
Cdc20_Cdc14	Clb1_2 | Mcm1_SFF
Clb5_6	(MBF | !Sic1) & !Cdc20_Cdc14
Sic1	(Swi5 & Cdc20_Cdc14) | !(Cln1_2 | Clb5_6 | Clb1_2)
Clb1_2	!(Cdh1 | Sic1 | Cdc20_Cdc14) | (Clb5_6 & Mcm1_SFF)
Mcm1_SFF	Clb1_2 | Clb5_6

Toward the Modeling of Mammalian Cell Cycle Networks

Proper modeling of the regulatory network controlling mammalian cell cycle remains a daunting challenge. Most published models concentrate on the control of the G1-S transition (see Fauré et al. 2006, for a generic Boolean model). Various groups are currently developing models of human cell cycle in connection with cancer studies, with the aim of identifying intervention points to block uncontrolled proliferation. One difficulty is that the actors to consider depend to some extent on the cellular context (cell type, environmental signals, presence of mutations).

For example, Sahin et al. developed a Boolean model of the control of the G1-S transition by the EGF pathway to predict potential targets for the treatment of trastuzumab-resistant breast cancer (Sahin et al. 2009). Predictions yielded by loss-of-function

simulations were assessed using RNA interference and proteomic (RPPA) essays. A refined version of the model was then derived to cope with observed inconsistencies.

In order to account for the spatial aspects of tumor formation, logical models of regulatory networks can be combined with other formal frameworks to generate multiscale models. For example, a model for avascular tumor growth has been proposed (Jiang et al. 2005), where the upper, extracellular layer consists of reaction-diffusion equations for growth and inhibitory factors; the middle, cellular layer is a lattice Monte Carlo model (▶ Monte Carlo Simulation); and the lower layer is a simplified Boolean model of the regulatory network that controls the G1-S transition simulated synchronously. To each updating corresponds a fixed amount of time, during which simulation of the upper levels proceeds proportionately. This integration of a simple logical model of the molecular cell cycle network with the other two layers proved accurate enough to match observed upper-level tumor properties. Furthermore, based on this model, the authors where able to infer that tumor cell quiescence is more likely due to failure to enter the S phase rather than to mechanical constraints exerted by the surrounding cells.

Outlook

Arguably, logical approaches are well suited to cope with large regulatory networks, in particular when most available data are qualitative or semi-quantitative. Furthermore, consistent logical models can serve as templates to define more quantitative models using more differential or stochastic formalisms, as exemplified by the recent modeling study devoted to DNA repair mechanism by Abou-Jaoudé et al. (2009). Although logical models have yielded only limited impact onto experimental studies, this situation should change in the future as collaborations between modelers and experimentalists intensify.

Cross-References

- ▶ Cell Cycle
- ▶ Cell Cycle Checkpoints
- ▶ Cell Cycle, Budding Yeast
- ▶ Cell Cycle, Fission Yeast
- ▶ Cyclins and Cyclin-dependent Kinases
- ▶ Logical Model
- ▶ Monte Carlo Simulation

References

Abou-Jaoudé W, Ouattara DA, Kaufman M (2009) From structure to dynamics: frequency tuning in the p53-mdm2 network. I. logical approach. J Theor Biol 258(4):561–577

Bähler J, Svetina S (2005) A logical circuit for the regulation of fission yeast growth modes. J Theor Biol 237(2):210–218

Bornholdt S (2008) Boolean network models of cellular regulation: prospects and limitations. J R Soc 5(Suppl 1):S85–S94

Fauré A, Thieffry D (2009) Logical modelling of cell cycle control in eukaryotes: a comparative study. Mol Biosyst 5:1569–1581

Fauré A, Naldi A, Chaouiya C, Thieffry D (2006) Dynamical analysis of a generic Boolean model for the control of the mammalian cell cycle. Bioinformatics 22(14):e124–e131

Fauré A, Naldi A, Lopez F, Chaouiya C, Ciliberto C, Thieffry D (2009) Modular logical modelling of the budding yeast cell cycle. Mol Biosyst 5:1787–1796

Irons DJ (2009) Logical analysis of the budding yeast cell cycle. J Theor Biol 257(4):543–559

Jiang Y, Pjesivac-Grbovic J, Cantrell C, Freyer JP (2005) A multiscale model for avascular tumor growth. Biophys J 89(6):3884–3894

Li F, Long T, Lu Y, Ouyang Q, Tang C (2004) The yeast cell-cycle network is robustly designed. Proc Natl Acad Sci USA 101(14):4781–4786

Mangla K, Dill DL, Horowitz MA (2010) Timing robustness in the budding and fission yeast cell cycles. PLoS ONE 5(2):e8906

Naldi A, Berenguier D, Fauré A, Lopez F, Thieffry D, Chaouiya C (2009) Logical modelling of regulatory networks with GINsim 2.3. BioSyst 97(2):134–139

Sahin O, Fröhlich H, Löbke C, Korf U, Burmester S, Majety M, Mattern J, Schupp I, Chaouiya C, Thieffry D, Poustka A, Wiemann S, Beissbarth T, Arlt D (2009) Modeling erbb receptor-regulated g1/s transition to find novel targets for de novo trastuzumab resistance. BMC Syst Biol 3:1

Cell Cycle Modeling, Differential Equation

John J. Tyson
Department of Biological Sciences, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA

Synonyms

Kinetic equations; Rate equations; Reaction-diffusion equations

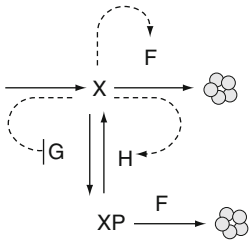
Definition

The physiological properties of cells – their growth, division, movement, signaling, metabolism, differentiation, and death – are all controlled by gene-protein regulatory networks of considerable complexity. Thanks to the revolutionary advances of molecular genetics in the latter part of the twentieth century, much is known now about the genes and proteins that constitute these networks and about their interactions, as well as the meso-scale topology of particular regulatory networks and the global topology of genome-wide surveys of gene, mRNA and protein interactions. The analysis of genome-wide ("omics") data is still very much dominated by statistical methods, but at the local and meso-scale of network complexity it is possible to build detailed, accurate, and predictive models of the dynamics of network behavior by using differential equations.

The applicability of differential equations to modeling network dynamics in general, and cell cycle regulation in particular, is based on the following logic (Tyson et al. 2001). A molecular regulatory network can be described, depending on the level of detail available from experimental investigations, by (1) a system of biochemical reactions or (2) an "influence" diagram or (3) a hybrid of the two types. These types of descriptions are illustrated in Fig. 1. Realistic networks are, of course, much more complex than the example given. Whether the network is described by chemical *reactions* or *influences* or a *hybrid* of the two, the network diagram is trying to tell us how one biochemical species is affecting the rates of production or removal of another species. As such, the network diagram can be converted into a set of ordinary differential equations (ODEs), one ODE for each time-varying biochemical concentration:

$$\frac{\mathrm{d}[\mathrm{X}_i]}{\mathrm{d}t} = \sum_j P_{ij}\left([\mathrm{X}_{\ldots}], [\mathrm{M}_{\ldots}], k_j\right) - \sum_l R_{il}([\mathrm{X}_{\ldots}], [\mathrm{M}_{\ldots}], k_l) \qquad (1)$$

In this equation, $[\mathrm{X}_i]$ = concentration of species i, P_{ij} = rate of the j-th reaction that produces species i, $[\mathrm{X}_{\ldots}]$ = concentrations of the time-varying biochemical species that participate in each of these reactions, $[\mathrm{M}_{\ldots}]$ = constant concentrations of the time-invariant biochemical species ("modifiers") that participate in each of these reactions (e.g., the total concentration of the enzyme that catalyzes reaction j), and k_j = the rate constant(s) needed to express the material flux through reaction j. Similarly, R_{il} = rate of the l-th reaction that removes species i, etc.



Cell Cycle Modeling, Differential Equation, Fig. 1 A typical molecular regulatory network. In this diagram, letters denote chemical species (proteins) and solid arrows represent chemical reactions transforming substrate(s) into product(s). A letter sitting next to an arrow denotes the enzyme catalyzing the reaction. Dashed arrows represent "influences" of one protein on another. Barbed arrows denote "activation" and blunt arrows denote "inhibition." The dynamics of this reaction network is represented by the system of five ODEs in Eq. 2

Regulatory networks like Fig. 1 are sometimes called "wiring" diagrams, in analogy to the schematic diagrams of electrical devices. Just like the dynamical behavior of an electrical device can be predicted (in practice) from Kirchoff's Laws (ODEs for the voltages at various points in the circuit), so the dynamics of a molecular regulatory network can be predicted (in principle) from ODEs (Eq. 1). Unfortunately, the analogy to electrical engineering goes no further. For an electrical device, we can obtain a schematic wiring diagram from the manufacturer, as well as a specification of the numerical values of the parameters that characterize each component (resistances, capacitances, etc.). For the living cell, we must guess the wiring diagram, and we must estimate the rate constants from the very experiments we are trying to explain. In essence, we must "reverse engineer" the cell's circuitry by performing well-designed experiments to probe the input-output (signal-response) characteristics of the cell under normal and contrived conditions (including mutations which scramble the wiring diagram in controlled ways).

For the many ways that differential equations have been used in molecular and cell biology, see the classic books by Edelstein-Keshet (1988), Murray (1989), Goldbeter (1996), Fall et al. (2002), and Keener and Sneyd (2009).

Characteristics

Wiring Diagrams and Rate Equations

Figure 1, which will serve as our example of modeling by ODEs, can be interpreted as a model of MPF dynamics in a fertilized egg (▶ Cell Cycle Dynamics, Bistability and Oscillations.) In this example, X = MPF = dimer of Cdk1 and cyclin B, XP = preMPF = phosphorylated (inactive) form of MPF, G = Wee1 = kinase that inactivates MPF, H = Cdc25 = phosphatase that converts preMPF into active MPF, F = APC/Cdc20 = ubiquitin-ligase that labels cyclin B for proteolysis. The production and removal of X and XP are described by chemical reactions (synthesis, degradation, phosphorylation, dephosphorylation), and kinetic equations for the rates of change of X and XP can be written by standard principles of biochemical kinetics:

$$\frac{dX}{dt} = k_{sx} - k_{dx}FX - \frac{k_g GX}{K_{mg} + X} + \frac{k_h HX_p}{K_{mh} + X_p}$$
$$\frac{dX_p}{dt} = -k_{dx}FX_p + \frac{k_g GX}{K_{mg} + X} + \frac{k_h HX_p}{K_{mh} + X_p} \quad (2a, b)$$

In each case, the rate of a reaction is given either by the law of mass action (for synthesis and degradation reactions) or by the Michaelis-Menten rate law (for the phosphorylation and dephosphorylation reactions). (Which rate law we use for an enzyme-catalyzed reaction depends on whether the enzyme tends to work in its "linear" regime or in its "saturated" regime.) Regulation of the enzymes, F, G, and H in Fig. 1, is only specified as "influences": X "activates" F and H, and X "inhibits" G. We choose to describe these influences by generic ODEs:

$$\frac{dF}{dt} = \lambda_f[\Psi(\sigma_f \cdot \{\omega_{f0} + \omega_{f1}X\}) - F]$$
$$\frac{dG}{dt} = \lambda_g[\Psi(\sigma_g \cdot \{\omega_{g0} + \omega_{g1}X\}) - G] \quad (2c, d, e)$$
$$\frac{dH}{dt} = \lambda_h[\Psi(\sigma_h \cdot \{\omega_{h0} + \omega_{h1}X\}) - H]$$

where $\Psi(\xi) = 1/(1 + e^{-\xi})$ is a "soft Heaviside" function; $\Psi(\xi)$ varies smoothly from 0 for $\xi << -1$ to 0.5 for $\xi = 0$ to +1 for $\xi >> 1$. According to Eq. 2c, d, e, F, G, and H are continually changing to keep up with the soft Heaviside functions, which are changing in response to the dynamical variable X. In return, the dynamical variables X and X_P are changing in response to F, G, and H according to Eq. 2a, b. It would be impossible to keep track of the implications of all these changes in one's head; it is the job of the ODEs to track the variables for us.

Numerical Simulation of ODEs: Parameter Values and Initial Conditions

Before we can solve the ODEs (Eq. 2) we must specify numerical values for all the "parameters" (rate constants, Michaelis constants, ω's and σ's); see Table 1. Usually, these parameters must be estimated from experimental data, but in this example we assign values to illustrate some interesting and physiologically relevant solutions of the ODEs. In addition to parameter assignments, we must also specify "initial conditions" (values at $t = 0$) for the five time-varying species: $X(0)$, $X_P(0)$, $F(0)$, $G(0)$, $H(0)$; see Table 2.

With this information, we can now instruct a computer, using the ODEs (Eq. 2), to take small time steps, dt, and update the values of the dynamical variables:

$$X(t + dt) = X(t) + \left[k_{sx} - k_{dx}X(t) - \frac{k_g GX(t)}{K_{mg} + X(t)} + \frac{k_h HX_P(t)}{K_{mh} + X_P(t)}\right] \cdot dt$$
$$X_P(t + dt) = X_P(t) + \left[-k_{dx}X_P(t) + \frac{k_g GX(t)}{K_{mg} + X(t)} - \frac{k_h HX_P(t)}{K_{mh} + X_P(t)}\right] \cdot dt$$

etc. (3a, b, c, d, e)

The computer starts at $t = 0$, with the given initial conditions, computes the instantaneous rates of change (the functions in [. . .] above), and then uses Eq. 3 to compute the values of the five dynamical variables at $t = 0 + dt$. The computer then repeats the process to get the values of the dynamical variables at $t = 2dt$, $3dt$, etc. For dt small enough, this procedure gives an accurate numerical solution of the ODEs. Of course, there are more sophisticated and efficient algorithms for solving nonlinear ODEs, but they are all based on the fundamental procedure just described.

In Fig. 2, we present numerical simulations of ODEs (Eq. 2) for the parameter values in Table 1, with modifications given in the figure legend. For the case in Fig. 2a, the ODEs have a single stable steady

Cell Cycle Modeling, Differential Equation, Table 1 Parameter values for the simulations in Fig. 2

$k_g = k_h = 10$		$K_{mg} = K_{mh} = 0.05$	$\lambda_f = \lambda_g = \lambda_h = 1$		$\omega_{f1} = 1$
$\sigma_g = 3$	$\omega_{g0} = 0.2$	$\omega_{g1} = -0.7$	$\sigma_h = 3$	$\omega_{h0} = -0.2$	$\omega_{h1} = 0.8$
Fig. 2a	*Fig. 2b*	*Fig. 2c*	*Fig. 2a*	*Fig. 2b*	*Fig. 2c*
$k_{sx} = 0.04$	$k_{sx} = 0.1$	$k_{sx} = 0.1$	$k_{dx} = 1$	$k_{dx} = 1$	$k_{dx} = 0.5$
$\sigma_f = 20$	$\sigma_f = 20$	$\sigma_f = 5$	$\omega_{f0} = -0.3$	$\omega_{f0} = -0.3$	$\omega_{f0} = -0.4$

Cell Cycle Modeling, Differential Equation, Table 2 Initial conditions for the simulations in Fig. 2

	X	X_P	F	G	H
Fig. 2a	0.1149	1.5459	0.0241	0.5887	0.4197
Fig. 2b	0.1036	1.0602	0.0181	0.5961	0.4111
Fig. 2c(low)	0.1058	0.9649	0.1868	0.5933	0.4143
Fig. 2c(med)	0.1946	0.526	0.318	0.544	0.471
Fig. 2c(high)	0.3327	0.1472	0.4167	0.4753	0.5495

state solution. In Fig. 2b the steady state solution is unstable and the system of ODEs exhibits spontaneous oscillations of all the variables. In Fig. 2c, the system exhibits a phenomenon called "bistability," i.e., two stable steady states separated by an unstable steady state.

Analysis of Nonlinear Ordinary Differential Equations

Why does the system of nonlinear ODEs in Eq. 2 show the many different sorts of behavior illustrated in Fig. 2? Might the system show other, qualitatively different sorts of behavior? For what values of the parameters are each of the types of solutions expected? The answers to these sorts of questions are provided by bifurcation theory, which is described in the article "cell cycle model analysis, bifurcation theory."

Parameter Estimation

If we have experimental measurements of some of the dynamical variables at a sequence of time points, under a variety of experimental conditions, both natural and contrived (e.g., in mutant cells), then it is sometimes possible to estimate the parameters in a dynamical model (and to test the adequacy of the wiring diagram) by least-squares fitting of numerical simulations of the model to the experimental data. For instance, the curves in Fig. 2b look very much like the measurements of Pomerening et al. (2005; their Fig. 1V). However, even for a modest network such as our example, with five dynamical variables and 18 parameters, fitting simulations to experimental data can be a very difficult task. It requires careful choice of experimental conditions and sophisticated methods of searching the parameter space. A lead-in to this extensive literature is provided by Apgar et al. (2010).

Alternative Modeling Strategies

In this chapter, we have discussed modeling by nonlinear ODEs, assuming that the cell is a well-mixed chemical reactor. Surprisingly, in some cases this is not a bad approximation. For example, the time it takes for a typical protein to diffuse across a yeast cell (diameter ~5 μm) is only about 10 s, which is very short compared to the interdivision time (at least 90 min) of a yeast cell. Hence, for models of the yeast cell cycle, the cytoplasm is essentially well-mixed. Of course, one might want to distinguish between nuclear and cytoplasmic compartments, but this situation can be handled with nonlinear ODEs by including fluxes of components into and out of the nucleus.

In other situations, where the time scale is shorter and/or the space scale is larger, one must take into account the coupling of local chemical reactions with molecular diffusion (and possibly vectorial transport processes, e.g., along microtubules). In these cases, the correct modeling approach might be partial differential equations.

In some cases, when very little is known about the underlying biochemistry of a control system, systems biologists prefer to model the system with a Boolean network, an approach described in the article on "cell cycle modeling, logical rules."

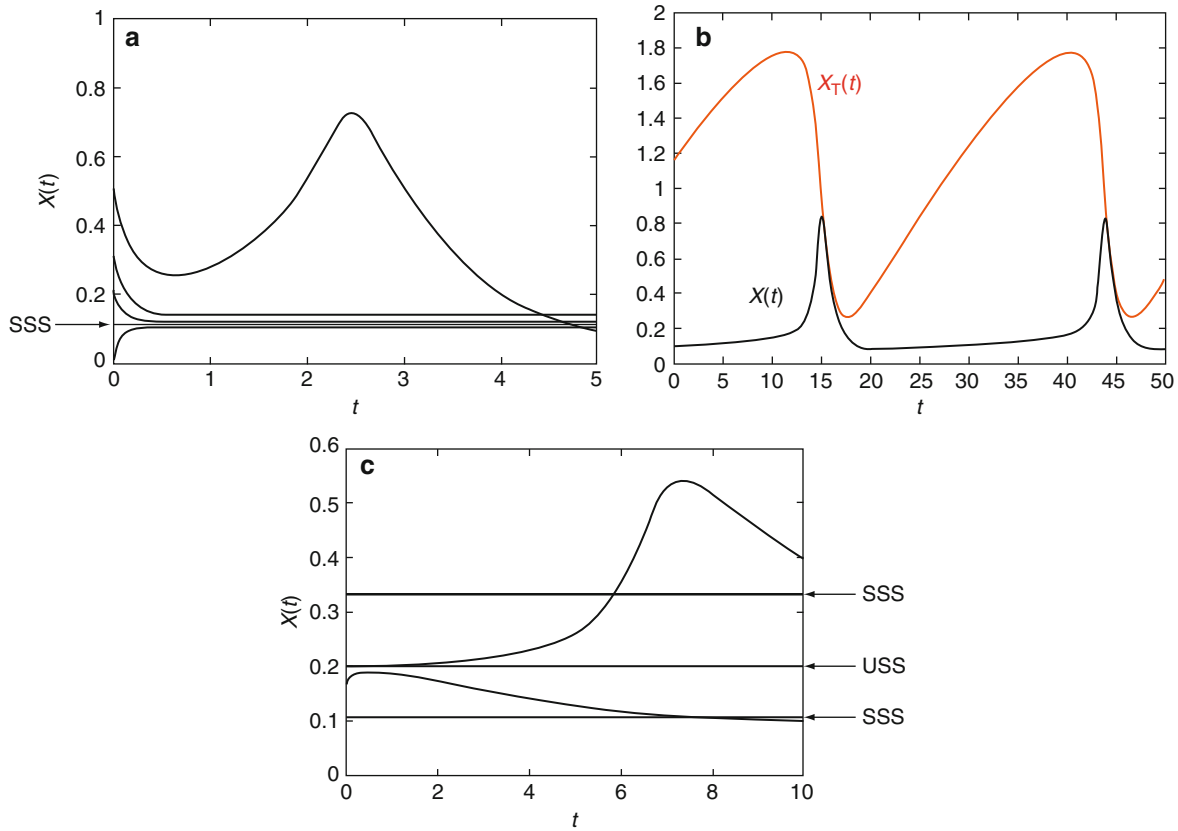
Software for Dynamic Modeling

There are several convenient software packages for modeling molecular regulatory networks with differential equations:

Copasi www.copasi.org

XPP http://www.math.pitt.edu/~bard/xpp/xpp.html

Madonna http://www.berkeleymadonna.com/download.html



Cell Cycle Modeling, Differential Equation, Fig. 2 Representative simulations of ODEs (Eq. 2). The parameter values and initial conditions for these simulations are given in Tables 1 and 2. (**a**) A unique stable steady state (sss). Small perturbations away from the steady state return immediately. A larger perturbation, $X(0) = 0.5$, exhibits a transient pulse of MPF activity before returning to the steady state. This behavior is called "excitability." (**b**) A stable limit cycle oscillation. In addition to $X(t) = [\text{MPF}]$, we also plot $X_T(t) = X(t) + X_P(t) = [\text{total cyclin}]$. The period of oscillation is 29 min. Compare to Fig. 1V in Pomerening et al. (2005). (**c**) Two coexisting stable steady states separated by an unstable steady state (uss)

Reaction-diffusion modeling with partial differential equations is best done by

Virtual Cell http://www.nrcam.uchc.edu/

Cross-References

- ▶ Cell Cycle Dynamics, Bistability and Oscillations
- ▶ Cell Cycle Dynamics, Irreversibility
- ▶ Cell Cycle Model Analysis, Bifurcation Theory
- ▶ Cell Cycle Modeling Using Logical Rules
- ▶ Cell Cycle Modeling, Stochastic Methods
- ▶ Cell Cycle Models, Sensitivity Analysis
- ▶ Cell Cycle of Early Frog Embryos
- ▶ Cell Cycle, Budding Yeast
- ▶ Cell Cycle, Fission Yeast

References

Apgar JF, Witmer DK, White FM, Tidor B (2010) Sloppy models, parameter uncertainty, and the role of experimental design. Mol Biosyst 6:890–900

Edelstein-Keshet L (1988) Mathematical models in biology. Random House, New York

Fall CP, Marland ES, Wagner JM, Tyson JJ (2002) Computational cell biology. Springer-Verlag, Berlin

Goldbeter A (1996) Biochemical oscillations and cellular rhythms. Cambridge University Press, Cambridge

Keener J, Sneyd J (2009) Mathematical physiology, 2nd edn. Springer-Verlag, Berlin

Murray JD (1989) Mathematical biology. Springer-Verlag, Berlin

Pomerening JR, Kim SY, Ferrell JE Jr (2005) Systems-level dissection of the cell-cycle oscillator: bypassing positive feedback produces damped oscillations. Cell 122:565–578

Tyson JJ, Chen K, Novak B (2001) Network dynamics and cell physiology. Nat Rev Mol Cell Biol 2:908–916

Cell Cycle Modeling, Petri Nets

Ivan Mura
The Microsoft Research - University of Trento Centre for Computational and Systems Biology, Trento, Italy

Synonyms

Place/transition net

Definition

Petri Nets are a graphical modeling formalism with numerous applications to the modeling and analysis of systems composed by multiple concurrent processes, including the molecular kinetics of cell cycle regulation networks. They are named after Carl Adam Petri, who originally defined them in his dissertation thesis in 1962 (Petri 1962).

Petri Net models consist of four types of elements:

- **Places**, depicted as hollow circles, represent variables of the model.
- **Tokens**, depicted as black dots, are contained into places and provide the numerical value associated to the variable.
- **Transitions**, depicted as bars, represent events affecting the variables. The occurrence of the event associated to a transition is referred as the *firing* of the transition.
- **Arcs**, linking transitions to places and places to transitions (but not places to places nor transitions to transitions), carry multiplicities defining the changes on variables as a result of transitions firings. Incoming arcs to a place add tokens to the place, whereas outgoing arcs remove tokens.

The intuitive graphical formalisms of Petri Nets coupled with the considerable amount of theoretical results supporting their analysis (see for instance the work of Murata 1989) have favored the spread of this modeling formalism in various scientific and technical fields, among them being Systems Biology. There is quite an immediate mapping between reaction-oriented descriptions of biological systems and places, arcs, and transitions. We show in Fig. 1 a Petri Net model composed by four places, seven transitions, and 18 arcs, encoding the interactions between the APC molecule Cdh1 and the CycB-Cdk1 dimer. The modeled reactions are shown in Eqs. 1–5, and include two possible reversible bindings of Cdh1 with the cyclin dimer (Eqs. 1 and 2), two possible degradation ways of the Cdh1 molecule (Eqs. 3 and 4), and the CycB-Cdk1 catalysis reaction operated by Cdh1 (Eq. 5):

$$Cdh1 + CycB - Cdk1 \leftrightharpoons Cdh1 - CycB - Cdk1 \quad (1)$$

$$CycB - Cdk1 + Cdh1 \leftrightharpoons CycB - Cdk1 - Cdh1 \quad (2)$$

$$Cdh1 - CycB - Cdk1 \rightarrow CycB - Cdk1 \quad (3)$$

$$CycB - Cdk1 - Cdh1 \rightarrow CycB - Cdk1 \quad (4)$$

$$Cdh1 - CycB - Cdk1 \rightarrow Cdh1 \quad (5)$$

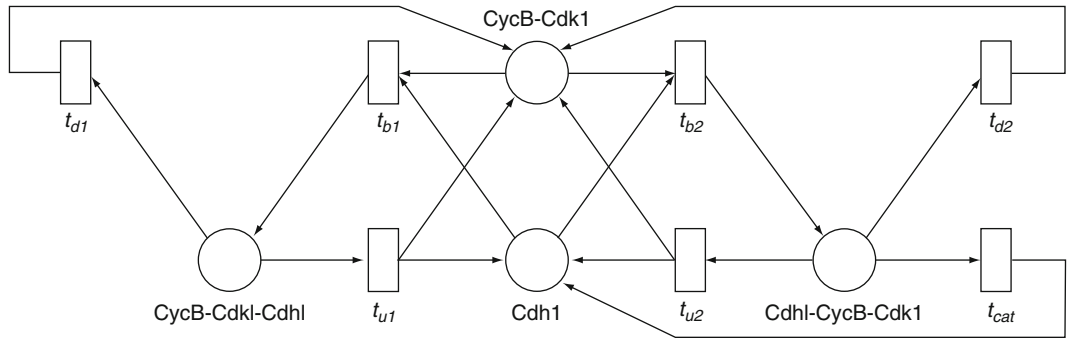
Tokens are not represented in the Petri Net model in Fig. 1, neither are arc cardinalities, which are all unitary.

A very good reference textbook to the applications of Petri Nets for the modeling of living systems is (Koch et al. 2010). We concisely describe in the following several features of Petri Nets, providing in particular details and references to their applications to the modeling of cell cycle regulation networks.

Characteristics

Semantics of Transitions Firing

The marking of a Petri Net is the amount of tokens contained in the places. It evolves as a consequence of transition firings, i.e., the occurrence of events modeled in a Petri Net. A transition of the model is said to be *enabled* if and only if each place connected to the transition by an input arc contains a number of tokens greater or equal to the multiplicity of the arc. An enabled transition may fire. Upon firing, the marking of each input place is updated by subtracting a number of tokens equal to the arc multiplicity. Simultaneously, the marking of each place connected to the firing transition by an output arc is updated by adding to it a number of tokens equal to the multiplicity of the connecting arc.



Cell Cycle Modeling, Petri Nets, Fig. 1 Petri net encoding of the interactions between Cdh1 and the CycB-Cdk1 dimer

Starting from the initial marking, the net can reach new markings, which represent the evolution of the state of the modeled variables. In our context, the firing of transitions usually corresponds to the occurrence of reactions and the changes in the marking of the places caused by the firing corresponds to the updates of the abundance of reactants and products of reactions.

It is quite immediate to realize that several concepts need a more precise definition to reach a complete characterization of the firing semantics. Consider for instance the case when multiple transitions are enabled: obviously, rules are needed for defining the order of firings, as different orders may result in different final markings of the net.

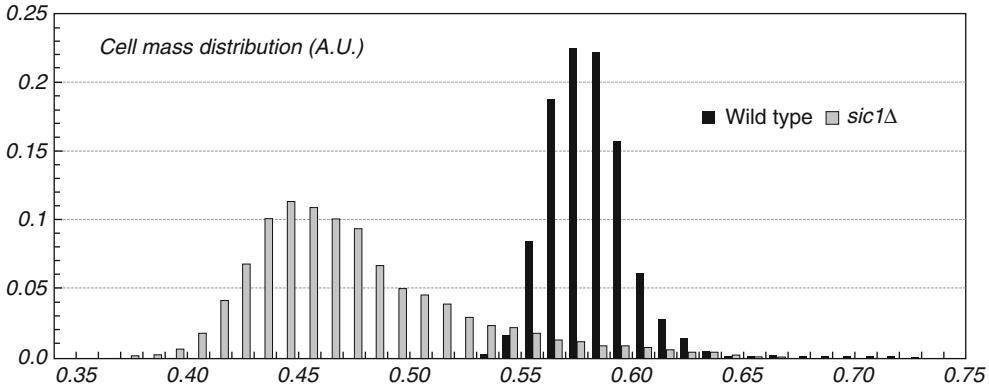
Qualitative Modeling with Petri Nets

The earliest definition of Petri Nets introduced a nondeterministic choice of the firing, which is the same as to say that each order is possible, compatibly with the enabling rules of the firing. Basically, a nondeterministic firing policy assigns an equal occurrence priority to each transition, irrespective of the time dimension of net evolution, as determined by the relative speed with which reactions take place in the modeled system. This type of firing semantics is useful when no quantitative information is available about the speed of reactions, and can nevertheless provide a view of the possible markings a given Petri Net model can reach. Thus, it can help in showing that some features of the model exist no matter which speed the real kinetics of reactions possesses. An example of this Petri Net qualitative modeling can be found in (Sackmann et al. 2006), where the mating pheromone response pathway of budding yeast, which is responsible for cell cycle arrest in G1 phase, is studied through the techniques of p-invariant and t-invariant analysis (Murata 1989).

Petri Nets with Stochastic Firing Times

The quantitative information about the speed of reactions is introduced into Petri Net models by various extensions of the firing semantics. The first one we consider here proposes a stochastic formulation of reaction occurrence times (Marsan 1990), where the competition among concurrently enabled transitions is broken down through a random choice modulated by the speed of the competing transitions. More precisely, each transition firing occurs with a time that follows a probability distribution law, and in a set of simultaneously enabled transitions the smallest random time determines which transition will fire first. This firing semantics is called *race policy*. In the case of continuous-time distributions of the firing times, no simultaneous firings of transitions are thus allowed. Petri Nets adhering to this semantics of firing are commonly called Stochastic Petri Nets.

Of particular interest in Systems Biology are the Stochastic Petri Nets whose firing times are distributed as negative exponential ▶ random variables, in which case the marking evolution process over time is equivalent to a discrete-space continuous-time Markov process (▶ Cell Cycle Modeling, Stochastic Methods; ▶ Stochastic Processes, Fokker-Planck Equation). This class of models lend themselves to naturally represent the stochastic molecular kinetics as described by Gillespie (1977), and their evolution can be studied through ▶ Gillespie stochastic simulation algorithm (Gillespie 1977).



Cell Cycle Modeling, Petri Nets, Fig. 2 Cell mass spread in wild type and *sic1*Δ mutant of budding yeast

Among the modeling works that resort to stochastic methods (▶ Cell Cycle Modeling, Stochastic Methods), we found in the literature a few applications of Stochastic Petri Nets to the modeling and analysis of yeast cell cycle. The work in (Mura and Csikász-Nagy 2008) provides a model of budding yeast (▶ Cell Cycle, Budding Yeast) cell cycle regulation network. The model validation is performed with a comparison of simulated results against a set of experimental results about wild type as well as budding yeast mutant cells. The authors demonstrate how the Stochastic Petri Net model can be used to obtain statistics of the key cell cycle parameters such as duration and cellular mass, beyond the average one. This information provides insights on cell cycle characteristics. For instance, Fig. 2 compares the simulated cell mass distribution of wild type budding yeast with that of the *sic1*Δ mutant, where the Cdk inhibitor of cyclins is removed. The chart clearly shows that the spread of the cell mass distribution is quite different in the two modeled organisms. The mutant mass distribution exhibits a higher variability, in agreement with experimental observations. Thus, the stochastic model reveals a characteristic of the phenotype of *sic1*Δ cells that could not be anticipated by deterministic models.

In (Csikász-Nagy and Mura 2010), Stochastic Petri Nets are used to explore the effects of intrinsic noise propagation on the variability of cell cycle duration and cell mass at division time in fission yeast (▶ Cell Cycle, Fission Yeast). In particular, the authors show that the variance of RNA synthesis and RNA degradation times of the key regulator Cdh1, and not only their average values, are critical in determining the overall variability of cell cycle statistics. This result provides an important argument for supporting the validity of pursuing a stochastic modeling of cell cycle regulation networks.

Hybrid Petri Nets

Hybrid Functional Petri Nets are another popular timed extension of Petri Nets that has been used to model cell cycle regulation. This modeling formalism allows the coexistence of both continuous and discrete processes. Both the set of places and that of transitions is divided in a discrete and a continuous part. Enabled discrete transitions fire after a determined delay, consuming and producing discrete numbers of tokens as in regular Petri Nets. Enabled continuous transitions fire continuously at a given rate. This latter feature provides a convenient abstraction mechanism for directly representing into the model processes that are difficult to model with a discrete number of tokens, such as very high molecular concentrations, as well as external driving variables such as cell mass.

The cell cycle regulatory network of fission yeast (▶ Cell Cycle, Fission Yeast) has been modeled and analyzed by Hybrid Functional Petri Nets in (Fujita et al. 2004). This study demonstrate the versatility of the Hybrid Functional Petri Net modeling formalism in representing a broad range of molecular regulation mechanisms and the intuitiveness of the modeling process. A comprehensive model of the

cell division process of Xenopus frog embryo (▶ Cell Cycle of Early Frog Embryos), which includes the molecular dynamics of cyclins and of the S-phase and M-phase promoting factors as well as the DNA damage (▶ Cell Cycle Arrest After DNA Damage) checkpoint, has been proposed in (Matsui et al. 2004). This latter model is able to reproduce the changes in the cell division cycles from synchronous to asynchronous.

Cross-References

- ▶ Cell Cycle Arrest After DNA Damage
- ▶ Cell Cycle Modeling, Stochastic Methods
- ▶ Cell Cycle of Early Frog Embryos
- ▶ Cell Cycle, Budding Yeast
- ▶ Cell Cycle, Fission Yeast
- ▶ Gillespie Stochastic Simulation
- ▶ Random Variable
- ▶ Stochastic Processes, Fokker-Planck Equation

References

Csikász-Nagy A, Mura I (2010) Role of mRNA gestation and senescence in noise reduction during the cell cycle. In Silico Biol 10:81–88. doi:10.3233/ISB-2010-0416

Fujita S, Matsui M, Matsuno H, Miyano S (2004) Modeling and simulation of fission yeast cell cycle on Hybrid Functional Petri Net. IEICE Trans Fundam 87:2919–2928

Gillespie DT (1977) Exact stochastic simulation of coupled chemical reactions. J Phys Chem 81:2340–2361

Koch I, Reisig W, Schreiber F (eds) (2010) Modeling in systems biology, vol 16, Computational biology. Springer, London. doi:10.1007/978-1-84996-474-6_7

Marsan MA (1990) Stochastic Petri nets: an elementary introduction. In: Advances in Petri nets 1989. Springer-Verlag, New York, pp 1–29

Matsui M, Fujita S, Suzuki S, Matsuno H, Miyano S (2004) Simulated cell division processes of the Xenopus cell cycle pathway by Genomic Object Net. J Integr Bioinform 1:3. doi:10.2390/biecoll-jib-2004-3

Mura I, Csikász-Nagy A (2008) Stochastic Petri Net extension of a yeast cell cycle model. J Theor Biol. doi:10.1016/j.jtbi.2008.07.019

Murata T (1989) Petri Nets: properties, analysis and applications. IEEE Proc 77:541–580

Petri CA (1962) Kommunikation mit automaten. Ph. D. Thesis, University of Bonn

Sackmann A, Heiner M, Koch I (2006) Application of Petri net based analysis techniques to signal transduction pathways. BMC Bioinform 7:482. doi:10.1186/1471-2105-7-482

Cell Cycle Modeling, Process Algebra

Alida Palmisano[1] and Corrado Priami[2]
[1]Department of Biological Sciences and Department of Computer Science, Department of Biological Sciences Virginia Tech, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA
[2]Microsoft Research-University of Trento Centre for Computational and Systems Biology and DISI, University of Trento, Povo, Trento, Italy

Synonyms

Algorithmic modeling languages; Computational modeling languages; Rule-based languages

Definition

Process algebras are formal languages originally conceived for modeling concurrent systems and are also a conceptual tool for the high-level description of interactions, communications, and synchronizations between a collection of independent processes. Concurrent computing systems do not allow assumptions on the relative speed of their components, and therefore, a certain degree of non-determinism emerges in systems behavior.

Several approaches have been developed and used to model and study complex interaction mechanisms in biological systems, and different process algebra–derived languages have been used, in particular, to build models of cell cycle at different levels of abstraction.

In seminal work, Regev and Shapiro (2001) suggested an abstraction of *cells-as-computation* and proposed that process algebras formerly used in the study of interacting computational entities could be usefully employed and extended to model biological processes (Table 1).

The strategy used by process algebras diverges from classical mathematical modeling because it can describe the flow of control between species and reactions, i.e., not only the time, but also the causality relation among the events that constitute the very essence of the mechanistic steps describing the emergent behavior of the modeled systems.

Cell Cycle Modeling, Process Algebra, Table 1 A concise picture of the mapping between biology and process algebras. In the process algebras interpretation, a biological entity (e.g., a protein) is seen as a computation unit (i.e., a process) with interaction capabilities abstracted as communication channel names. Similarly to biological entities, which interact/react through complementary capabilities, processes synchronize/communicate on communication channels complementary send/receive actions. The modifications/evolutions of molecules after reactions are represented by state changes following communications

Biology	Process algebras
Entity	Process
Interaction capability	Communication channel name
Interaction	Synchronization/communication
Modification/evolution	State change

This interpretation is very similar to programming the behavior of a system rather than describing only its outcome with respect to time (Priami 2009).

A number of process algebras have been adapted or newly developed for building biological models and performing stochastic simulations (i.e., ▶ Stochastic pi-calculus, ▶ BioSPI, ▶ κ-Calculus, ▶ Bio-PEPA, ▶ BlenX). For a comprehensive picture of the state of the art of process algebras abstraction for biology, see (Guerriero et al. 2009). It is important to notice that each modeling language requires a different modeling style because it maps molecular components to interactions, process communication, and process composition in a quite different way. So by choosing a formalism, a modeler is often making implicit choices (we refer the reader to (Calder and Hillston 2009) for details about the abstractions used by many of the aforementioned process algebras).

Characteristics

Process Algebras, the Approach

A system is represented by a collection of entities (processes) that define the possible behaviors of the components of the system. Algebras are equipped with an operational semantics, which consists in an inductive system of logical rules that implicitly define a state transition function (Plotkin 1981) and allow users to automatically derive the possible future states of the system. The fact that process P evolves into process Q is generally written as $P \rightarrow Q$.

The basic steps of process algebras are encoded by *actions*. In the simplest case, actions are input/output operations on communication channels. Actions can be composed sequentially, meaning that a process can perform an action after the other one ("a.Q"). Processes can also be composed in parallel: ("P1 | P2") meaning that P1 and P2 run simultaneously, although they can synchronize and communicate if they share a communication channel. Moreover, most of the algebras are equipped with some sort of restriction operator which defines the scope of actions. For instance, assuming "*a*" to be an action, and using "v" to denote restrictions, "P1 | v*a*.P2" means that action *a* is private to process P2.

The interaction policy assumed by each algebra is one of its main distinguishing features, and it drives the design of its primitive operators. This is a list of the most used ones:

- Two-way synchronization (see, e.g., [Milner 1999]). Each action has a complementary action, typically called co-action. Actions and co-actions can synchronize with each other. For instance, assuming a and $\bar{a}$ to be complementary actions, the following interaction would be possible: $a.P1|\bar{a}.P2 \rightarrow P1|P2$ where "." is the sequential operator.
- Multi-way synchronization (see, e.g., [Hoare 1985]). The parallel composition operator is parametric with respect to a set of actions (sometimes called cooperation set) on which all the parallel processes are obliged to synchronize. In this case, there is no need to resort to complementary actions. For example, $a.P1\big|_{\{a\}} a.P2\big|_{\{a\}} a.P3 \rightarrow P1\big|_{\{a\}} P2\big|_{\{a\}} P3.....$
- Name-passing (see, e.g., [Milner 1999]). In this case, actions have either the form $\bar{a}\langle b\rangle$ or the form $a(c)$. The first one stands for "output the channel name b along the channel named a," and the second one stands for "input any name from channel a and then use it instead of the parameter name c." The name-passing interaction policy is a specific instance of the two-way communication paradigm: The actions $\bar{a}\langle b\rangle$ and $a(c)$ are complementary to each other, and they can be involved in an interaction. In this case, more than simple synchronization occurs: Channel names flow from senders to receivers, thus dynamically changing the intercommunication topology of processes. For instance, $\bar{a}\langle b\rangle.P1|a(c).P2 \rightarrow P1|P2\{b/c\}$, where $\{b/c\}$

denotes the substitution of the free occurrences of c by the actually received name b.

Since names can be transmitted in interactions, the restriction operator plays a special role in name-passing algebras. Restricted names cannot be used as transmission media; they can, however, be used as transmitted data and, once transmitted, they become private resources shared by the sender and the receiver.

The mentioned operators are those common to various process algebras; in addition to these, each calculus adopts a few specific operators (see specific entries ▶ Stochastic pi-calculus, ▶ κ-Calculus, ▶ Bio-PEPA, ▶ BlenX for more details)

Process Algebras, Comparing Different Languages

Table 2 schematically reports a comparison between different process algebras with the availability of simulation platforms (usually implementing some variants of the ▶ Stochastic simulation algorithm) and the investigation of some complex case studies.

Cell Cycle Modeling, Process Algebra, Table 2 Summary on tools and case studies of process algebras used for modeling biological systems

	Simulation	Case studies
Biochemical pi-calculus	BioSPI, SPiM	Autoreactive lymphocyte recruitment, gene regulation, cell cycle, Rho GTP-binding in phagocytosis, FGF pathway
Bio-PEPA	PEPA workbench, eclipse plug-in	ERK signaling pathway, cell cycle, epidemiological models, genetic networks, NF-kB
BlenX	CoSBiLab (BetaWorkbench)	Cell cycle, circadian clock, actin polymerization, NF-kB, MAPK, EGFR
k-calculus	Kappa factory	Cellular signals

Process Algebras, Applications in Biology (Cell Cycle)

As the best of our knowledge, the only three process algebras used to study different aspects of the cell cycle are ▶ BioSpi, ▶ Bio-PEPA, and ▶ BlenX.

Here we just sketch the main characteristic of the chosen model with respect to the peculiar feature of those languages just to give to the reader a flavor of how different process algebras handle various aspect of the cell cycle, usually studied in the deterministic framework. We refer the reader to the original publications for more detailed discussions.

A Cell Cycle Model in BioSpi

Lecca and Priami (2003) chose to implement the control mechanism of the cell cycle (modeled by Novak in 1999) that is accounting for the antagonistic interaction between cyclin-dependent kinase dimers and the anaphase-promoting complex: The APC extinguishes cdk activity by destroying its cyclin partners, whereas cyclin/cdk dimers inhibit APC activity by phosphorylating Cdh1 (the interaction is also mediated by a cyclin-dependent kinase inhibitor).

The system is composed by six concurrent processes, corresponding to the main five species of proteins, which regulate the cell cycle: CYCLIN, CDK, CDH1, CKI, CDC14 plus an auxiliary process CLOCK. First, cyclin subunits bind to CDK monomers and make them active; then the dimers CYCLIN/CDK, the activator CDC14, and the CDH1 are involved in a negative feedback loop: CYCLIN/CDK turns on CDC14, which activates CDH1, which inhibits the CYCLIN/CDK activity, destroying the cyclin subunits. The model includes also another possibility of inhibition of CYCLIN/CDK: the stoichiometric binding with CKI.

So the main events that the code simulates are: the dimers CYCLIN/CDK formation (by the usage of private channels between the two processes), phosphorylation, and dephosphorylation of CDH1 by CDC14 (by changing of the current state of the process) and the protein degradation (by changing the current state of the process to the constant empty process). All the events occur on global channels at different suitable rates (with the exception of the dimer formation that occurs on a private channel between the two processes). The different reactions in which the components of the system are involved are implemented as a multiple non-deterministic choice, that is then turned into a probabilistic one by the BioSpi simulator (See ▶ stochastic simulation algorithm for more details about the logic behind the simulator engine).

A Cell Cycle Model in Bio-PEPA

Ciocchetta and Hillston (2009) chose the model presented by Goldbeter in 1991 and was later extended by Gardner et al. in 1998. Broadly speaking, the model describes the negative feedback loop obtained by the fact that the cyclins promote the activation of

a cdk (cdc2) which in turn activates a cyclin protease which in turn promotes cyclin degradation. In order to represent a control mechanism for the cell division cycle, Gardner and colleagues introduced a protein that initiates and concludes the cell division phase and modulates the frequency of oscillations.

The Bio-PEPA model of the system has been built following this simple sequence of steps:

- Definition of the initial and the maximum concentrations (information derived from the paper) and choice of the appropriate level/step size that define the level of granularity at which the user wants to study the system.
- Definition of functional rates and parameters of the model. Concerning the kinetic laws, some of the reactions in the input model follow mass-action kinetics, whereas all the others have Michaelis–Menten kinetics.
- Definition of species components (i.e., the reagents/products of the different reactions) and of the model component (i.e., the composition of the species components).

The Bio-PEPA model has then been analyzed, and some interesting properties about different parameter sets and their effect on the underlying Continuous Time Markov Chain have been investigated.

A Cell Cycle Model in BlenX

Palmisano et al. (2009) illustrates an easy translation to BlenX of one of the most popular deterministic model of the budding yeast cell cycle developed by Novak and Tyson in 2003. This model contains species involved in reactions at very different levels of abstraction: There are elementary reactions following mass-action kinetics and more complex interaction mechanisms abstracted as cooperative reactions with Michaelis–Menten and Hill kinetics. Moreover, continuous variables (e.g., the mass of the cell) are introduced and they influence the rate of many other reactions.

In BlenX, it is easy to code all the features listed above considering the following ideas:

- Species can be modeled as simple boxes, with interaction capabilities defined only between the species involved in pure mass-action reactions.
- Complex mathematical formula used as kinetic laws can be simply copied in the input file of the model and assigned as rate of BlenX events. The simulator takes care of calculating their appropriate values at each simulation step and put them in the classical race condition with the elementary reaction rate, following the ▶ stochastic simulation algorithm strategy.
- Continuous variables can be defined in the input file and their value is updated by the simulator so that the user can define functional rates that refer to them and that are updated by the simulator whenever the continuous variable is recalculated.

The BlenX model has been used for performing stochastic simulations of wild type and mutants cells, showing how the effect of the noise is important to characterize partially viable mutants at the border of life and death.

Cross-References

- ▶ Bio-PEPA
- ▶ BioSPI
- ▶ BlenX
- ▶ κ-Calculus
- ▶ Stochastic Pi-calculus
- ▶ Stochastic Simulation Algorithm

References

Calder M, Hillston J (2009) Process algebra modeling styles for biomolecular processes. Lect Notes Comput Sci 5750:1–25

Ciocchetta F, Hillston J (2009) Bio-PEPA: a Framework for the Modeling and Analysis of Biological Systems. Theor Comput Sci 410(33–34):3065–3084

Guerriero ML, Prandi D, Priami C, Quaglia P (2009) Process calculi abstractions for biology, process calculi abstractions for biology in algorithmic bioprocesses. Springer, Heidelberg

Hoare CAR (1985) Communicating sequential processes, Prentice hall international series in computer science. Prentice Hall, UK

Lecca P, Priami C (2003) Cell cycle control in eukaryotes: a BioSpi model. Proceedings of BioConcur 03. Electron Notes Theor Comput Sci 180(3):51–63

Milner R (1999) Communicating and mobile systems: the pi-calculus. Cambridge University Press, Cambridge

Palmisano A, Mura I, Priami C (2009) From odes to language-based, executable models of biological systems. Proceedings of the pacific symposium on biocomputing (PSB 2009)

Plotkin GD (1981) A structural approach to operational semantics. Tech Rep DAIMI FN-19, Computer Science Department, Aarhus University, Aarhus, Denmark

Priami C (2009) Algorithmic systems biology. An opportunity for computer science. Commun ACM 52:80–88

Regev A, Shapiro E (2001) Cellular abstractions: cells as computation. Nature 419:343

Cell Cycle Modeling, Stochastic Methods

Ivan Mura
The Microsoft Research - University of Trento Centre for Computational and Systems Biology, Trento, Italy

Synonyms

Probabilistic methods

Definition

Stochastic methods of modeling include randomness as a way to represent the occurrence of events that, because of their very nature or due to practical impossibilities, can only be predicted in probabilistic terms. Thus, the diffusion of molecules through a membrane, the dynamic instability of microtubules, and the spontaneous switch from the non-lytic to the lytic behavior of the λ-phage can all be modeled as stochastic processes.

Various modeling studies of cell cycle regulation networks have been proposed in the literature, which include stochastic parts of behavior. The abstraction level at which stochasticity is included in the model can provide a convenient classification criterion. Going down to the highest to the lowest level we distinguish the following four classes:

- *Stochasticity at cell division time*. The elements of randomness considered are those related to the process of cell division, which is the asymmetric division of cells with unequal partitioning of cytoplasmic and/or nuclear material between the mother and daughter cells. The weights determining the relative partitioning of the cellular content are characterized by discrete or continuous random variables.
- *Stochasticity at cell cycle phases duration*. The duration of the various cell cycle phases depends on a number of factors, including genotypic differences, intrinsic molecular noise, and external factors such as the abundance of nutrients and DNA damage inducing events. Therefore, cell cycle phase duration is not constant in a population of cells, neither is across subsequent rounds of duplication of the same cell. The duration of cell cycle phases can be modeled as continuous time random variables.
- *Stochasticity as molecular noise*. The molecular dynamics of the proteins that participate to the cell cycle regulation network can be modeled by adding to a pure deterministic process that represents the average change, a stochastic process that represents molecular noise. In line with this approach we find probabilistic versions of Boolean networks, where the deterministic causality dictated by the structure of the model is relaxed through the introduction of probabilities of state change. Another way to represents stochastic changes in the molecular concentrations is through a differential ▶ Langevin equation with multiplicative noise, as shown in Eq. 1:

$$\frac{d}{dt}x(t) = f(x(t)) + N(0,1)\sqrt{2 \cdot D \cdot x(t)} \qquad (1)$$

where $x(t)$ is the concentration of species x at time t, $f(\cdot)$ is a deterministic function, $N(0,1)$ is a standard Gaussian noise (null expectation and unitary variance), and D is the noise amplitude.
- *Stochastic kinetics*. The randomness is directly introduced at the level of molecular interactions, according to the formulation of stochastic kinetics used in ▶ Gillespie stochastic simulation. Under a set of hypotheses about the homogeneity and the thermal equilibrium of the system, the times to the occurrence of reactions in the system can be faithfully modeled by negative exponential ▶ random variables. This allows describing the evolution of the system over time through a simple system of first order linear differential equations, known as the Chemical ▶ Master equation, or equivalently, as a discrete space continuous-time Markov process (▶ Markov Chain). The cell cycle regulation network is thus modeled by assigning to each reaction, for each possible state of the system, a rate of occurrence that is interpreted as the rate of a negative exponential distribution. Various high-level languages for specifying stochastic kinetics models exist, we cite here the stochastic Petri net (▶ Cell Cycle Modeling, Petri Nets) and stochastic process algebra (▶ Cell Cycle Modeling, Process Algebra) formalisms as they have been recently applied to the modeling of cell cycle regulation networks.

Characteristics

The four stochastic modeling approaches presented above have been applied, sometimes combined with each other and mixed with deterministic parts of modeling, to define *in-silico* models of cell cycle regulation networks. The vast majority of modeling studies that are of interest here deal with the cell cycle of yeast. The corpus of deterministic models of yeast cell cycle, based on Ordinary Differential Equations (▶ Ordinary Differential Equation (ODE)), has been extended in various ways with the stochastic methods to represent extrinsic and intrinsic molecular variability. A few stochastic models of higher organisms cell cycle regulation are also found in the literature.

Fission Yeast

In a pursuit to determine the origin of the observed variability in cell cycle duration, Sveiczer et al. (2001) modified a pure deterministic model of fission yeast (▶ Cell Cycle, Fission Yeast) cell cycle by introducing into it the asymmetric results of ▶ cytokinesis. The cell size at birth is modeled as the product of a Gaussian random variable – with average value = 0.5 and standard deviation = 0.016 – and the size of the mother cell, whereas the initial nuclear volume is assumed to be Gaussian, with average value = 1.0 (in arbitrary units) and standard deviation = 0.07 (determined from the fit with experimental data, due to the unknown distribution of real nucleus size). The inclusion of this stochasticity allowed the model to reproduce the experimentally observed levels of variability for both wild type and *wee1*$^-$ mutant fission yeast cell population. An extension of a deterministic model of fission yeast cell cycle molecular machinery through Langevin equation noise terms has been proposed in (Steuer 2004). The author rewrites a model composed by ordinary differential equations by adding stochastic noise to each of them. The simulation results demonstrate that the introduction of noise makes the model able to account for emerging properties, such as existence of quantized cycle times, thus showing that fluctuations cannot always be treated as small perturbations to the deterministic behavior. The effects of both intrinsic (stochastic molecular kinetics) and extrinsic (asymmetrical split at division time) are studied by Kar et al. (2009) with a Gillespie simulation scheme. The model shows that both intrinsic and extrinsic noises contribute significantly to the variability of cell cycle progression. Nonetheless, intrinsic molecular fluctuations in the control system are considerably noisier, mostly due to the low numbers of mRNA molecules reported for yeast cells.

Budding Yeast

The cell cycle of budding yeast (▶ Cell Cycle, Budding Yeast) has also been the subject of modeling studies employing stochastic methods. A modeling work based on a ▶ Boolean network by Zhang et al. (2006) encodes the activation/repression relationships among the key molecular species, and a perturbing probabilistic noise is considered when evaluating the state update rules. The authors show that the model is quite robust and that the main cycling behavior is conserved with high probability even for very noisy models. Similar conclusions were also drawn by Sabouri-Ghomi et al. (2008), in a study that explored the effects of molecular intrinsic noise in cell cycle progression of budding yeast by a stochastic kinetic model a la Gillespie. They demonstrated that the molecular switches controlling the Start (entry into S phase) and the Finish (exit from M phase) transitions are robust even for low molecular counts of the key regulating molecules. The works in (Mura and Csikász-Nagy 2008) and in (Palmisano et al. 2009) introduce, through the application on the budding yeast cell cycle regulation networks, a procedure to translate deterministic models based on ordinary differential equations into stochastic kinetic models for Gillespie simulation specified as stochastic Petri net and stochastic process algebra models, respectively. The authors demonstrate how stochastic models can be used to obtain statistics of key cell cycle parameters, such as duration and cellular mass, beyond the average one, information useful to revealing characteristics of the phenotypes that could not be anticipated by deterministic models. Finally, a stochastic kinetics model of the core of budding yeast cell cycle regulation network was studied through a combination of stochastic simulation and model checking techniques in (Ballarini et al. 2009) to characterize the elements of the network determining the irreversibility (▶ Cell Cycle Dynamics, irreversibility) of phase transitions.

Higher Organisms

A hybrid Petri net model of the cell division process of Xenopus frog embryo (▶ Cell Cycle of Early Frog Embryos) has been proposed in (Matsui et al. 2004). This model output is able to reproduce the changes in the cell division cycles from synchronous to asynchronous. The cell cycle desynchronization in a culture of proliferating mammalian cells (▶ Cell Cycle of Mammalian Cells) is the subject of Olofsson and McDonald (2010), where the stochastic characterization of phase duration is taken as an input to a branching process model of replicating cells, able to match the experimentally observed percentages of cells in the various phases of the cycle. We finally mention the modeling study of Zamborsky et al. (2007), which deals with the cross-talking upon the Wee1 kinase between circadian rhythm and cell cycle regulation networks in mammalian cells. A stochastic model of cell cycle regulation network expressed through Langevin equation is coupled with circadian clock (▶ Cell Cycle, Coupled with Circadian Clock) deterministic model. The model suggests that the circadian clock may contribute to enforce a cell size control on the cell cycle when the mass doubling time is quite different from the circadian period.

Cross-References

- ▶ Boolean Networks
- ▶ Cell Cycle Dynamics, Irreversibility
- ▶ Cell Cycle Modeling, Petri Nets
- ▶ Cell Cycle Modeling, Process Algebra
- ▶ Cell Cycle Modeling, Stochastic Methods
- ▶ Cell Cycle of Early Frog Embryos
- ▶ Cell Cycle of Mammalian Cells
- ▶ Cell Cycle, Budding Yeast
- ▶ Cell Cycle, Coupled with Circadian Clock
- ▶ Cell Cycle, Fission Yeast
- ▶ Cytokinesis
- ▶ Gillespie Stochastic Simulation
- ▶ Langevin Equation
- ▶ Master Equation
- ▶ Ordinary Differential Equation (ODE)
- ▶ Random Variable
- ▶ Stochastic Processes, Fokker-Planck Equation

References

Ballarini P, Mazza T, Palmisano A, Csikász-Nagy A (2009) Studying irreversible transitions in a model of cell cycle regulation. Electron Notes Theor Comput Sci 232:39–53. doi:10.1016/j.entcs.2009.02.049

Kar S, Baumann WT, Paul MR, Tyson JJ (2009) Exploring the roles of noise in the eukaryotic cell cycle. Proc Natl Acad Sci USA 106:6471–6476. doi:10.1073/pnas.0810034106

Matsui M, Fujita S, Suzuki S, Matsuno H, Miyano S (2004) Simulated cell division processes of the Xenopus cell cycle pathway by Genomic Object Net. J Integr Bioinform 1:3. doi:10.2390/biecoll-jib-2004-3

Mura I, Csikász-Nagy A (2008) Stochastic Petri Net extension of a yeast cell cycle model. J Theor Biol 254:859–860. doi:10.1016/j.jtbi.2008.07.019

Olofsson P, McDonald TO (2010) A stochastic model of cell cycle desynchronization. Math Biosci 223:97–104. doi:10.1016/j.mbs.2009.11.003

Palmisano A, Mura I, Priami C (2009) From ODEs to language-based, executable models of biological systems. In: Proceedings of the pacific symposium on biocomputing 14. Kohala Coast, Hawaii, USA, pp 239–250.

Sabouri-Ghomi M, Ciliberto A, Kar S, Novak B, Tyson JJ (2008) Antagonism and bistability in protein interaction networks. J Theor Biol 250:209–218. doi:10.1016/j.jtbi.2007.09.001

Steuer R (2004) Effects of stochasticity in models of the cell cycle: from quantized cycle times to noise-induced oscillations. J Theor Biol 228:293–301. doi:10.1016/j.jtbi.2004.01.012

Sveiczer A, Tyson JJ, Novak B (2001) A stochastic, molecular model of the fission yeast cell cycle: role of the nucleocytoplasmic ratio in cycle time regulation. Biophys Chem 92:1–15. doi:10.1016/S0301-4622(01)00183-1

Zámborszky J, Hong CI, Csikász-Nagy A (2007) Computational analysis of mammalian cell division gated by a circadian clock: quantized cell cycles and cell size control. J Biol Rhythms 22:542–553. doi:10.1177/0748730407307225

Zhang Y, Qian M, Ouyang Q, Deng M, Li F, Tanga C (2006) Stochastic model of yeast cell-cycle network. Phys D 219:35–39. doi:10.1016/j.physd.2006.05.009

Cell Cycle Models, Sensitivity Analysis

Tamás Turányi
Institute of Chemistry, Eötvös University (ELTE), Budapest, Hungary

Synonyms

Sensitivity analysis; Uncertainty analysis

Definition

Sensitivity analysis is the name of a family of mathematical methods that investigate the relation between the values of the parameters and the output of models. Local sensitivity analysis explores the effect of small parameter changes, while the methods of global sensitivity analysis may explore large parameter domains.

Characteristics

Local Sensitivity Analysis

A dynamical model can be characterized by the following initial value problem

$$\mathrm{d}\mathbf{Y}/\mathrm{d}\,t = \mathbf{f}(\mathbf{Y},\mathbf{p}) \quad \mathbf{Y}(0) = \mathbf{Y}^0 \tag{1}$$

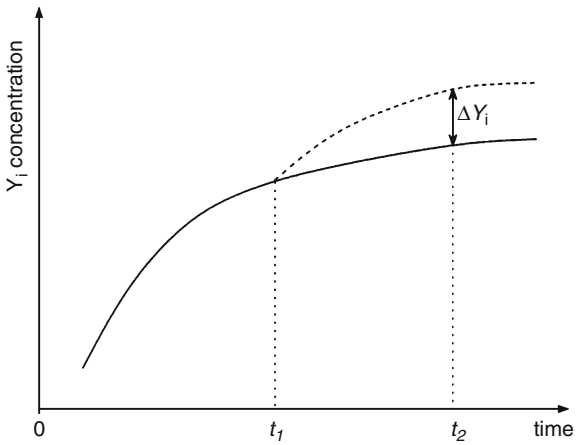
where t is time, $\mathbf{Y}$ is the n-vector of variables, $\mathbf{p}$ is the m-vector of parameters, $\mathbf{Y}^0$ is the vector of the initial values of the variables, and $\mathbf{f}$ is the right-hand-side of the differential equations.

Local sensitivity analysis means the calculation of partial derivative $\partial Y_i/\partial p_j$, which is called the first order local sensitivity coefficient. The meaning of a local sensitivity coefficient is visualized in Fig. 1. The concentration–time curve, as obtained by the solution of the original model, is plotted with solid line. If parameter j is changed by Δp_j at time t_1, then the solution of the modified model will deviate (see the dashed line) from the original solution. Denote ΔY_i the change in the calculated solution at time t_2. The finite difference approximation can be used for the calculation of the local sensitivity coefficient:

$$\frac{\partial Y_i}{\partial p_j} \approx \frac{\Delta Y_i}{\Delta p_j} \tag{2}$$

Local sensitivity coefficient $\partial Y_i/\partial p_j$ depends on the time of parameter perturbation t_1 and the time of observation t_2. If the time of perturbation t_1 is fixed, then the local sensitivity coefficient is a function of time $t = t_2 - t_1$.

The error of the finite difference approximation cannot be controlled effectively, therefore the local sensitivity function $s_{ik}(t)$ is usually calculated by solving the following initial value problem (Tomović and Vukobratović 1972):



Cell Cycle Models, Sensitivity Analysis, Fig. 1 Concentration – time solution of a model (*solid line*) and the modified solution when one of the parameters is changed at time t_1 (*dashed line*)

$$\dot{\mathbf{S}} = \mathbf{JS} + \mathbf{F} \quad \mathbf{S}(t_1) = \mathbf{0} \tag{3}$$

where $\mathbf{S}(t) = \{s_{ij}(t)\} = \{\partial Y_i/\partial p_j\}$ is the time-dependent local sensitivity matrix, matrix $\mathbf{J}$ is the Jacobian ($\mathbf{J} = \{\partial f_i/\partial Y_k\}$), and matrix $\mathbf{F}$ contains the derivatives of the right-hand-side of the ODE with respect to the parameters ($\mathbf{F} = \{\partial f_i/\partial p_j\}$). Matrices $\mathbf{J}$ and $\mathbf{F}$ are functions of time t_2. The sensitivity coefficients are usually used in normalized $(p_j/Y_i)\,(\partial Y_i/\partial p_j)$ or semi-normalized $p_j\,(\partial Y_i/\partial p_j)$ form. The normalized sensitivity coefficients are dimensionless and show the percentage change of model solution i as a result of +1% change of the value of parameter j.

Global Sensitivity Analysis

Textbooks about the various methods of global sensitivity analysis and their applications in science have been published by Saltelli et al. (2004, 2008). The most frequently applied methods of global sensitivity analysis are the Monte Carlo method with Latin hypercube sampling, application of pseudo random numbers (e.g., a Sobol' sequence), Fourier Amplitude Sensitivity Test (FAST), and High Dimensional Model Representation (HDMR). These methods are able to determine the

uncertainty of each model output (like the concentration at a given time or the period time) knowing the uncertainty of each parameter. The ultimate information (usually obtained by the combined application of several methods) is the calculation of the joint probability density function (*pdf*) of the model outputs and the contribution of each parameter to it knowing the joint *pdf* of the parameters.

Raw and Cleaned Local Sensitivity Functions

The concentration–time curves of a cell cycle model are always nearly periodic and in most models the concentration–time curves are truly periodic having elapsed a transition time. In this case one period time τ later the concentrations are identical:

$$\mathbf{Y}(t) = \mathbf{Y}(t+\tau) \tag{4}$$

It is possible to investigate how period time τ depends on the values of the parameters. Using a local sensitivity approach, this information is provided by the period time sensitivities $\partial\tau/\partial p_j$.

One might assume that the sensitivity functions of a model having periodic solution are also periodic. However, it is true only for the local sensitivity functions belonging to parameter p_k, if the corresponding period time sensitivity is zero ($\partial\tau/\partial p_k = 0$). If $\partial\tau/\partial p_k$ is not zero, then the maxima of the sensitivity–time curves are continuously increasing. It has been shown by several authors (see (Tomović and Vukobratović 1972), (Edelson and Thomas 1981), (Larter 1983), (Zak et al. 2005)) that the original ("raw") sensitivity functions $\left(\partial Y_i/\partial p_j\right)$ can be separated to truly periodic ("cleaned") sensitivity functions $\left(\partial Y_i/\partial p_j\right)_\tau$ and a secular term:

$$\frac{\partial Y_i}{\partial p_j} = \left(\frac{\partial Y_i}{\partial p_j}\right)_\tau - \frac{t}{\tau}\frac{\partial \tau}{\partial p_j}\frac{\partial Y_i}{\partial t} \tag{5}$$

These authors suggested several methods for the accurate determination of the period time sensitivity $\partial\tau/\partial p_j$ and the cleaned sensitivity functions from the raw sensitivity functions (Fig. 2).

The Similarity of the Local Sensitivity Functions

In the case of a general mathematical model, no relation is expected among the rows and/or the columns of the sensitivity matrix. However, in several chemical kinetic systems the following relations have been observed (see Rabitz 1989; Zsély et al. 2003):

1. Local similarity: Value

$$\lambda_{ij}(t) = \frac{s_{ik}(t)}{s_{jk}(t)} \tag{6}$$

depends on time t and the model results Y_i and Y_j selected, but is independent of the parameter p_k perturbed.

2. Scaling relation: Equation

$$\frac{(dY_i/dt)}{(dY_j/dt)} = \frac{s_{ik}(t)}{s_{jk}(t)} \tag{7}$$

is valid for any parameter p_k. Existence of scaling relation includes the presence of local similarity.

3. Global similarity: Value

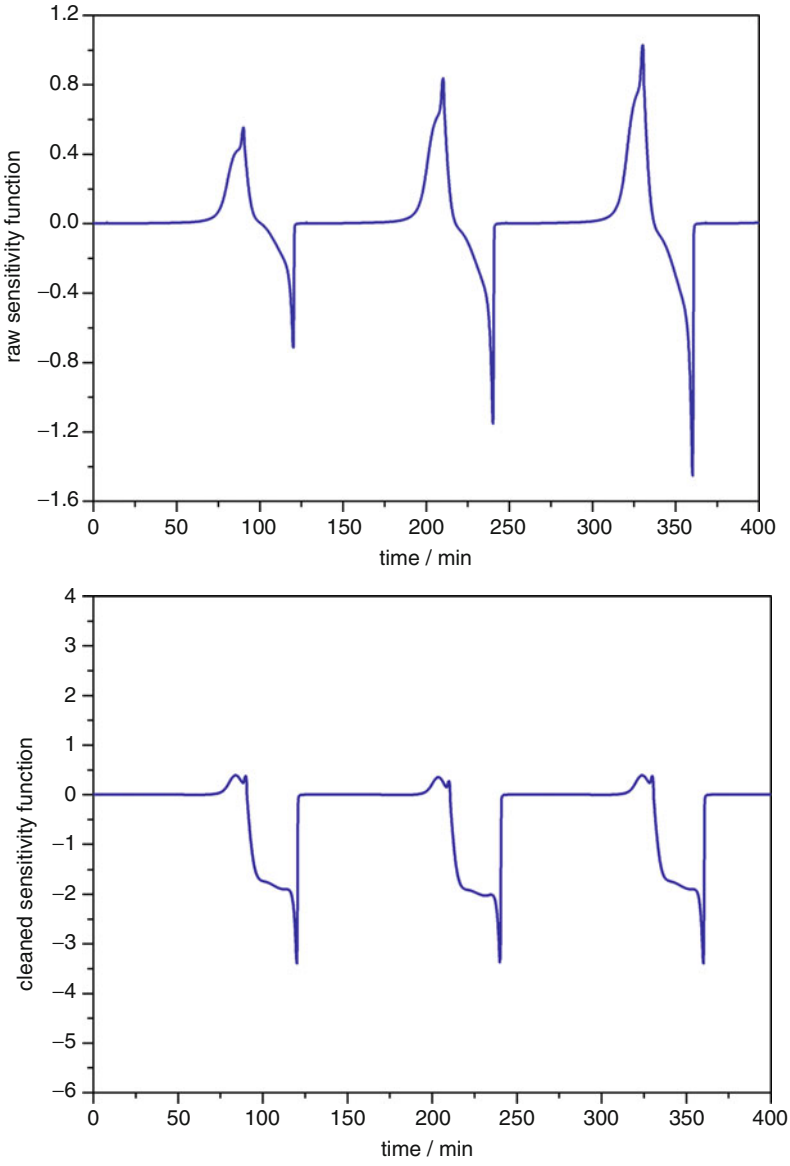
$$\mu_{ikm} = \frac{s_{ik}(t)}{s_{im}(t)} \tag{8}$$

is independent of t (within an interval).

Zsély et al. (2003) have shown that the existence of low-dimensional manifolds in variable space may cause local similarity. Global similarity emerges if local similarity is present and the sensitivity differential equations are pseudo-homogeneous. The latter is related to the existence of excitation periods, like the autocatalytic runaways of concentrations.

In cell cycle models low-dimensional manifolds and excitation periods may occur (Lovrics et al. 2006), and accordingly all the three types of relations between the sensitivity functions were found (Lovrics et al. 2008). If the maxima of the sensitivity functions are continuously increasing, then the cleaned sensitivity functions $\left(\partial Y_i/\partial p_j\right)_\tau$ should be investigated.

Detection of the global similarity of sensitivity functions has a special importance. If sensitivity functions $\partial Y_i/\partial p_j$ and $\partial Y_i/\partial p_k$ are globally similar, then changing parameter j causes exactly the same effect on concentration–time curve $Y_i(t)$ during the whole time period than a (different extent) change of parameter k. This means that the effect of the change of parameter j can be compensated by an appropriate modification of the value of parameter k. As a consequence, an infinite number of parameter sets



Cell Cycle Models, Sensitivity Analysis, Fig. 2 Raw (**a**) and cleaned (**b**) local sensitivity functions of a cell cycle model

may result in the same simulation results. Also, the effect of a physical change of a parameter in a biological system (e.g., due to the change of the environment) can be fully compensated by the appropriate change of another parameter, which is a form of ► robustness in biological systems.

Cross-References

► Robustness

References

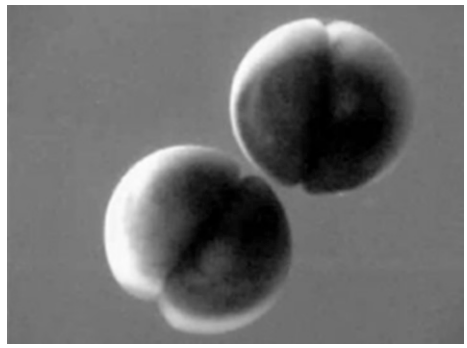
Edelson D, Thomas VM (1981) Sensitivity analysis of oscillating reactions. J Phys Chem 85:1555–1558

Larter R (1983) Sensitivity analysis of autonomous oscillators. Separation of secular terms and determination of structural stability. J Phys Chem 87:3114–3121

Lovrics A, Csikász-Nagy A, Zsély IGy, Zádor J, Turányi T, Novák B (2006) Time scale and dimension analysis of a budding yeast cell cycle model. BMC Bioinformatics 7:494

Lovrics A, Zsély IGy, Csikász-Nagy A, Zádor J, Turányi T, Novák B (2008) Analysis of a budding yeast cell cycle model using the shapes of local sensitivity function. Int J Chem Kinet 40:710–720

Rabitz H (1989) Systems analysis at the molecular scale. Science 246:221–226
Saltelli A, Tarantola S, Campolongo F, Ratto M (2004) Sensitivity analysis in practice. A guide to assessing scientific models. Wiley, Chichester
Saltelli A, Ratto M, Andres T, Campolongo F, Cariboni J, Gatelli D, Saisana M, Tarantola S (2008) Global sensitivity analysis: the primer. Wiley, Chichester
Tomović R, Vukobratović M (1972) General sensitivity theory. Elsevier, New York
Zak DE, Stelling J, Doyle FJ III (2005) Sensitivity analysis of oscillatory (bio)chemical systems. Comput Chem Eng 29:663–673
Zsély IGy, Zádor J, Turányi T (2003) Similarity of sensitivity functions of reaction kinetic models. J Phys Chem A 107:2216–2238



Cell Cycle of Early Frog Embryos, Fig. 1 Early *Xenopus laevis* embryos after their first cell cleavage

Cell Cycle of Early Frog Embryos

Joseph R. Pomerening
Department of Biology, Indiana University, Bloomington, IN, USA

Synonyms

Cyclin-dependent kinase 1 (CDK1)/Anaphase-promoting complex (APC) oscillator; Early embryonic oscillator; Xenopus laevis embryonic cell cycle

Definition

Early frog embryos – including those of the well-studied genus *Xenopus* – exhibit a cell cycle with unique characteristics. These include a lack of growth phases, as well as an inability to arrest despite chemical or physical insult, including DNA-damaging or cytoskeleton-disrupting agents, as well as nuclear ablation. Alternation of genome replication (S-phase) and daughter chromosome segregation followed by cell cleavage (M-phase) occurs quickly in frog embryos (Fig. 1).

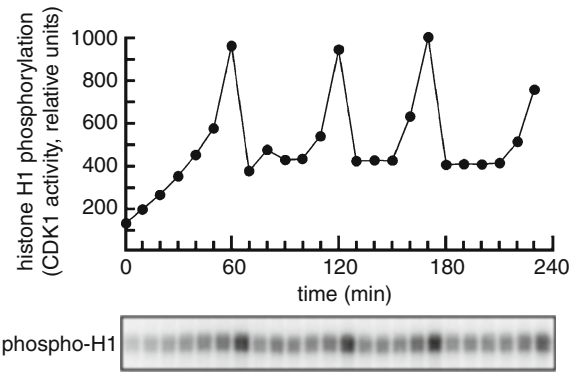
Egg-borne maternal stores of mRNA and protein drive these oscillations, with the duration of cycles 2–12 being roughly 30 min, following an initial 75–90 min cycle. Due to the rapid and independent nature of these frog early embryonic cycles, the enzymatic tandem that drives it – the stimulatory phospho-transferase, cyclin-dependent kinase 1 (CDK1), and the inhibitory ubiquitin ligase, anaphase-promoting complex (APC) – is known to behave as an autonomous, clock-like cell cycle oscillator. Following the first 12 cleavages, at a stage known as the mid-blastula transition (MBT), the frog embryo cell cycle is modified to include zygotic transcription, and slows due to the incorporation of growth phases, lengthened S-phases, and the ability to induce cell cycle arrest due to the activation of checkpoints.

Characteristics

No Gap or Growth Phases, and Zygotic Transcription is Inhibited

Frog eggs and early embryos have been a long-utilized experimental system due to the ease of care of parental populations, their large size lending to the ease of manipulation and observation, and the ability to easily collect an abundance of eggs suitable for use in biochemical studies – primarily in the pseudo-tetraploid *Xenopus laevis* – or for genetic manipulation in the diploid *Xenopus tropicalis*. Cell cycles in the early embryos of these frogs – from the period of fertilization up until MBT, when the zygotic genetic program is initiated – exhibit unique characteristics that distinguish it from the cycles that occur in somatic cells later in development, as well as in their mammalian embryonic counterparts (Murray and Hunt 1993). The frog early embryonic cell cycle, however, does share similarities to the early cycles of other amphibians, fish, and insects. One similarity is that the earliest cleavages that occur following fertilization do not accommodate periods for growth, while rapid exchanges between the periods of DNA replication and mitosis occur.

Xenopus laevis eggs – and developing one-cell blastomeres – are roughly 1 mm in diameter, and remain this size until their cell cycle is modified to include growth phases after the onset of MBT (Bernardini et al. 1999). The maternal contribution of protein and mRNAs to the frog egg is critical to these early cell cleavages, because its genome remains transcriptionally inactive until the embryo reaches MBT. Experimental evidence points to this transition correlating strongly with the nuclear-to-cytoplasmic volume (N-to-C) ratio in cells, but what triggers the cell cycle to behave more deliberately with its inclusion of growth phases and checkpoints remains unknown (Newport and Kirschner 1982).



Cell Cycle of Early Frog Embryos, Fig. 2 Oscillations of CDK1 activity in a *Xenopus laevis* cycling egg extract

No Checkpoints in Response to Unreplicated DNA, DNA Damage, or Mitotic Defects

After fertilization, frog early embryos continue their rapid cell cycles unimpeded, even if exposed to agents that block the processes required for DNA replication and cell division (Morgan 2006). Treatment with DNA synthesis inhibitors prevents S-phase completion, but cell cleavages will continue. In fact, the presence of a nucleus is not a requirement: enucleation of fertilized eggs has been shown inconsequential to oscillations of the biochemical clock that drives the frog early embryonic cell cycle. Poisons including nocodazole and taxol inhibit the assembly and disassembly of the mitotic spindle, respectively, but even these cell division inhibitors fail to block the activities of the oscillating enzymes that drive the cycles. Frog early embryos treated with these drugs do not undergo cleavages, but do exhibit waves of cortical/cell surface contractions that mirror the timing when mitosis would be expected to occur. DNA-damaging agents – including UV irradiation – are also insufficient to halt progression through the cell cycle in these early embryos. Altogether, these traits exemplify the fact that the cell cycle and its control system in frog early embryos is a clock-like, autonomous oscillator that is not susceptible to perturbations that affect the genome, or the subcellular structures that are necessary for cell division.

Frog Early Embryonic Cell Cycles Are Driven by Alternating Oscillations of CDK1 and APC Activities

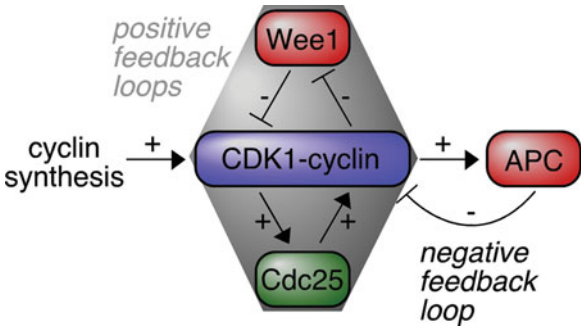
During the late 1980s, extensive physiological and biochemical studies led to the purification of the M-phase-promoting factor from unfertilized *Xenopus laevis* eggs (Lohka et al. 1988). This activity was ascribed to a heterodimeric complex of proteins comprising the cyclin-dependent kinase 1 (CDK1), and its activating counterpart, cyclin B. In 1995, the enzyme responsible for targeting cyclin B for proteolysis by the proteasome – called the anaphase-promoting complex (APC) – was also purified from Xenopus eggs and clams (King et al. 1995; Sudakin et al. 1995). Together, the M-phase stimulatory CDK1 protein kinase and the CDK1-inhibiting APC ubiquitin ligase were found to be the system that drives the clock-like oscillations during the early embryogenesis of frogs, as well as other amphibians, marine invertebrates (such as clams and starfish), and insects. Cycling extracts derived from parthenogenetically activated Xenopus eggs exhibit rapid and sustained oscillations of CDK1 activity, and provide a useful tool to study the dynamical behaviors of this enzyme and others within the frog early embryonic oscillatory system (Murray 1991). CDK1 activity is assayed in these cytoplasmic extracts by measuring its phosphorylation of a preferred substrate, histone H1 (Fig. 2).

Positive Feedback Generates Bistability in the Response of CDK1 to Cyclin, and Confers a Relaxation Oscillator Design in the Frog Early Embryo

The activation of CDK1 is under the control of inhibitory kinases and activating phosphatases (Morgan 2006). Upon binding of cyclin, phosphorylation of CDK1 on residues of its ATP-binding site (Thr14/Tyr15) by the kinase Wee1 (as well as another, called Myt1) inhibits its activity. Activation of the phosphatase Cdc25 counteracts this inhibition, but also initiates two

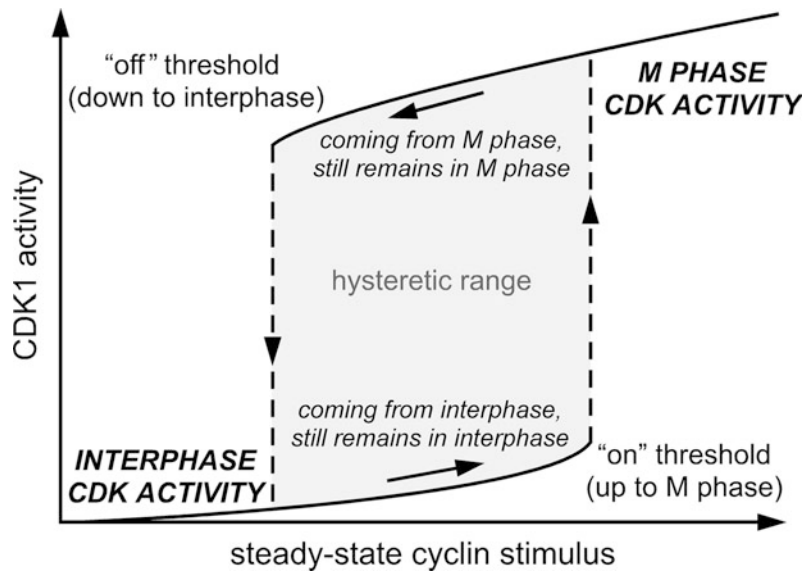
critical positive-feedback loops. CDK1 phosphorylates and inhibits the Wee1 and Myt1 kinases – causing a double-negative (positive)-feedback loop – while also further activating additional Cdc25, and inducing a second positive-feedback loop (Fig. 3).

This activation, in turn, stimulates the APC-mediated negative-feedback loop, and induces mitotic exit and interphase by resetting CDK1 to an inactive state through targeting the cyclin subunit for proteolysis. Together, these positive- and negative-feedback subcircuits provide the basis for CDK1 activation and inactivation, respectively, during the cell cycle. Systems-level studies into the dynamical nature of CDK1 activation – that included a fusion of both computation and experiment – revealed that the steady-state threshold for CDK1 activation is higher (when leaving the off-state), than the threshold level of cyclin that must be surpassed to switch CDK1 from on to off (Pomerening et al. 2003; Sha et al. 2003). This steady-state behavior is known as hysteresis, and this means that there is bistability in response of CDK1 to cyclin stimulus: depending on where the system starts – either in the off or on state – it can exist in either of those two states for a range of cyclin concentrations (Fig. 4).



Cell Cycle of Early Frog Embryos, Fig. 3 Positive-feedback loops in CDK1 activation

In other words, within the hysteretic range of cyclin stimulus, if CDK1 starts in the on state (mitosis) and the stimulus decreases, CDK1 will remain active (though its activity would be decreased, since there would be less cyclin present) and the cell would remain in mitosis. If the system starts in the off state and the stimulus increases to the same level as mentioned previously, CDK1 would remain off and the cell would not leave interphase. Combining positive feedback and bistability with the APC-driven negative-feedback loop, the CDK1-APC oscillator in frog early embryos behaves as a relaxation oscillator – producing rapid and spike-like bursts of CDK1 activity – followed by its rapid APC-mediated inactivation. Lessening of this positive feedback through the addition of a non-phosphorylatable (Wee1-insensitive) mutant of CDK1 was found to cause damped CDK1 activity oscillations (Pomerening et al. 2005). This caused defects in cell cycle progression – namely, the premature inhibition of DNA synthesis – by rendering egg extracts incapable of sustaining periods of low



Cell Cycle of Early Frog Embryos, Fig. 4 Hysteresis and bistability in CDK1 activation

CDK1 activity. Therefore, on the systems-level, positive feedback and bistability in this oscillator help to ensure sustained, pulsatile oscillations of CDK1 activity to properly drive cell cycle progression during frog early embryogenesis.

References

Bernardini G, Prati M, Bonetti E, Scari G (1999) Atlas of Xenopus development. Springer, New York

King RW, Peters JM, Tugendreich S, Rolfe M, Hieter P, Kirschner MW (1995) A 20S complex containing CDC27 and CDC16 catalyzes the mitosis-specific conjugation of ubiquitin to cyclin B. Cell 81(2):279–288

Lohka M, Hayes MK, Maller JL (1988) Purification of maturation-promoting factor, an intracellular regulator of early mitotic events. Proc Natl Acad Sci USA 85:3009–3013

Morgan DO (2006) The cell cycle: principles of control. New Science Press, London

Murray AW (1991) Cell cycle extracts. Methods Cell Biol 36:581–605

Murray AW, Hunt T (1993) The cell cycle: an introduction. Oxford University Press, New York

Newport J, Kirschner M (1982) A major developmental transition in early Xenopus embryos: I. Characterization and timing of cellular changes at the midblastula stage. Cell 30(3):675–686

Pomerening JR, Sontag ED, Ferrell JE Jr (2003) Building a cell cycle oscillator: hysteresis and bistability in the activation of Cdc2. Nat Cell Biol 5(4):346–351

Pomerening JR, Kim SY, Ferrell JE Jr (2005) Systems-level dissection of the cell-cycle oscillator: bypassing positive feedback produces damped oscillations. Cell 122(4): 565–578

Sha W, Moore J, Chen K, Lassaletta AD, Yi CS, Tyson JJ, Sible JC (2003) Hysteresis drives cell-cycle transitions in Xenopus laevis egg extracts. Proc Natl Acad Sci USA 100(3):975–980

Sudakin V, Ganoth D, Dahan A, Heller H, Hershko J, Luca FC, Ruderman JV, Hershko A (1995) The cyclosome, a large complex containing cyclin-selective ubiquitin ligase activity, targets cyclins for destruction at the end of mitosis. Mol Biol Cell 6(2):185–197

Cell Cycle of Mammalian Cells

Gerard J. Ostheimer
Department of Agriculture, Washington, DC, USA

Synonyms

Cell division; Cell growth; Proliferation

Definition

The cell cycle is the regulated, ordered progression of steps by which a cell commits to proliferation, replicates its DNA, and divides into two daughter cells. The mammalian cell cycle is coupled to normal and aberrant biological processes unique to multicellular organisms including development, differentiation, tissue formation, wound healing, and cancer. Numerous signal transduction pathways link these processes to mammalian cell cycle entry and progression. Cancer is the disease state that results from excessive cellular proliferation. Cancer results when mechanisms for regulating the cell cycle are lost, as such the complexities of the mammalian cell cycle have received significant attention from the research community. Mammalian cells have signal transduction pathways that monitor aberrant proliferation and in response commit the cell to either senescence or apoptosis as a barrier to tumorigenesis. The complex, multivariate relationship between the cell cycle and the context of the cell necessitates a systems biology approach to understanding how the cell cycle is regulated and importantly for understanding how this regulation fails in cancer cells.

Characteristics

Fundamentals

Mammals are complex, multicellular organisms possessing diverse tissues and organs containing a plethora of cell types. All of these diverse cell types are generated during development and arise from the totipotent fertilized egg. The process that generates distinct cell types is known as cellular differentiation. The overwhelming majority of cells in a mature mammal are somatic cells, which are terminally differentiated and no longer proliferate. Mammals maintain pluripotent stem cells that drive tissue regeneration and wound healing.

The cell cycle of a proliferating mammalian cell consists of four phases G1, S, G2, and M. Mitosis (M) is the ordered series of events that result in the replicated chromosomes being physically separated, i.e., segregated, and cytokinesis in which two daughter cells with equal amounts of genetic material are produced. Replication occurs during the synthesis (S) phase. G1 is the gap period between M and S phase and G2 is the gap period between S and M phase. Actively proliferating

mammalian cells, such as cancer cells in tissue culture, move through the cell cycle in 12–24 h. Cells with the potential to proliferate but which are not proliferating are in the G0 phase of the cell cycle and are described as quiescent. As in yeast, progression through the cell cycle is driven and regulated by cyclin–cyclin-dependent kinase (CDK) pairs responsible for promoting and regulating each of the four transitions G0-G1, G1-S, S-G2, G2-M. Mammalian cells have a larger and more diverse set of cyclins than yeast and these cyclins allow for tissue and context-specific regulation of the cell cycle (see article by Malumbres).

Cell Cycle Entry

It is extremely difficult to monitor cell growth and proliferation in vivo. As such, much of our understanding of the mammalian cell cycle is from experiments performed in tissue culture where it is possible to control cellular exposure to many of the variables that determine whether cells proliferate including exposure to growth factors and nutrient availability. Removal of nutrients and growth factors, aka serum starvation, drives many immortal cell lines out of the cell cycle and into quiescence (G0). Quiescence is characterized by hypophosphorylated Rb that binds E2F1 thereby blocking transcription of genes essential or cell cycle entry. Addition of nutrients and growth factors activates signal transduction pathways that upregulates gene expression that results in phosphorylation of Rb and causes Rb to no longer bind E2F1. Released E2F1 upregulates the expression of Cyclin D and Cdk1, and this cyclin-CDK pair regulates the transition from G0 to G1. Once cells have passed the restriction point (link to article by You et al.) then they move into S and are committed to moving through the entire cell cycle even if serum starved again.

A complex set of signal transduction pathways communicates extracellular cues to the cell cycle machinery. A canonical example is the extracellular growth factor (EGF) pathway. EGF binds its receptor, EGFR, activating the mitogen-activated protein kinase (MAPK aka ERK) pathway, which in turn activates transcriptional programs (via c-Myc and CREB) that induce cell cycle entry. The Akt pathway also transduces extracellular growth factor signal into cell cycle entry. Many components of the MAPK and Akt pathways are oncogenes. Mutations in these proteins can drive oncogenesis, and cancer cells often exhibit oncogene addiction (link to article by Frick et al.). Nutrient availability, a prerequisite for growth and proliferation, is communicated via the mTor pathway (Hay and Sonenberg 2004). The EGFR signaling pathway has received considerable attention from systems and computational biologists and highly detailed mass-action kinetic models have been published (Chen et al. 2009). These models are being used to generate molecular targets for anticancer therapeutics (Schoeberl et al. 2009).

Cell Cycle Quality Control

For proliferation to succeed both daughter cells must receive accurate and complete copies of the genome. Replicating the genome by unwinding, transcribing, and accurately segregating chromosomes (S phase) causes numerous sites of DNA damage such as single and double strand breaks. In addition, there are numerous additional sources of DNA damage including free radicals from the cell's own metabolism and environmental sources of damaging agents such as background radiation. DNA damage is detected and responded to by an extensive molecular machinery of proteins that recognize DNA damage and signal its presence and extent to the cell and DNA repair machinery. In mammalian cells the protein p53 functions as a signal integration node coupling DNA damage to cell cycle arrest, senescence, and apoptosis. In addition, the signaling kinases of the DNA damage response, Chk1 and Chk2, directly couple the DNA damage response to the cell cycle machinery. For more details on how DNA damage is coupled to the cell cycle, see Cell cycle signaling, DNA damage by Toettcher in this volume.

Cell fate after DNA damages is a function of the amount of damage, the dynamics of the damage, and interplay with extracellular cues. Low levels of transient DNA damage induce a temporary cell cycle arrest that permits repair. In the G2 phase of the cell cycle the cell possesses two copies of each chromosome, which allows the cell to correct damage generated in S phase through coupled homologous recombination and repair. High and/or persistent levels of DNA damage signal through p53 to induce either senescence or apoptosis. Differentiated mammalian cells do not express telomerase and as such each round of replication truncates their telomeres. Telomeres completely eroded by excessive rounds of replication are recognized by the

DNA damage machinery which induces a persistent DNA damage response that drives the cell into a state of permanent cell cycle referred to as replicative senescence. These cellular responses are unique departures from the mammalian cell cycle and function as barriers to tumorigenesis (Bartkova et al. 2006). Interestingly, cell fate determination after DNA damage is strongly influenced by extracellular cues, such as cytokines and growth factors, implying that a multivariate systems approach is necessary for understanding how cells couple extracellular signal transduction to the DNA damage response and the cell cycle machinery to commit to these cell fates.

Cross-References

- ▶ Apoptosis
- ▶ Cdk Inhibitors
- ▶ Cell Cycle
- ▶ Cell Cycle Checkpoints
- ▶ Cell Cycle Model Analysis, Bifurcation Theory
- ▶ Cell Cycle Modeling, Differential Equation
- ▶ Cell Cycle Models, Sensitivity Analysis
- ▶ Cell Cycle Transitions, G2/M
- ▶ Cell Cycle Transitions, Mitotic Exit
- ▶ Cell Cycle, Budding Yeast
- ▶ Cell Cycle, Cancer Cell Cycle and Oncogene Addiction
- ▶ Cell Cycle, Fission Yeast
- ▶ Cytokines
- ▶ DNA Repair
- ▶ DNA Replication
- ▶ Endoreplication
- ▶ Mitosis
- ▶ Mitotic Kinases
- ▶ Senescence
- ▶ Signal Transduction Pathway
- ▶ Transcription

References

Bartkova J et al (2006) Oncogene-induced senescence is part of the tumorigenesis barrier imposed by DNA damage checkpoints. Nature 444:633–637

Chen WW, Schoeberl B, Jasper PJ, Niepel M, Nielsen UB, Lauffenburger DA, Sorger PK (2009) Input-output behavior of ErbB signaling pathways as revealed by a mass action model trained against dynamic data. Mol Syst Biol 5:239

Hay N, Sonenberg N (2004) Upstream and downstream of mTOR. Genes Dev 18:1926–45

Schoeberl B, Pace EA, Fitzgerald JB, Harms BD, Xu L, Nie L, Linggi B, Kalra A, Paragas V, Bukhalid R, Grantcharova V, Kohli N, West KA, Leszczyniecka M, Feldhaus MJ, Kudla AJ, Nielsen UB (2009) Therapeutically targeting ErbB3: a key node in ligand-induced activation of the ErbB receptor-PI3K axis. Sci Signal 2:ra31

Cell Cycle Ontology (CCO)

Erick Antezana, Vladimir Mironov and Martin Kuiper
Department of Biology, Systems Biology Group, Norwegian University of Science and Technology, Trondheim, Norway

Synonyms

Application ontology; Domain ontology

Definition

Application ontologies are engineered for a specific research area or domain. They often use canonical ontologies to construct ontological classes and relationships between classes. Ontologies (▶ Ontology) are a means by which knowledge is captured according to a common conceptualization of a domain, and they are commonly used within the molecular biology subdiscipline of the life sciences (Stevens and Lord 2008). The cell cycle ontology (CCO, Antezana et al. 2009) is an application ontology, developed to integrate knowledge of the eukaryotic cell cycle process (▶ Cell Cycle). It facilitates scientific discovery in the area of cell cycle research.

Characteristics

Requirements of CCO

CCO brings together a variety of knowledge from the cell cycle domain, following a standard ontology-based integration paradigm. The knowledge in this

domain is constantly evolving, and covers a variety of organisms. This poses a range of requirements:

1. CCO integrates cell cycle knowledge across several model organisms: *H. sapiens, S. cerevisiae, S. pombe*, and *A. thaliana*.
2. It presents a common conceptual view of this knowledge, permitting sophisticated querying in a variety of forms, across cell cycle knowledge.
3. CCO accommodates frequent changes in data, by frequent rebuilds through an automated pipeline.
4. CCO performs automatic mapping of biological entities.

Contents of CCO

The cell cycle ontology covers the following topics:

1. The cell cycle proteins
2. The genes coding for cell cycle proteins
3. The molecular function of cell cycle proteins
4. The cellular localization of cell cycle proteins
5. Cellular processes associated with cell cycle proteins
6. Post-translational modifications of cell cycle proteins
7. Protein-protein interactions among cell cycle proteins
8. Phylogenetic information for the supported organisms through orthology relationships among cell cycle proteins

The cell cycle ontology integrates the following resources:

1. Genes, proteins, and their post-translational modifications from the UniProt database
2. The cell cycle, cell division, cell proliferation, DNA replication branches of the gene ontology biological process branch (▶ Gene Ontology)
3. The complete molecular function branch from the gene ontology
4. The complete cellular component branch from the gene ontology
5. The relation ontology provided by the Open Biomedical Ontologies Consortium (OBO) (Smith et al. 2007)
6. The gene ontology annotations (Camon et al. 2004) of UniProt proteins for the cell cycle (these are the attachments of GO Cellular Component concepts to protein to describe their location); GO biological process and molecular function concepts to describe their functionality
7. Biological species relationships as described by the NCBI taxonomy (▶ NCBI BioProject Genome Resources)
8. Protein-protein interactions involving cell cycle proteins from the IntACT database (▶ Protein–Protein Interaction Databases)

Core Concepts of CCO

Protein, Modified protein, Gene, Molecular interaction, Molecular function, Biological process, Cellular component, and Organism. All these entities in the CCO (proteins – including their modified forms, genes, interactions, etc.) are modeled as classes since they gather shared commonalities that are present in all the individuals (instances) they represent (Fig. 1).

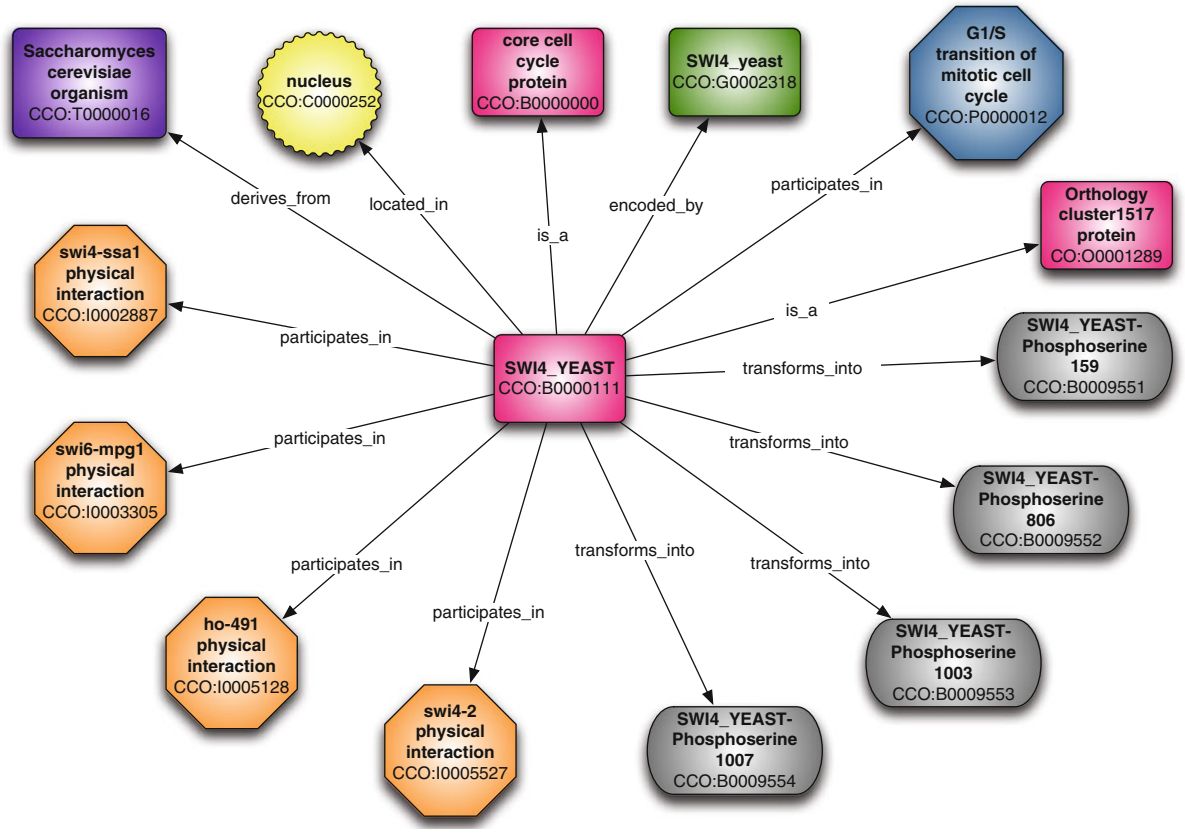
Upper Level Ontology

An upper level ontology (ULO) connects the integrated resources. A ULO is an ontology that structures general types of concepts (such as a process) in generic as well as specific domains to provide an integration scaffold for including other ontologies. The CCO-ULO is based on the basic formal ontology (BFO, Grenon et al. 2004) to ensure interoperability of the CCO with ontologies from OBO. The CCO-ULO has been customized for the CCO by inclusion of a few high-level concepts, such as "cell cycle protein" (see Fig. 2).

Construction of CCO

The cell cycle ontology is constructed in several steps:

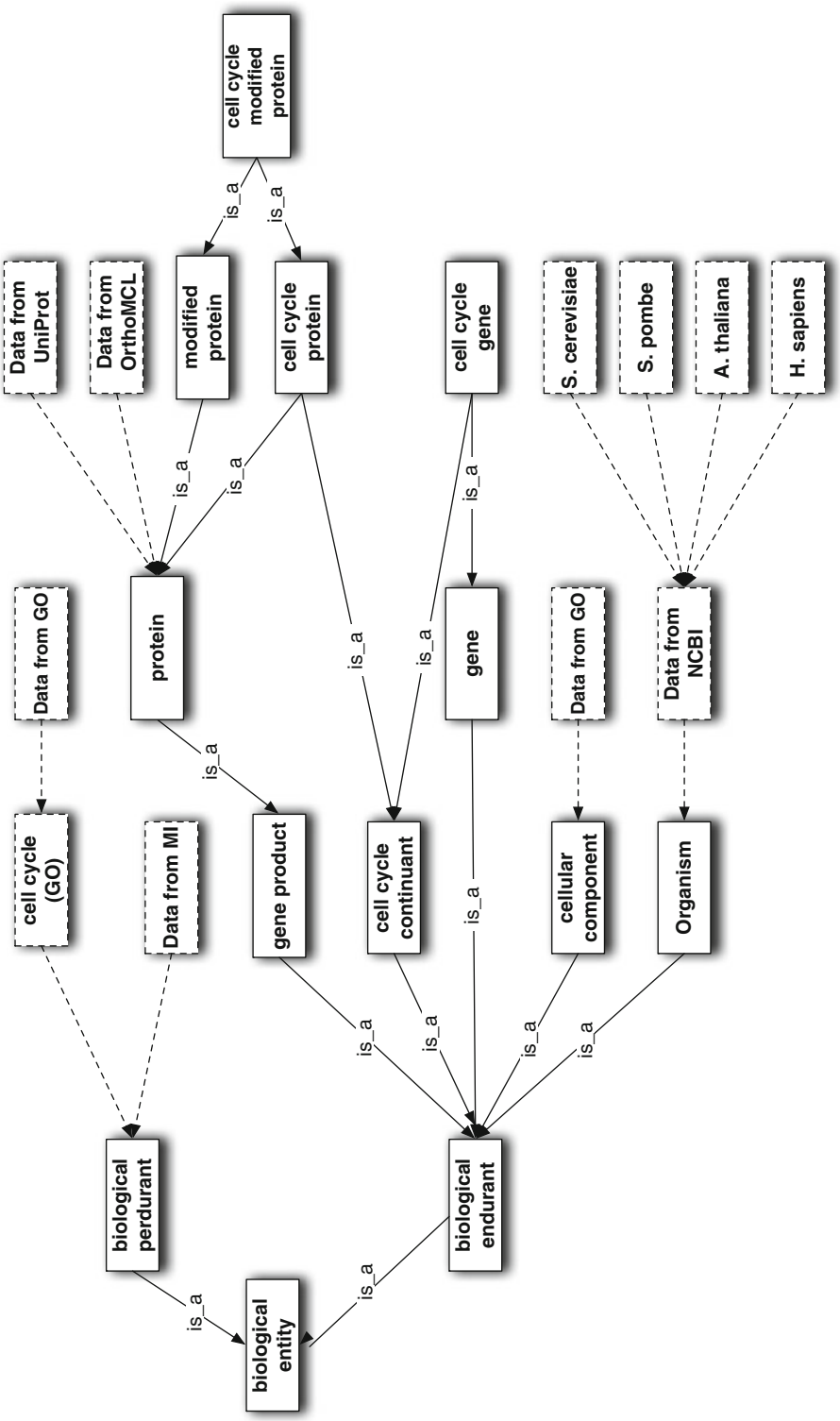
1. The process starts by selecting portions from various OBO source ontologies. This "pre-cell cycle ontology" constitutes a backbone for the complete CCO and the four species-specific sub-ontologies thereof.
2. The OBO relations ontology is fully incorporated, together with the "interaction type" branch from the molecular interaction ontology from OBO.
3. A specific taxonomy is built on the basis of the NCBI taxonomy for *H. sapiens, S. cerevisiae, S. pombe*, and *A. thaliana*.
4. Protein "hubs" that connect their relevant annotation data (such as the molecular functions in which they participate) are generated. All proteins described with cell cycle concepts in gene ontology annotation (GOA) files (Camon et al. 2004) are named "core cell cycle proteins."



Cell Cycle Ontology (CCO), Fig. 1 Protein-centric knowledge in CCO. The local neighborhood of the cell cycle protein wee1 (human) is shown. Entities (*color coded*) and relationships (*arrows*) illustrate the types of knowledge in CCO

These proteins are added to the CCO as the children of the concept core cell cycle protein (CCO: B0000000) and used as seeds for the data integration process.

5. The "core cell cycle" protein section of the CCO is expanded to include the proteins known to interact with the core cell cycle proteins, as documented in IntACT (Kerrien et al. 2007). These new protein classes are included as subclasses of the class cell cycle protein. Then, protein information (such as synonym names, encoding genes, and cross-references) is fetched from the UniProt knowledge base. In addition, post-translational modification data, when available in UniProt, are also added by creating new concepts defined by their specific modification.
6. Clusters of putative orthologs of the four species are generated with the OrthoMCL clustering utility (Li et al. 2003). This tool is used to infer evolutionary relationships between classes of proteins. The proteins from the clusters containing at least one core cell cycle protein are added to the CCO as subclasses of the class cell cycle protein. All the imported entries from the sources are cross-referenced so that the data can be traced back to its source.
7. The four organism-specific ontologies and the composite cell cycle ontology are checked and made available producing the official release of the system.
8. Now that the CCO is available with all the knowledge in place, semantic enrichment can occur. We use the Ontology PreProcessor Language (OPPL, http://oppl2.sourceforge.net/) to further transform the CCO with application of ontology design patterns.



Cell Cycle Ontology (CCO), Fig. 2 Upper level ontology (ULO) for the CCO. The ULO is a hierarchical scaffold including generic concepts (e.g., cell cycle gene) which connects the integrated resources. The *dashed rectangles* represent the type of data residing below the parental concepts: the node labeled "data from GO" shows where the concept from GO's cellular component ontology goes under the concept cellular component (e.g., nucleus), the node "data from UniProt" under the concept protein shows where protein data from UniProt resides, and so forth

Cell Cycle Ontology (CCO), Table 1 Numbers of classes, breadth, and depth of the CCO hierarchy

Structure and cohesion of CCO	
Metric	Asserted CCO
No. of classes	89,532
No. of leaf classes	85,235
Max. depth	33
Mean depth	15.35
Mean number of subclasses per class	4.61
Max. number of subclasses	24,021
No. of properties	52
Property with maximum usage	has_source used 65 110 times
Mean usage per property	6,673.69
Mean property usage per class	3.87

9. Finally, validation and verification processes are carried out while building the CCO to ensure its soundness. The ontologies generated in the previous phase are manually and automatically checked by ontology editors, validators, and reasoners. In addition, the pipeline log execution files are inspected in detail. These files are sufficiently detailed to point out any possible problem. This has allowed us to assemble a fully automated pipeline that uploads the ontologies and their exports in the different formats to the CCO website and to all related and supporting repositories (Table1).

Examples of Other Application Ontologies

The experimental factor ontology (EFO, Malone et al. 2010), developed by the European Bioinformatics Institute (EBI), represents sample variables from gene expression experimental data. The NIFSTD ontology, developed by the NeuroInformatics Framework (NIF, Bug et al. 2008), has produced an inventory of Web-based neuroscience resources, integrating an ontology with separate modules covering major domains of neuroscience: anatomy, cell, subcellular, molecule, function and dysfunction, and concepts for describing experimental techniques, instruments, and other resources.

Cross-References

- ▶ Cell Cycle
- ▶ Gene Ontology
- ▶ NCBI BioProject Genome Resources
- ▶ Ontology
- ▶ Protein–Protein Interaction Databases

References

Antezana E, Egaña M, Blondé W, Illarramendi A, Bilbao I, De Baets B, Stevens R, Mironov V, Kuiper M (2009) The cell cycle ontology: an application ontology for the representation and integrated analysis of the cell cycle process. Genome Biol 10:R58

Bug WJ, Ascoli GA, Grethe JS, Gupta A, Fennema-Notestine C, Laird AR, Larson SD, Rubin D, Shepherd GM, Turner JA, Martone ME (2008) The NIFSTD and BIRNLex vocabularies: building comprehensive ontologies for neuroscience. Neuroinformatics 6:175–194

Camon E, Magrane M, Barrell D, Lee V, Dimmer E, Maslen J, Binns D, Harte N, Lopez R, Apweiler R (2004) The gene ontology annotation (GOA) database: sharing knowledge in uniprot with gene ontology. Nucleic Acids Res 32(Database issue):D262–266

Grenon P, Smith B, Goldberg L (2004) Biodynamic ontology: applying BFO in the biomedical domain. Stud Health Technol Inform 102:20–38

Kerrien S, Alam-Faruque Y, Aranda B, Bancarz I, Bridge A, Derow C, Dimmer E, Feuermann M, Friedrichsen A, Huntley R, Kohler C, Khadake J, Leroy C, Liban A, Lieftink C, Montecchi-Palazzi L, Orchard S, Risse J, Robbe K, Roechert B, Thorneycroft D, Zhang Y, Apweiler R, Hermjakob H (2007) IntAct - open source resource for molecular interaction data. Nucleic Acids Res 35(Database issue):D561–565

Li L, Stoeckert CJ Jr, Roos DS (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res 13:2178–2189

Malone J, Holloway E, Adamusiak T, Kapushesky M, Zheng J, Kolesnikov N, Zhukova A, Brazma A, Parkinson H (2010) Modeling sample variables with an experimental factor ontology. Bioinformatics 26:1112–1118

Smith B, Ashburner M, Rosse C, Bard J, Bug W, Ceusters W, Goldberg LJ, Eilbeck K, Ireland A, Mungall CJ, OBI Consortium, Leontis N, Rocca-Serra P, Ruttenberg A, Sansone SA, Scheuermann RH, Shah N, Whetzel PL, Lewis S (2007) The OBO foundry: coordinated evolution of ontologies to support biomedical data integration. Nat Biotechnol 25:1251–1255

Stevens R, Lord P (2008) Application of ontologies in bioinformatics. In: Staab S, Studer R (eds) Handbook on ontologies in information systems. Springer, Berlin

Cell Cycle Progression in Bacteria

- ▶ Cell Cycle, Prokaryotes

Cell Cycle Regulation, microRNAs

Baltazar D. Aguda[1] and Gregory Riddick[2]
[1]Mathematical Biosciences Institute, Ohio State University, Columbus, OH, USA
[2]Neuro-Oncology Branch, National Cancer Institute, Bethesda, MD, USA

Synonyms

Post-transcriptional regulation of gene expression

Definition

MicroRNAs are short (~21 nucleotides) non-coding RNAs that are endogenous in the genomes of animals and plants, and have been shown to induce cleavage or suppression of translation of their target mRNAs. Thus, microRNAs are generally known as post-translational negative regulators of gene expression, although there are a few recent reports that some microRNAs act as positive regulators. The mRNA target is recognized through binding of a seed sequence in the microRNA to a complementary sequence in the mRNA. Using this sequence complementarity, several computer programs have been created to predict microRNA targets. Some examples of the more popular programs publicly available on the Internet are the following:

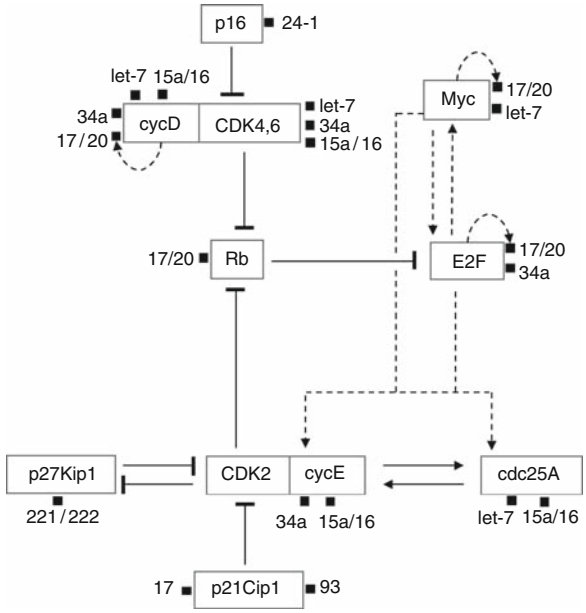
TargetScan (http://www.targetscan.org/)
Microcosm (http://www.ebi.ac.uk/enright-srv/microcosm/htdocs/targets)
mirWalk (http://www.ma.uni-heidelberg.de/apps/zmf/mirwalk/).

It has been estimated that the expression of at least a third of human genes is post-transcriptionally regulated by microRNAs. To date, there are more than 700 microRNAs identified in the human genome. The complexity of the regulation of gene expression by microRNAs stems from the possibility of multiple targets per microRNA (in some cases, a single microRNA is predicted to target up to several hundred mRNAs) and the possibility of multiple microRNAs targeting per gene.

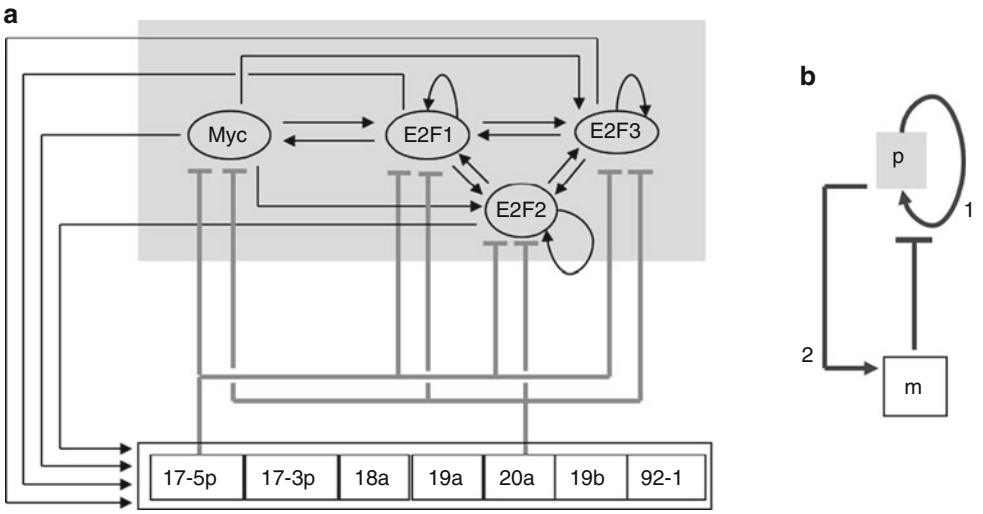
Characteristics

Cell Cycle Genes Directly Regulated by microRNAs

Shown in Fig. 1 are some of the experimentally validated microRNAs targeting key cell cycle genes involved in the early phases (G1 and S) of the mammalian cell cycle (Chivukula and Mendell 2008; Yu et al. 2010). Myc and E2F are transcription factors that induce expression of cell cycle genes such as cyclins (e.g., cyclin E) and cyclin-dependent kinases (e.g., CDK4). Note that Myc and E2F promote each other's expression. As shown in Fig. 1, these transcription factors also induce the expression of miR-17/20 (which is the abbreviated name for two microRNAs, namely, miR-17-5p and miR-20a). In return, miR-17/20 target and inhibit the expression of Myc and E2F, thus forming negative



Cell Cycle Regulation, microRNAs, Fig. 1 MicroRNAs targeting key cell cycle genes in G1 and S phases. Cell cycle genes are shown in boxes. Gene abbreviations: cycD (cyclin D), cycE (cyclin E), Myc (c-myc), Rb (Retinoblastoma protein), E2F (E2F1, E2F2, E2F3). Small black squares indicate microRNAs targeting the genes beside them. Labels beside black squares give the shortened names of the microRNAs (e.g., 17/20 means miR-17 and miR-20). Dashed arrows mean transcriptional induction of gene expression (e.g., cycD, Myc, and E2F induce the expressions of miR-17, and miR-20). Qualitative interactions (edges between genes) are symbolized by hammerheads (inhibition) and arrows (activation)



Cell Cycle Regulation, microRNAs, Fig. 2 Modeling the negative feedback loop between miR-17/20 and Myc/E2F. **(a)** Shown inside the gray box are the proteins Myc, E2F1, E2F2, and E2F3 inducing each other's expression. All of these proteins also induce the expression of the miR-17-92 cluster pictured by the row of seven microRNAs at the bottom of the figure. Two members of the cluster, miR-17-5p and miR-20a, target and inhibit Myc and E2F1-3. **(b)** A reduced model involving an autocatalytic protein module, p, and a microRNA module, m

feedback loops. A third negative feedback shown in Fig. 1 is between miR-17/20 and cyclin D.

Mathematical Modeling of the Negative Feedback Loops Involving miR-17/20 and Myc/E2F

The microRNA cluster called miR-17-92 is composed of seven microRNAs as shown in the bottom of Fig. 2a. Members of this cluster, particularly miR-17-5p and miR-20a, have been implicated in various human cancers – that is, their specific deletions in the genome or overexpressions are associated with cancer initiation (Chivukula and Mendell 2008). The negative feedback loop between miR-17/20 and Myc/E2F was modeled by Aguda et al. (2008) to explore how changes in miR-17/20 expression affect Myc/E2F activation (and thereby the rate of G1-to-S transition in the cell cycle). The detailed network shown in Fig. 2a was abstracted to the two-variable model shown in Fig. 2b. This simple model predicts that there is an all-or-none switching behavior of Myc/E2F activity as miR-17/20 expression is increased. This bistable switch is generated by the autocatalytic nature of the protein module, p, as indicated by step 1 in Fig. 2b. The model further predicts that oscillations in the activities of p and m are possible due to the presence of the negative feedback loop between p and m.

References

Aguda BD, Kim Y, Piper-Hunter MG, Friedman A, Marsh CB (2008) MicroRNA regulation of a cancer network: consequences of the feedback loops involving miR-17-92, E2F, and Myc. Proc Natl Acad Sci USA 105(50):19678–19683

Chivukula RR, Mendell JT (2008) Circular reasoning: microRNAs and cell-cycle control. Trends Biochem Sci 33(10):474–481

Yu Z, Baserga R, Chen L, Wang C, Lisanti MP, Pestell RG (2010) microRNA, cell cycle, and human breast cancer. Am J Pathol 176(3):1058–1064

Cell Cycle Signaling, Hypoxia

Bernhard Schmierer
Department of Biochemistry, Oxford Centre for Integrative Systems Biology (OCISB), University of Oxford, Oxford, UK

Definition

Hypoxia refers to a state of inadequate oxygen supply in cells, tissues, or organisms. Among many other cellular processes, hypoxia affects cell cycle progression.

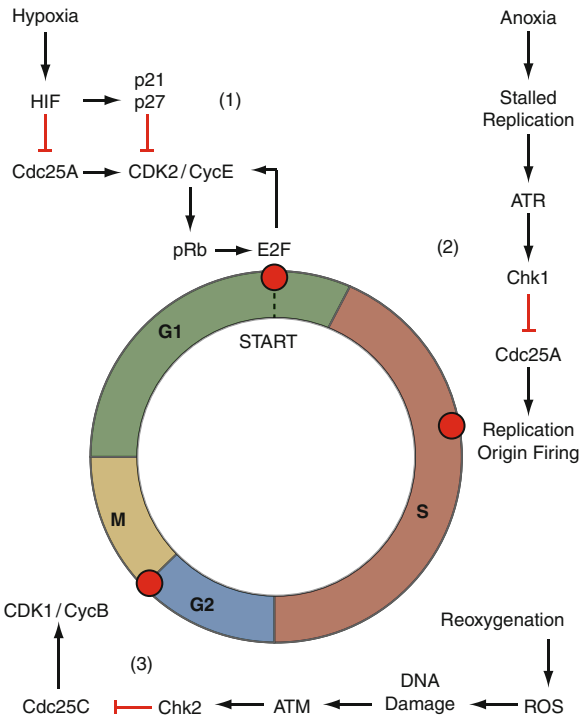
Characteristics

Oxygen is the terminal electron acceptor of the mitochondrial respiratory chain, and an adequate oxygen supply is a fundamental requirement for the survival and proper function of all metazoan cells (Semenza 2007). Higher animals have evolved sophisticated organ systems such as the cardiovascular and respiratory systems to supply all body tissues with sufficient amounts of oxygen at all times. Disruption of adequate oxygen supply leads to tissue hypoxia, a condition which features prominently in pathophysiological contexts such as anemia, ▶ ischemic disease, and tumor formation. However, hypoxic episodes may also be encountered in physiological contexts such as embryonic development, exercise, or altitude adaptation. At the cellular level, the intracellular oxygen tension is continuously monitored. Many cell types are able to adapt to hypoxic conditions by systemic strategies such as metabolic reprogramming and energy preservation. The former strategy is exemplified by the upregulation of the glycolytic pathway and by the shutdown of mitochondrial respiration; the latter strategy includes the inhibition of protein translation, cell growth, and cell proliferation. By secreting pro-angiogenic growth factors, cells signal their hypoxic status to the surrounding tissue to induce blood vessel formation and to restore the oxygen supply to the hypoxic area. In cases where these adaptive strategies fail, prolonged hypoxia can lead to either cell necrosis or programmed cell death.

A substantial proportion of the cellular adaptation to hypoxia in animals is regulated by the αβ-heterodimeric transcription factor hypoxia-inducible factor (HIF), which is thought to be a master regulator of the hypoxic response (Semenza 1999). HIF is activated by low oxygen tensions. Briefly, in normoxic conditions, the HIFα-subunit is highly unstable owing to its post-translational hydroxylation, which is catalyzed by the prolyl-hydroxylase domain (PHD or EGLN) family of oxygen-dependent dioxygenases (Schofield and Ratcliffe 2004). Hydroxylation of two proline residues in HIFα by these enzymes very substantially increases the interaction of HIFα with the tumor suppressor protein von Hippel Lindau (▶ *PVHL*), the targeting component of an E3-ubiquitin ligase complex that triggers rapid proteasomal degradation of HIFα. In hypoxic conditions, when oxygen becomes limiting for the PHD-dependent hydroxylation reaction, HIFα escapes degradation and accumulates. Binding of HIFα to HIFβ, which is constitutively present and identical to the aryl hydrocarbon receptor nuclear translocator (ARNT) then yields the transcriptionally active, heterodimeric transcription factor. This mechanism is evolutionarily conserved in all metazoans and has been put forward as the primary cellular oxygen-sensing mechanism (Kaelin and Ratcliffe 2008).

In most primary and also immortalized cell lines, one of the functional consequences of the HIF-induced transcriptional program is the inhibition of cell proliferation, caused by a cell cycle arrest. The point in the cell cycle at which the arrest occurs seems to depend on the severity of the hypoxic stress. Broadly speaking, moderate hypoxia leads to an arrest in the G1 phase of the cell cycle, which is mostly reversible upon ▶ reoxygenation. Severe hypoxia or anoxia can additionally arrest cells in S-phase as a consequence of the hypoxic activation of the DNA damage checkpoint. Interestingly, this occurs in the absence of any detectable DNA damage. Finally, reoxygenation after stringent hypoxia can cause DNA single- and double-strand breaks via the formation of reactive oxygen species, leading to a G2 arrest downstream of the DNA damage checkpoint. To date, insight into mechanistic aspects of the link between hypoxia and reduced cell cycle progression is limited, as is the understanding of its functional consequences. The following section briefly reviews the available information and attempts to connect isolated observations into a more coherent picture (Fig. 1).

The best characterized effect of hypoxia on the regulatory network controlling cell cycle progression is the HIF-dependent transcriptional induction of the cyclin-dependent kinase inhibitors (CDKIs) (▶ CDK Inhibitors) p21 (also known as CDKN1A) and p27 (also known as CDKN1B). In the G1 phase of the cell cycle, these CDKIs inhibit CDK2/Cyclin E, and prevent the phosphorylation of the Retinoblastoma protein and thus the release of the E2F-transcription factors, arresting the cells at the restriction point (▶ Cell Cycle Transition, Principles of Restriction Point; ▶ Cell Cycle Transition, Detailed Regulation of Restriction Point) in the G1 phase of the cell cycle (Morgan 2007). The restriction point is the point in G1 after which mammalian cells are committed to completing their division cycle, even in the absence of growth factors. Mechanistically, HIF is thought to



Cell Cycle Signaling, Hypoxia, Fig. 1 Cartoon representation of the signaling pathways causing cell cycle arrest at different points in the cell cycle (*red circles*) in response to hypoxia and reoxygenation. (**1**) Moderate hypoxia arrests cells at the restriction point (START) of the G1-phase in a HIF-dependent manner. pRB indicates the phosphorylated Retinoblastoma protein. (**2**) Severe hypoxia or anoxic conditions additionally cause an arrest in S-phase. This is most likely caused by an inhibition of replication origin firing downstream of a signaling cascade triggered by stalled replication forks, which themselves may be a consequence of insufficient metabolite concentrations or an unfavorable energy charge. (**3**) Finally, reoxygenation after severe hypoxia produces significant levels of reactive oxygen species (ROS), which can lead to DNA damage and activation of the DNA damage checkpoint. This prevents entry into mitosis and arrests cells in the G2-phase. For details, see text

upregulate both the *CDKN1A* and *CDKN1B* genes by counteracting ▶ *c-Myc*, which normally represses these genes (Dang et al. 2008). Conversely, the phosphatase CDC25A, an activator of CDK/cyclin complexes and thus a positive regulator of cell cycle progression, is normally activated by Myc. In hypoxic conditions, HIF represses *CDC25A* by counteracting Myc (Huang 2008). All these regulatory events contribute to a reversible G1 arrest in hypoxia: Upon restoration of sufficiently high oxygen levels, HIFα is rapidly degraded, its effects are abolished, and the cell cycle proceeds through the restriction point. In addition to this direct regulation of the cell cycle machinery by HIF, a number of other HIF-dependent, but less direct, effects likely contribute to this G1 arrest. Examples are a reduced protein translation, slowed cellular mass gain, and extensive metabolic reprogramming, processes which, at least in part, have been attributed to the HIF/Myc antagonism.

In addition to the reversible G1 arrest observed in response to moderate hypoxia, very stringent hypoxia or ▶ *anoxia* additionally activates components of the DNA damage checkpoint. This occurs in the absence of any detectable DNA damage and is possibly a DNA damage-independent process, most likely triggered by a hypoxia-induced replication arrest. Among other possible mechanisms, stringent hypoxia has the potential to cause a DNA replication arrest by HIF-dependent repression of the gene *CAD*, which codes for multiple enzymatic activities that are required for pyrimidine biosynthesis (Gordan et al. 2007), a metabolic pathway essential for DNA replication. Interestingly, another enzyme required for pyrimidine biosynthesis is strictly oxygen dependent (Loffler 1989). Thus, repression of this biosynthetic pathway has the potential to contribute to a replication arrest in response to severe hypoxia in both a HIF-dependent and a HIF-independent manner. Stalled replication forks then trigger the activation of the kinase ATR, which causes an S-phase or replication arrest in response to stringent hypoxia. This arrest is accomplished by the prevention of replication origin firing by CHK1-dependent inhibition of CDC25A and the consequent high activity of cyclin/CDK2 complexes, which is required for the initiation of replication forks (Morgan 2007). Signaling downstream of stalled replication can also induce apoptosis by activating p53 (Hammond et al. 2007). The evolutionary advantage of an activation of the DNA damage checkpoint under hypoxic conditions is disputed; however, it has been speculated that this occurs in anticipation of the DNA damage caused by the production of reactive oxygen species (ROS) during reoxygenation. Such reoxygenation-induced DNA damage triggers ATM-dependent CHK2 activation, and phosphorylation and inactivation of the phosphatase CDC25C, which normally allows mitotic entry by activating CDK1/CycB complexes. In this way, ROS ultimately arrest cells in the G2 phase of the cell cycle (Hammond et al. 2007).

In summary, insufficient oxygen levels block cell cycle progression in distinct phases of the cycle by various mechanisms. The consequences on cell fate depend on the cell type and the degree and duration of the hypoxic stress, ranging from a perfectly reversible, transient arrest to the induction of apoptotic cell death.

Cross-References

- Anoxia
- Cdk Inhibitors
- Cell Cycle Arrest After DNA Damage
- Cell Cycle Checkpoints
- Cell Cycle of Mammalian Cells
- Cell Cycle Signaling, Metabolic Pathway
- Cell Cycle Transition, Principles of Restriction Point
- Cell Cycle, Cancer Cell Cycle and Oncogene Addiction
- c-Myc
- Ischemic Disease
- PVHL
- Reoxygenation

References

Dang CV, Kim JW, Gao P, Yustein J (2008) The interplay between MYC and HIF in cancer. Nat Rev Cancer 8(1):51–56

Gordan JD, Thompson CB, Simon MC (2007) HIF and c-Myc: sibling rivals for control of cancer cell metabolism and proliferation. Cancer Cell 12(2):108–113

Hammond EM, Kaufmann MR, Giaccia AJ (2007) Oxygen sensing and the DNA-damage response. Curr Opin Cell Biol 19(6):680–684

Huang LE (2008) Carrot and stick: HIF-alpha engages c-Myc in hypoxic adaptation. Cell Death Differ 15(4):672–677

Kaelin WG Jr, Ratcliffe PJ (2008) Oxygen sensing by metazoans: the central role of the HIF hydroxylase pathway. Mol Cell 30(4):393–402

Loffler M (1989) The biosynthetic pathway of pyrimidine (deoxy)nucleotides: a sensor of oxygen tension necessary for maintaining cell proliferation? Exp Cell Res 182(2):673–680

Morgan DO (2007) The cell cycle. Principles of control. New Science Press Ltd., London

Schofield CJ, Ratcliffe PJ (2004) Oxygen sensing by HIF hydroxylases. Nat Rev Mol Cell Biol 5(5):343–354

Semenza GL (1999) Regulation of mammalian O_2 homeostasis by hypoxia-inducible factor 1. Annu Rev Cell Dev Biol 15:551–578

Semenza GL (2007) Life with oxygen. Science 318(5847):62–64

Cell Cycle Signaling, Metabolic Pathway

Fabian Rudolf and Joerg Stelling
Department of Biosystems Science and Engineering, ETH Zurich, Basel, Switzerland

Definition

Cell proliferation requires growth and division. Growth is connected to cell cycle progression on multiple levels, such as the metabolic state, nutrient signaling programs, and cell homeostasis. An illustration of this linkage is the Warburg effect. Cancer cells propagate by exclusively using an anaerobic metabolic program. Forcing them to an aerobic program slows down the cell cycle and thereby inhibits cancer progression (Hsu and Sabatini 2008). This effect is not only used for diagnostic purposes, but also as a cancer treatment.

Cells seem to segregate their metabolic activity to different phases in the cell cycle. *S. cerevisiae* cells grown under carbon limitation in a chemostat can show robust oscillations in their oxygen consumption as they alternate between glycolytic and respiratory metabolism, and their cell cycle is synchronized to these oscillations (Tu et al. 2005). The metabolic cycle can be divided into three major phases: oxidative, reductive/building, and reductive/charging, where S phase of the cell cycle is restricted to the reductive phases. Interestingly, individual *S. cerevisiae* cells grown under similar conditions show the same coupling (Silverman et al. 2010). This synchrony provides a unique experimental approach for further studying the coupling of metabolism and cell cycle.

Characteristics

Nutrient Availability and Cell Cycle Progression

Metabolism can regulate the cell cycle by different means, mostly by controlling the availability of crucial molecules. For example, the cycle arrests or slows down in S phase upon depletion of any nucleotide via drugs or removal of nucleotide biosynthesis pathways (Brauer et al. 2008). Similar effects are observed during M phase upon reduced lipid biosynthesis. However, we have to consider

possible laboratory artifacts since *S. cerevisiae* mutants in nucleotide biosynthesis show a defective metabolic regulation similar to the Warburg effect when the corresponding nucleotide is limiting. Interestingly, *S. cerevisiae* cells limited for any nutrient in its basic form (P, N, S = inorganic salts, C = carbohydrates) correctly regulate metabolism and therefore sensing mechanisms link metabolism and cell cycle progression (Brauer et al. 2008).

Nutrient signaling influences the cell cycle state at multiple points. Unicellular organisms (and presumably stem cells) measure their size before they commit to a new round of division (Fantes et al. 1975). While the nature of the size signal remains unclear, cells are able to sense key metabolites; in bacteria, genetic evidence suggests that cells measure their nucleotide content (Weart et al. 2007) while simple eukaryotes such as *S. cerevisiae* measure stored carbohydrates in the form of trehalose (Paalman et al. 2003). As for higher eukaryotes not much is known. Interestingly, transcriptional regulation of metabolic pathways seems to be coupled with the cell cycle. In *S. cerevisiae*, the P, N, and S pathways are upregulated around "Start" (the equivalent of the restriction point in mammalian cells) in G1 (Tu et al. 2005). It is not clear whether this is directly required for progression, and if and how metabolite sensing mechanisms link to the cell-size checkpoint.

Moreover, cells adjust the timing of the cell cycle to nutrient availability (Fantes and Nurse 1977). Often, the phase where the nutrient limitation is sensed is affected and extended. For example, *S. cerevisiae* cell cycle length differs on respiratory compared to fermentative carbon sources, and the timing correlates with the energy gained from these sources. In general, slower mass accumulation under limiting C, S, or P sources predominantly either expands G1 (*S. cerevisiae* and cultured cells from multicellular organisms) or G2 (*S. pombe*), but effects on other phases cannot be ruled out (especially in *S. pombe*). In contrast, N limitation affects all phases and the cell cycle arrest point in *S. pombe* is less well defined than for the other nutrients (Sveiczer et al. 2004).

All those nutrients are sensed within the cell cycle at the cell-size checkpoint called "Sizer." In *S. cerevisiae*, *D. melanogaster*, and human cells, size control is effective at "Start," whereas in *S. pombe* it is in G2 before entering M phase. The molecular details of "Sizer" are unknown in all systems. In *S. cerevisiae*, ribosome biogenesis, translational efficiency, and carbohydrate storage appear to be sensed. This is consistent with the observation that cell cycle progression depends on translational regulation of key cell cycle proteins. In addition, *S. pombe* employs a length-sensing network by limiting key cell cycle inhibitors to the growing tips (Moseley et al. 2009).

Nutrient-Controlled Signaling Pathways

Metabolism and the available nutrients generate signals that are received by the cell cycle, with the cAMP/PKA, AMPK/Snf, and TOR pathways being the most prominent signaling pathways. It is assumed that the first two pathways are controlled by carbon sources (Santangelo 2006) while the TOR pathway is downstream of different nitrogen sources (Wullschleger et al. 2006). For S, the signal seems to be directly linked to a misregulation of cystein and methionin biosynthesis. Regulation of intracellular polyphosphate levels or direct control of cyclin translation conveys P availability.

AMP-kinase is activated by high AMP levels, which occur upon energy shortage. AMPK and cell cycle are linked around "Start" and overexpression of hyperactivated AMPK leads to a G1/S arrest, while loss of AMPK function results in small cells and ultimately cell death (Baena-González et al. 2007). In higher eukaryotes, AMPK controls p53 and thereby delays the G1/S transition. In *S. cerevisiae*, the evidence is less clear, but AMPK seems to directly control the transcription factor (TF) necessary for B type cyclin synthesis and therefore S phase entry. It is unclear whether this direct transcriptional effect exists in higher eukaryotes.

High cAMP levels activate PKA through direct binding to its regulatory domain. In contrast to AMPK, energy shortage inactivates PKA through a drop in cAMP. *S. cerevisiae* cells with defective cAMP metabolism arrest at "Start," and addition of exogenous cAMP rescues this defect (Tamaki 2007). PKA influences the cell cycle by multiple processes: regulation of cyclin transcription, of translation efficiency, and of protein degradation. Interpreting the PKA effects is not trivial because the phenotypes are often pleiotropic and not necessarily related to cell cycle defects. Moreover, in *S. cerevisiae*, PKA can be bypassed by deleting TFs not involved in cell cycle progression. Interestingly, carbon-limited *S. cerevisiae*

cells activate PKA predominately in G1 as cAMP increases shortly before "Start." This seems to be necessary for cell cycle progression and it is controlled by the trehalose pathway.

TOR regulates protein synthesis. While high doses of the TOR inhibitor rapamycin lead to a phase-independent cell cycle arrest, lower concentrations and conditional TOR loss-of-function alleles arrest the cycle at the respective nutrient restriction point. Similarly, cells with low TOR activity cannot commit to a new cycle but they still grow. While the exact molecular mechanism is unclear, data from *S. cerevisiae* suggest that translation efficiency is sensed because overexpression of key cell cycle and translational regulators can compensate for deficient TOR signaling (Danaie et al. 1999).

Systems Biology Approaches

Many genome-wide studies of the cell cycle were performed. Most of them do not specifically address the connection between metabolism and cell cycle, except the above studies on the metabolic cycle of budding yeast that combine transcriptome and metabolome data.

Most computational studies of the cell cycle either assume a direct control of the cell cycle by key metabolites such as nucleotides or energy equivalents, or they include a growth-dependent effect on general translational (Csikász-Nagy et al. 2008). While these approaches capture the cellular behavior, they await further detailing of the metabolic input. Therefore, key challenges are to mechanistically capture the connection between metabolic state and cell cycle as well as to unravel the cell-size checkpoint.

Cross-References

- ▶ Bifurcation
- ▶ Cell Cycle
- ▶ Cell Cycle Checkpoints
- ▶ Cell Cycle Modeling, Differential Equation
- ▶ Cell Cycle Transition, Detailed Regulation of Restriction Point
- ▶ Cell Cycle, Budding Yeast
- ▶ Cell Cycle, Cell Size Regulation
- ▶ Cell Cycle, Fission Yeast
- ▶ Disease Ontology
- ▶ DNA Replication
- ▶ Model Organism
- ▶ Single Cell Experiments
- ▶ Signal Transduction Pathway

References

Baena-González E, Rolland F, Thevelein JM, Sheen J (2007) A central integrator of transcription networks in plant stress and energy signalling. Nature 448:938–942

Brauer MJ, Huttenhower C, Airoldi EM, Rosenstein R, Matese JC, Gresham D, Boer VM, Troyanskaya OG, Botstein D (2008) Coordination of growth rate, cell cycle, stress response, and metabolic activity in yeast. Mol Biol Cell 19:352–367

Csikász-Nagy A, Novák B, Tyson JJ (2008) Reverse engineering models of cell cycle regulation. Adv Exp Med Biol 641:88–97

Danaie P, Altmann M, Hall MN, Trachsel H, Helliwell SB (1999) CLN3 expression is sufficient to restore G1-to-S-phase progression in Saccharomyces cerevisiae mutants defective in translation initiation factor eIF4E. Biochem J 340(Pt 1):135–141

Fantes P, Nurse P (1977) Control of cell size at division in fission yeast by a growth-modulated size control over nuclear division. Exp Cell Res 107:377–386

Fantes PA, Grant WD, Pritchard RH, Sudbery PE, Wheals AE (1975) The regulation of cell size and the control of mitosis. J Theor Biol 50:213–244

Hsu PP, Sabatini DM (2008) Cancer cell metabolism: Warburg and beyond. Cell 134:703–707

Moseley JB, Mayeux A, Paoletti A, Nurse P (2009) A spatial gradient coordinates cell size and mitotic entry in fission yeast. Nature 459:857–860

Paalman JWG, Verwaal R, Slofstra SH, Verkleij AJ, Boonstra J, Verrips CT (2003) Trehalose and glycogen accumulation is related to the duration of the G1 phase of Saccharomyces cerevisiae. FEMS Yeast Res 3:261–268

Santangelo G (2006) Glucose signaling in Saccharomyces cerevisiae. Microbiol Mol Biol Rev 70:253

Silverman SJ, Petti AA, Slavov N, Parsons L, Briehof R, Thiberge SY, Zenklusen D, Gandhi SJ, Larson DR, Singer RH et al (2010) Metabolic cycling in single yeast cells from unsynchronized steady-state populations limited on glucose or phosphate. Proc Natl Acad Sci USA 107:6946–6951

Sveiczer A, Novak B, Mitchison JM (2004) Size control in growing yeast and mammalian cells. Theor Biol Med Model 1:12

Tamaki H (2007) Glucose-stimulated cAMP-protein kinase a pathway in yeast Saccharomyces cerevisiae. J Biosci Bioeng 104:245–250

Tu BP, Kudlicki A, Rowicka M, McKnight SL (2005) Logic of the yeast metabolic cycle: temporal compartmentalization of cellular processes. Science 310:1152–1158

Weart RB, Lee AH, Chien AC, Haeusser DP, Hill NS, Levin PA (2007) A metabolic sensor governing cell size in bacteria. Cell 130:335–347

Wullschleger S, Loewith R, Hall MN (2006) TOR signaling in growth and metabolism. Cell 124:471–484

Cell Cycle Signaling, Spindle Assembly Checkpoint

Andrea Ciliberto[1] and Jagesh Shah[2]
[1]IFOM – The FIRC Institute of Molecular Oncology, Milan, Italy
[2]Department of Systems Biology, Harvard Medical School and Renal Division, Brigham and Women's Hospital, Boston, MA, USA

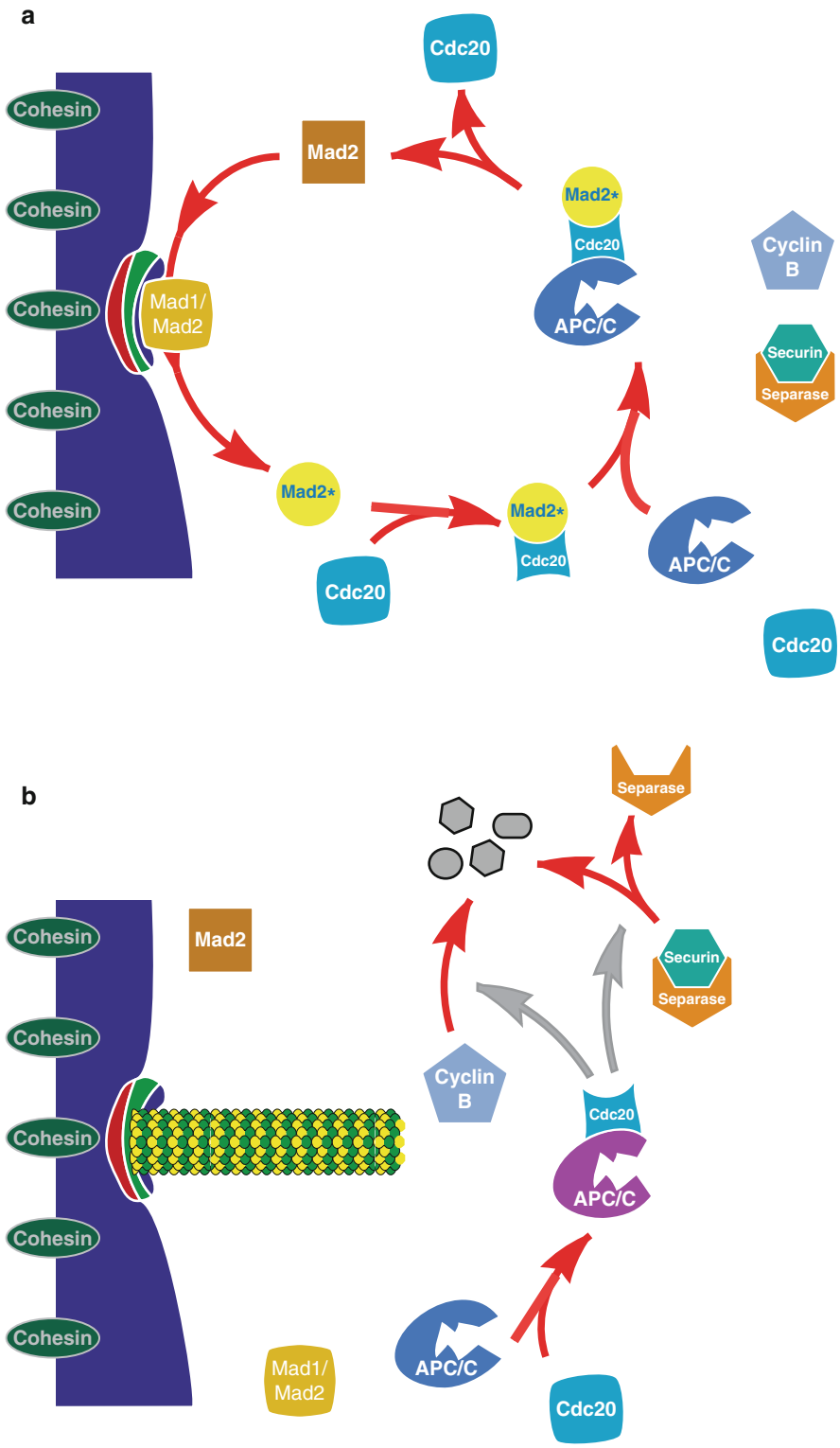
Definition

When a cell divides, each of the daughter cells receives the same number of chromosomes from the mother. When the process of chromosome segregation, or anaphase, goes awry daughter cells receive an unequal number of chromosomes, a condition known as aneuploidy. Although unbalanced gene dosage is not necessarily lethal, aneuploidy can seriously affect the viability of cells. Moreover, it is a typical trait in cancer cells where aneuploidy is thought to contribute to the progression to malignancy. Not surprisingly, cells have developed a signaling pathway that ensures sister chromatids are properly segregated to daughter cells. The pathway, known as the Spindle Assembly Checkpoint (SAC), delays sister chromatid separation until all kinetochores are attached to microtubules coming from the opposite spindle poles (a condition known as metaphase). When all kinetochores are attached in a bipolar manner, the checkpoint-mediated inhibition is lifted and cells undergo the transition from metaphase to anaphase.

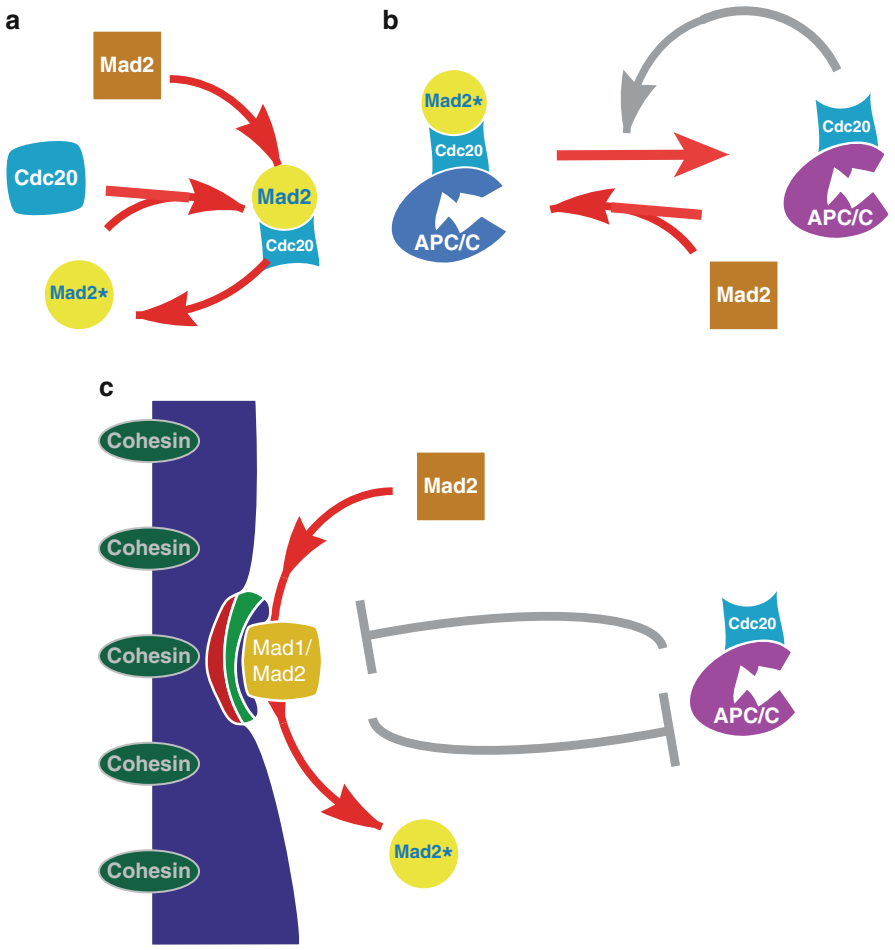
While the SAC pathway is not essential in budding yeast, in mammals the checkpoint is activated during every cell cycle, ensuring that mitosis lasts long enough to allow proper microtubule/kinetochore attachment. However, it can be artificially induced by drugs that either depolymerize (e.g., nocodazole) or stabilize (e.g., Taxol) microtubules of the mitotic spindle. Genes coding for proteins of this pathway were originally identified in budding yeast by treating cells with microtubule-depolymerizing drugs (Li and Murray 1991; Hoyt et al. 1991). Components of the SAC were identified when mutant cells did not arrest in response to microtubule depolymerization.

Consistent with the notion that the SAC monitors-microtubule/kinetochore attachment, the "wait anaphase" signal originates at the unattached kinetochores. The target of the pathway is the Anaphase Promoting Complex or Cyclosome (APC/C), an E3 ubiquitin ligase which ubiquitinates, thus tagging for degradation, two fundamental players of the metaphase-to-anaphase transition: Cyclin B and securin. The first is the mitotic cyclin, which in complex with Cyclin Dependent Kinase 1 (CDK1) drives the cell cycle into mitosis. The latter is a stoichiometric inhibitor of separase, a protease that cleaves cohesin, the molecular glue holding duplicated chromatids together, and in so doing allows them to be segregated. The degradation of the two proteins following their ubiquitination is carried out by the proteasome, and is required for the exit from mitosis (Cyclin B degradation) and the segregation of sister chromatids toward the opposite poles (securin degradation).

The SAC inhibits APC/C activity, and it does so indirectly. APC/C recognizes Cyclin B and securin only when it is bound to its co-activator Cdc20. As a result of SAC activation, Cdc20 cannot activate the APC/C, and Cyclin B and securin are stabilized. From a molecular standpoint, the inhibition requires the formation of a multiprotein complex, the Mitotic Checkpoint Complex or MCC, which comprises Cdc20 and several SAC components (Mad2, Bub3, BubR1). Another critical complex is the one formed by Mad1 and Mad2 that localizes at the unattached kinetochore where it catalyzes the formation of the MCC. Finally, several kinases have been implicated in SAC functioning: Aurora B, Mps1, and Cyclin B/CDK1 itself. Thus, the current model of the checkpoint (Fig. 1a) includes a sensor (Mad1:Mad2) which by inducing the structural rearrangement of Mad2 favors the formation of the Cdc20:Mad2 complex (Musacchio and Salmon 2007), the first step to the formation of the MCC. It is widely assumed that the MCC binds to APC/C with higher affinity than Cdc20 alone, and thus as long as MCC is formed it keeps APC/C inhibited in the APC/C:MCC complex. After the last kinetochore has attached, Mad1:Mad2 is displaced from the last kinetochore preventing further MCC production, and energy-dependent reactions drive the dissociation of the MCC and the ensuing activation of APC/C (Fig. 1b). Observations of SAC dynamics implicate the presence of feedback loops, in addition to the reactions-described above, that give the checkpoint interesting systems level properties.



Cell Cycle Signaling, Spindle Assembly Checkpoint, Fig. 1 (**a**) Spindle assembly checkpoint maintenance. Mad1:Mad2 localize at unattached kinetochores (layered structures) where they catalyze the conversion of Mad2 into an active conformation, capable of binding Cdc20. Cdc20:Mad2 binds to the APC/C, inhibiting its activity. As a consequence, Cyclin B and securin are stable, cohesin is not degraded, and cells are arrested in prometaphase. (**b**) When kinetochores attach, to microtubules Mad1:Mad2 are displaced, Mad2 is not converted into a form capable of fast binding to Cdc20 which can now bind and activate APC/C. As a consequence, Cyclin B and securin are degraded, and (not shown) cohesin can be proteolysed by separase driving sister chromatids separation



Cell Cycle Signaling, Spindle Assembly Checkpoint,
Fig. 2 Proposed positive feedback loops in the SAC network. (**a**) The Mad2-template model (De Antoni et al. 2005) proposes that Cdc20:Mad2 catalyzes the formation of more Cdc20:Mad2 complexes similarly to Mad1:Mad2. (**b**) Activated APC/C:Cdc20 has been proposed to induce the formation of more APC/C:Cdc20 complexes, possibly via the ubiquitination of Cdc20. (**c**) A double-negative feedback loop involving the SAC and APC/C:Cdc20 has also been suggested. The double negative behaves like a positive feedback loop and could be important for SAC maintenance and disengagement

Characteristics

Systems Level Properties and Models

The checkpoint is characterized by three key systems level properties: (1) as long as there is one single unattached kinetochore, the checkpoint is active and can arrest cells for several hours (Rieder et al. 1995); (2) as soon as the last kinetochore is attached the checkpoint is disengaged in a matter of few minutes, and cells transit into anaphase (Clute and Pines 1999); and (3) after it is disengaged, the checkpoint can be re-activated, and its key substrates (Cyclin B and securin) stabilized when cells are treated with microtubule stabilizing drugs (Clute and Pines 1999). To explain these properties, several positive feedback loops have been proposed. Although no existing model has been conclusively proven, it is worth going through the suggested loops, addressing first those involved in checkpoint maintenance and then those proposed to govern checkpoint inactivation (Fig. 2).

Checkpoint maintenance: The fact that one unattached kinetochore can fully arrest cell cycle progression for hours (systems level property (1)) naturally raises the question whether this is the outcome of one kinetochore only, or whether it is due to the additional contribution of circuits away from the kinetochore. Given that only a few thousand SAC molecules localize at the unattached kinetochore (Howell et al. 2000), simplified models (Doncic et al. 2005) of the checkpoint came to the conclusion that amplification away from kinetochores is required, a result further reinforced when the fluxes of checkpoint components at the unattached kinetochores are taken into account (Sear and Howard 2006). However, the nature of the amplification process still requires much clarification. Based on structural data, it was

proposed that the MCC itself (or better, one of its constituents, the Mad2:Cdc20 complex) is capable of propagating the formation of new MCC complexes, the so-called Mad2-template model (Fig. 2a and De Antoni et al. 2005). This idea, never confirmed in vivo, was initially criticized on the basis that such a loop could make the checkpoint independent from kinetochores and thus potentially impossible to be switched off, but it should be noticed that the positive feedback loop does not introduce energy in the system, and thus is conceivable there exists regimes of autocatalysis where the non-kinetochore-mediated MCC production is still dependent on the kinetochore-mediated catalysis (Simonetta et al. 2009).

Checkpoint inactivation: Positive feedback loops were also proposed to explain the rapid inactivation of the SAC (systems level property (2)). First, it was suggested that active APC/C:Cdc20 could help the transformation of APC/C:MCC complexes into active APC/C:Cdc20 via the ubiquitination of Cdc20 (Fig. 2b and Reddy et al. 2007). This positive feedback loop could accelerate the release of APC/C inhibition once MCC production is shut off. Later, it was proposed that a double-negative feedback loop comprising APC/C:Cdc20 and the checkpoint could explain the rapid checkpoint inactivation after the last kinetochore has attached (Fig. 2c and He et al. 2011; Zeng et al. 2010). While the SAC inactivating APC/C has been well documented, the mechanism whereby APC/C inactivates the SAC has not been so thoroughly examined. It should be noted that SAC components are substrates of APC/C (e.g., Mps1), and Cyclin B itself is widely recognized as required for SAC activation, and it is of course a substrate of APC/C:Cdc20. Importantly, the presence of feedback loops involved in SAC inactivation needs to be reconciled with its reported reversibility (system level property (3)).

Quantitative Measurements

The circuits depicted above provide only a qualitative description of the SAC pathway. In this section, we report the most relevant experimental measurements that are needed to understand how the SAC might actually work. Given the basic structure of SAC signaling, the most relevant parameters would seem to be the balance of MCC production and MCC:APC/C dissociation. Currently only the former has been measured quantitatively. This work stems from the measurement of the number of SAC molecules, specifically Mad2 molecules, at an unattached kinetochore (~1,500 molecules/unattached kinetochore) (Howell et al. 2000). Subsequent to this work, a number of groups have measured Mad2 turnover at the unattached kinetochore (t1/2 ~ 10–25 s) (Howell et al. 2000; Shah et al. 2004) providing an upper bound to the MCC generation rate (~35–65 molecules/kinetochore/s).

Cell Cycle Signaling, Spindle Assembly Checkpoint, Table 1 Concentrations of molecular species belonging to the SAC pathway in HeLa cells

Species	Value	Reference
APC/C	80 nM	Tang et al. (2001)
	65 nM	Sudakin et al. (2001)
Bub1	100 nM	Howell et al. (2004)
BubR1	90 nM	Tang et al. (2001)
	52 nM	Sudakin et al. (2001)
	127 nM	Fang (2002)
Cdc20	100 nM	Tang et al. (2001)
	285 nM	Fang (2002)
Mad1	20 nM	Shah et al. (2004)
	1/4 of Mad2	Luo et al. (2004)
Mad2	120 nM	Tang et al. (2001)
	400 nM	Sudakin et al. (2001)
	100–300 nM	Luo et al. (2004)
	200 nM	Shah et al. (2004)
	100 nM	Howell et al. (2000)
	230 nM	Fang (2002)

This generation rate must be matched or be greater than the dissociation rate of the MCC:APC/C complex to permit robust inhibition. The dissociation rate, at this time, has only been estimated from measurements of mitotic timing. Observations of live cells have provided detailed measurements of anaphase onset after the establishment of complete bipolar attachment. While the numbers vary, the consensus is that in mammals from last kinetochore attachment anaphase ensues in ~ 25 min, and from the loss of Mad2 at the last unattached kinetochore anaphase begins ~ 10 min later. From these measurements and known concentrations of SAC and APC/C proteins (see Table 1), we can estimate that the dissociation rate of the MCC:APC/C is near 0.0013/s. For ~100,000 APC/C molecules in the cell, this is a molecular dissociation rate of 130 molecules/s. This does not match the MCC production rate from the kinetochore, a fact pointing to the little knowledge we have about the ways cells inactivate the SAC. Although different mechanisms

have been proposed to explain SAC switching off (the stripping Mad1:Mad2 away from unattached kinetochores, the degradation of Cdc20, the phosphorylation of Mad2, the ubiquitination of Cdc20), no consensus has been reached yet, making SAC silencing a major field of investigation.

References

Clute P, Pines J (1999) Temporal and spatial control of cyclin B1 destruction in metaphase. Nat Cell Biol 1:82–87

De Antoni A, Pearson CG, Cimini D, Canman JC, Sala V, Nezi L, Mapelli M, Sironi L, Faretta M, Salmon ED, Musacchio A (2005) The mad1/mad2 complex as a template for mad2 activation in the spindle assembly checkpoint. Curr Biol 15:214–225

Doncic A, Ben-Jacob E, Barkai N (2005) Evaluating putative mechanisms of the mitotic spindle checkpoint. Proc Natl Acad Sci USA 102(18):6332–6337

Fang G (2002) Checkpoint protein BubR1 acts synergistically with Mad2 to inhibit anaphase-promoting complex. Mol Biol Cell 13(3):755

He E, Kapuy O, Oliveria RA, Uhlmann F, Tyson JJ, Novák B (2011) System-level feedbacks make the anaphase switch irreversible. Proc Natl Acad Sci USA 108(24):10016–10021

Howell BJ, Hoffman DB, Fang G, Murray AW, Salmon ED (2000) Visualization of Mad2 dynamics at kinetochores, along spindle fibers, and at spindle poles in living cells. J Cell Biol 150:1233–1250

Howell BJ et al. (2000) Visualization of Mad2 dynamics at kinetochores, along spindle fibers, and at spindle poles in living cells. J Cell Biol 150(6):1233

Howell BJ et al. (2004) Spindle checkpoint protein dynamics at kinetochores in living cells. Curr Biol 14(11):953

Hoyt MA, Totis L, Roberts BTS (1991) Cerevisiae genes required for cell cycle arrest in response to loss of microtubule function. Cell 66:507–517

Li R, Murray A (1991) Feedback control of mitosis in budding yeast. Cell 66:519–531

Luo X et al. (2004) The Mad2 spindle checkpoint protein has two distinct natively folded states. Nat Struct Mol Biol 11(4):338

Musacchio A, Salmon ED (2007) The spindle-assembly checkpoint in space and time. Nat Rev Mol Cell Biol 8:379–393

Reddy SK, Rape M, Margansky WA, Kirschner MW (2007) Ubiquitination by the anaphase-promoting complex drives spindle checkpoint inactivation. Nature 446:921–925

Rieder CL, Cole RW, Khodjakov A, Sluder G (1995) The checkpoint delaying anaphase in response to chromosome monoorientation is mediated by an inhibitory signal produced by unattached kinetochores. J Cell Biol 130:941–948

Sear RP, Howard M (2006) Modeling dual pathways for the me-tazoan spindle assembly checkpoint. Proc Natl Acad Sci USA 103:16758–16763

Shah JV, Botvinick E, Bonday Z, Furnari F, Berns M, Cleveland DW (2004) Dynamics of centromere and kinetochore proteins; implications for checkpoint signaling and silencing. Curr Biol 14:942–952

Shah JV et al. (2004) Dynamics of centromere and kinetochore proteins; implications for checkpoint signaling and silencing. Curr Biol 14(11):942

Simonetta M, Manzoni R, Mosca R, Mapelli M, Massimiliano L, Vink M, Novak B, Musacchio A, Ciliberto A (2009) The influence of catalysis on mad2 activation dynamics. PLoS Biol 7(1):e10

Sudakin V, Chan GK, Yen TJ (2001) Checkpoint inhibition of the APC/C in HeLa cells is mediated by a complex of BUBR1, BUB3, CDC20, and MAD2. J Cell Biol 154(5):925

Tang Z, Bharadwaj R, Li B, Yu H (2001) Mad2-Independent inhibition of APCCdc20 by the mitotic checkpoint protein BubR1. Dev Cell 1(2):227

Zeng X, Sigoillot F, Gaur S, Choi S, Pfaff KL, Oh DC, Hathaway N, Dimova N, Cuny GD, King RW (2010) Pharmacologic inhibition of the anaphase-promoting complex induces a spindle checkpoint-dependent mitotic arrest in the absence of spindle damage. Cancer Cell 18:382–395

Cell Cycle Transition, Detailed Regulation of Restriction Point

Laurence Calzone
Institut Curie – INSERM U900 – Mines ParisTech, Paris, France

Synonyms

G1 checkpoint; R-point

Definition

The restriction point is an irreversible transition occurring in late G1 phase after which the cell commits to DNA replication and cell cycle progression. Using a systems biology approach, detailed molecular maps are built and reveal a highly complex architecture governing the regulation of this transition.

Characteristics

Basic Concepts of the Restriction Point

Progression and coordination of the cell cycle are governed by complex molecular processes (▶ Cell Cycle). Throughout the cycle, different safety mechanisms verify that proper conditions are met for cell cycle events such as DNA replication, chromosome

alignment, etc. (▶ Cell Cycle Checkpoints). The restriction point (▶ Cell Cycle Transition, Principles of Restriction Point) is one of these checkpoints. It blocks entry into S phase upon growth-limiting constraints.

In the absence of mitogens, the cell enters a G0 (or quiescent) state. In the presence of mitogens, growth is stimulated and the cell advances through G1 phase. During the time preceding the restriction point transition, the cell is responsive to both positive (e.g., mitogenic growth factors) and negative (e.g., transforming growth factor β) external factors. However, when it passes through the restriction point in late G1, the cell commits to cell cycle events and proceeds to DNA synthesis. After that point, even if the mitogens are removed, the cell completes its cycle before stopping at the following round (▶ Cell Cycle Transition, Principles of Restriction Point).

What is the molecular machinery behind the cell decision to either halt and enter a quiescent state or replicate its DNA and progress through M phase?

Biochemical Interactions Involved in the Restriction Point Transition

To recapitulate all the key players and understand the intricate processes governing this decision, a thorough description of the molecular machinery controlling the restriction point is proposed.

Processing the Information

A specific family of kinases has first been identified as cell growth sensors (▶ Cyclins and Cyclin-dependent Kinases). They link external stimuli to the intracellular machinery. These sensors are composed of a kinase subunit, Cdk, and a cyclin partner, which determines the localization and the function of the complex. If these complexes can sense growth signals, how exactly is the signal transmitted to them?

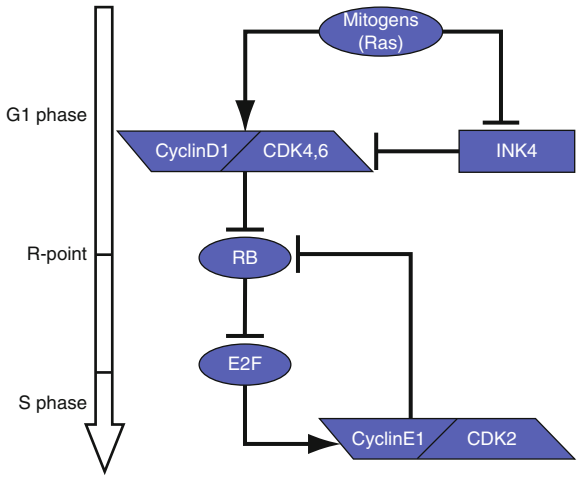
The sequence of events leading to the activation of these kinases can be characterized as follows: The binding of growth factors to their receptors lead to the activation of cytoplasmic proteins such as Ras, which passes on the signal to a cascade that ultimately turns on the transcription of cell cycle genes (▶ Transcription). The growth factor-mediated activation of these genes operates through two distinct paths. On one hand, it leads to the accumulation of the prototypical G1 cyclin, CyclinD1, in the nucleus and promotes its association with the cyclin-dependent kinases CDK4 or CDK6 to trigger the G1/S transition. On the other hand, it suppresses the activity of ▶ Cdk inhibitors (or CKI, such as p16INK4a or p15INK4b), which prohibit the association of the CyclinD1/CDK4,6 complexes. However, the transition is not governed by the sole activation of these complexes. As long as proper intracellular conditions are not met, a checkpoint prevents entry into S phase until the cell is ready.

RB, the Guardian of the Restriction Point Gate

The tumor suppressor gene *retinoblastoma* (RB) plays the role of gatekeeper. It is the main target of CyclinD1/CDK4,6 complexes. During the G1 phase, RB sequesters a family of key ▶ transcription factors, the E2Fs, that mediates the transcription of many genes involved in cell cycle events and in the apoptotic pathway (DeGregori and Johnson 2006). Upon mitotic stimulation, CyclinD1/CDK4,6 complexes start phosphorylating RB, which releases and activates E2Fs. E2Fs activation terminates G1 phase and leads to entry into S phase. CyclinE1, one of E2F target genes, binds to CDK2 and collaborates with CyclinD1/CDK4,6 to maintain the hyperphosphorylated state of RB, thereby ensuring the irreversibility of the transition. The processes described above are illustrated in Fig. 1.

The study of RB in the context of the restriction point is critical for several reasons. First, RB inactivation and the passage through the restriction point occur concomitantly (Bartek et al. 1996). Furthermore, RB is an anti-oncogene and its activity is deregulated in virtually all types of cancers (▶ Cell Cycle, Cancer Cell Cycle and Oncogene Addiction) (Weinberg 2006), including familial and sporadic forms (osteosarcomas, breast carcinomas, small cell lung carcinomas, bladder carcinomas, melanomas, etc.).

The deficiency of the RB pathway is observed in cancers when (1) the external stimulation is constant, for example, when the growth factors are always present or when their receptors are mutated and as a result the signal is always considered to be on, (2) the cell is insensitive to antimitogenic factors (e.g., TGFβ), or (3) when processes that control the passage through the restriction point are perturbed by mutations or diverse dysfunctions (Sherr 1996). More precisely, in cancer samples, dysfunctions of RB itself have been found to arise from the deletion of the gene, from epigenetic mutations, from the presence of viral



Cell Cycle Transition, Detailed Regulation of Restriction Point, Fig. 1 Simple representation of the interactions involved in the restriction point transition

oncoproteins, or as a result of a deregulation of the kinases – and the kinase inhibitors – that control its activity.

Comprehensive Maps of the Restriction Point Transition

The real events behind the restriction point transition are, of course, much more complex than the one depicted in Fig. 1. In this representation, a precise and explicit molecular characterization of the protein interactions and transformations is missing. A simple inhibition in Fig. 1 can be interpreted as an inhibitory sequestration, a degradation process, or a (de)phosphorylation. Such network representation may lead to ambiguous conclusions.

The complexity increases exponentially as more details are included in the description. For instance, let us consider only the interactions between RB and E2F: RB, along with its family members, binds to some but not all of the eight E2F genes (E2F1-8), some of them being activators and others inhibitors of cell cycle advance.

Complex Interactions Between RB, E2F and Their Family Members

- RB protein family
  RB (RB1) is present in quiescent and proliferating cells; RB has 16 sites of phosphorylation.
  p107 (RBL1) is present in proliferating cells; p107 has 10 sites of phosphorylation.
  p130 (RBL2) is present in quiescent cells; p130 has 22 sites of phosphorylation.
- E2F transcription factors
  Activators of transcription: E2F1, E2F2, E2F3a.
  Inhibitors of transcription: E2F3b, E2F4, E2F5, E2F6, E2F7a, E2F7b, E2F8.
  The E2F transcription factors are active as dimers.
  E2F1 to E2F6 dimerize with DP1 or DP2.
  E2F7 and E2F8 are active as homodimers.
- Association between RB family members and the E2Fs
  E2F1, E2F2 and E2F3 bind RB.
  E2F4 binds RB, p107 or p130.
  E2F5 binds p130 (and to some extent p107).
  E2F6 binds proteins from the polycomb group.
  E2F7 and E2F8 form homodimers and do not bind to other known proteins.

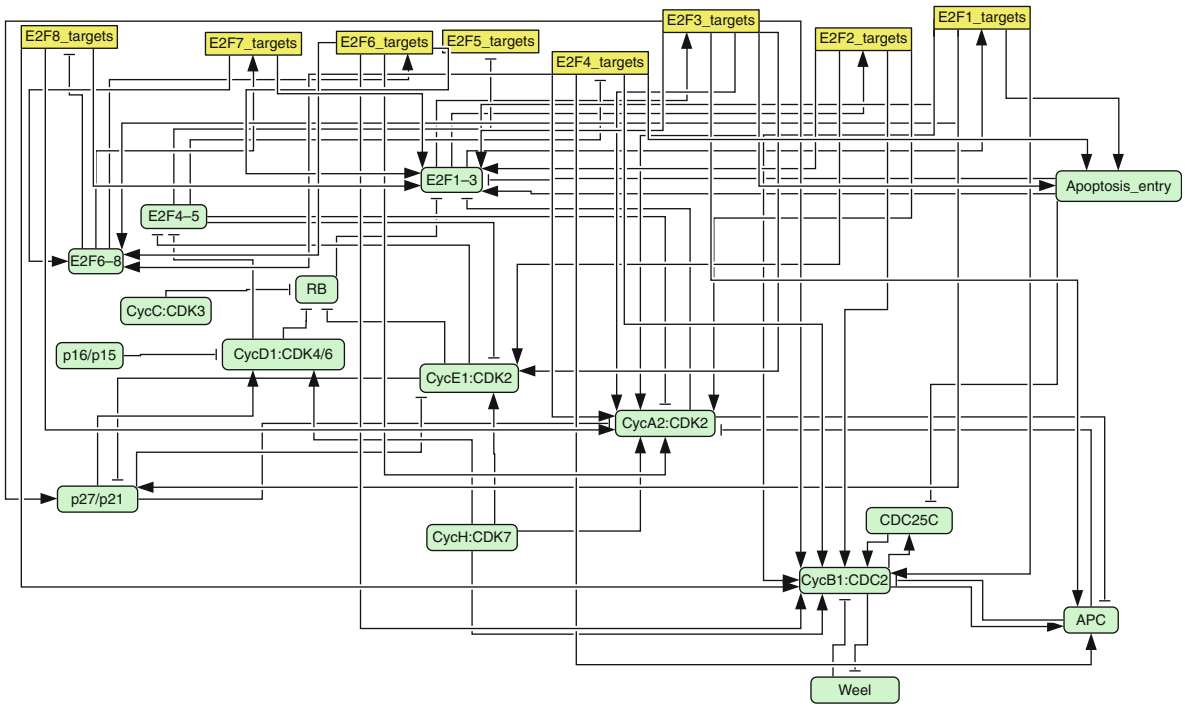
The Comprehensive Map

Network representations can be used to visualize what is known about a biological process. It offers the possibility to integrate heterogeneous data from sparse sources of information from both individual and high throughput studies and summarize them into a synthetic picture. It also provides a tool to replace a mechanism or a gene in its context, identify contradictions from the literature, and anticipate the effects of a local perturbation on the global system.

Networks illustrating any biological mechanisms need to be (1) biologically accurate, coherent, and unambiguous, (2) understandable and human-readable, (3) easily extensible when new information is added, (4) computable, i.e., ready for modeling, and machine-readable, (5) accessible in standard format (SBGN, Le Novère et al. 2009) and (6) hierarchical, with different levels of description (Kitano et al. 2005).

There exist different kinds of representations in numerous databases recapitulating heterogeneous types of information and for which the nodes and edges have different meanings (Pathway Interaction Database from Nature: http://pid.nci.nih.gov/; Biocarta: http://www.biocarta.com/; KEGG: http://www.genome.jp/kegg/; Reactome: http://www.reactome.org, etc.).

Considerable efforts have been made to produce detailed and accurate maps of the mammalian cell cycle molecular machinery (Kohn 1999 and in various databases cited above) and more specifically that of the



Cell Cycle Transition, Detailed Regulation of Restriction Point, Fig. 2 Detailed map of the *retinoblastoma* (RB) network emphasizing the passage through the restriction point. The map is constructed using standard SBGN language in CellDesigner (http://www.celldesigner.org). It comprises 80 proteins, 176 genes, 165 interactions, and established from circa 350 publications. *Upper panel*: biochemical reactions regulating the modifications of the major proteins involved in RB regulation. Blue rectangles refer to tumor suppressor genes and red rectangles to proto-oncogenes. *Lower panel*: mRNA and gene targets of the eight E2F transcription factors. A readable and clickable version of the map can be found at http://bioinfo-out.curie.fr/projects/rbpathway/interactive/rb_network.html

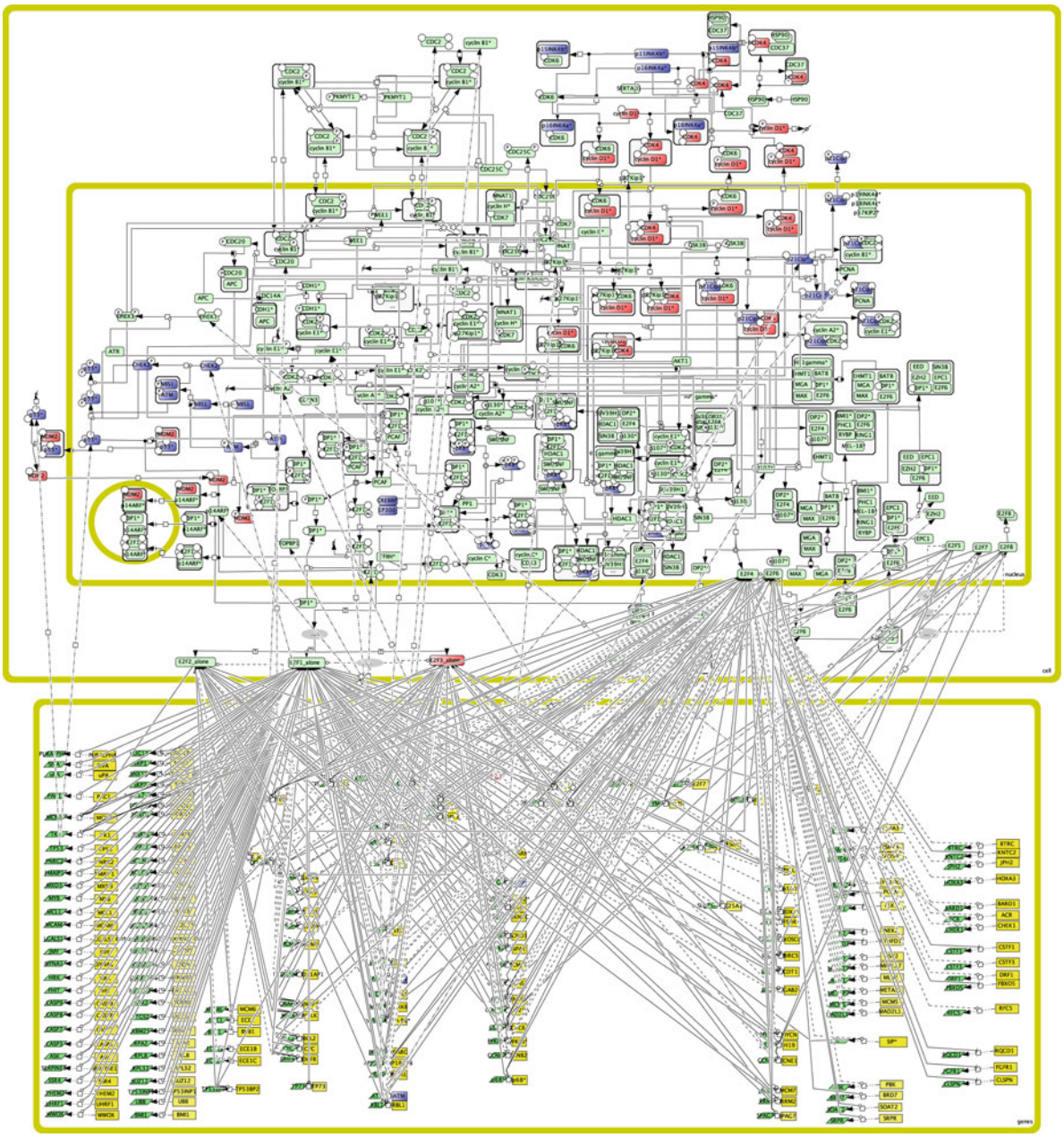
restriction point (Calzone et al. 2008). In the latter map, the focus is put on the biochemical reactions and transformations (edges) of species (nodes) that play an important role in the regulatory process (Fig. 2).

Use of the Map

Beyond the role of integrating knowledge in an unambiguous form, these molecular maps can be used to interpret biological data or reason on unexpected behaviors of the system in specific conditions. The map can be viewed as an electronic circuit and interrogated based on its sole topology. It can be compared to a map of the New York subway which allows to find, for example, the simplest way from one station to another or suggest an alternative route if one station is closed. All the important partners or catalyzers for a particular activation can be identified. All or one path(s) between two species can be isolated and the necessary partners or intermediaries of a transformation distinguished. Some links can be broken and alternative – or back-up – pathways, etc., proposed.

Modular View of the Comprehensive Map

To improve the readability of the map, one can simplify it without losing information about regulations. For that purpose, from the initial comprehensive map, groups of components can be identified from data ▶ clustering, when the data is available, or from graph theory techniques (▶ Module Network). Here, the reaction network of Fig. 2 is transformed into an influence network (Fig. 3) in which each node corresponds to a set of connected species in the comprehensive map. The sets of proteins are grouped into *modules*. The nodes are connected either by positive or negative arrows based on the available biological information derived from the complete map. The resulting graph is an abstraction of the initial complex and detailed map.



Cell Cycle Transition, Detailed Regulation of Restriction Point, Fig. 3 Modular view of RB pathway. The nodes of this graph represent modules – or groups of interacting proteins – derived from the comprehensive map (Fig. 2). The nodes are connected by influence arrows, either positive or negative, deduced from the initial network and based on biological knowledge of the nature of the interactions

The different versions of the RB network proposed here (Figs. 1–3) can then be translated into mathematical formulas. Ordinary differential equations are often used for a reaction-based network (▶ Cell Cycle Modeling, Differential Equation), whereas discrete modeling can be applied to influence network (▶ Cell Cycle Modeling Using Logical Rules), etc.

More generally, the construction of the map of a particular process and the gathering of biological knowledge from heterogeneous sources is the first

step toward the elaboration of a mathematical model (Novák and Tyson 2004; Alfieri et al. 2009). Mathematical modeling provides a formal tool to answer specific biological questions. More importantly, it allows to test and predict, *in silico*, different perturbations of a biological process that would be too complex to understand by simple intuitive reasoning.

Cross-References

- ▶ CDK Inhibitors
- ▶ Cell Cycle
- ▶ Cell Cycle Checkpoints
- ▶ Cell Cycle Modeling Using Logical Rules
- ▶ Cell Cycle Modeling, Differential Equation
- ▶ Cell Cycle Transition, Principles of Restriction Point
- ▶ Cell Cycle, Cancer Cell Cycle and Oncogene Addiction
- ▶ Clustering
- ▶ Cyclins and Cyclin-dependent Kinases
- ▶ Module Network
- ▶ Pathway
- ▶ Transcription
- ▶ Transcription Factor

References

Alfieri R, Barberis M, Chiaradonna F, Gaglio D, Milanesi L, Vanoni M, Klipp E, Alberghina L (2009) Towards a systems biology approach to mammalian cell cycle: modeling the entrance into S phase of quiescent fibroblasts after serum stimulation. BMC Bioinformatics 10(Suppl 12):S16

Bartek J, Bartkova J, Lukas J (1996) The retinoblastoma protein pathway and the restriction point. Curr Opin Cell Biol 8(6):805–814

Calzone L, Gelay A, Zinovyev A, Radvanyi F, Barillot E (2008) A comprehensive modular map of molecular interactions in RB/E2F pathway. Mol Syst Biol 4:173

DeGregori J, Johnson DG (2006) Distinct and overlapping roles for E2F family members in transcription, proliferation and apoptosis. Curr Mol Med 6(7):739–748

Kitano H, Funahashi A, Matsuoka Y, Oda K (2005) Using process diagrams for the graphical representation of biological networks. Nat Biotechnol 23(8):961–966

Kohn K (1999) Molecular interaction map of the mammalian cell cycle control and DNA repair systems. Mol Biol Cell 10(8):2703–2734

Le Novère N, Hucka M, Mi H, Moodie S, Schreiber F, Sorokin A, Demir E, Wegner K, Aladjem MI, Wimalaratne SM, Bergman FT, Gauges R, Ghazal P, Kawaji H, Li L, Matsuoka Y, Villéger A, Boyd SE, Calzone L, Courtot M, Dogrusoz U, Freeman TC, Funahashi A, Ghosh S, Jouraku A, Kim S, Kolpakov F, Luna A, Sahle S, Schmidt E, Watterson S, Wu G, Goryanin I, Kell DB, Sander C, Sauro H, Snoep JL, Kohn K, Kitano H (2009) The systems biology graphical notation. Nat Biotechnol 27(8):735–741

Novák B, Tyson JJ (2004) A model for restriction point control of the mammalian cell cycle. J Theor Biol 230(4):563–579

Sherr CJ (1996) Cancer cell cycles. Science 274(5293): 1672–1677

Weinberg R (ed) (2006) The biology of cancer. Garland Science, New York, pp 255–305

Cell Cycle Transition, Principles of Restriction Point

Tae Lee[1], Guang Yao[2] and Lingchong You[3]
[1]School of Biology and School of Physics, Georgia Institute of Technology, Atlanta, GA, USA
[2]Department of Molecular and Cellular Biology, University of Arizona, Tucson, AZ, USA
[3]Biomedical Engineering Department, Duke University, Durham, NC, USA

Synonyms

G1 phase checkpoint; Start

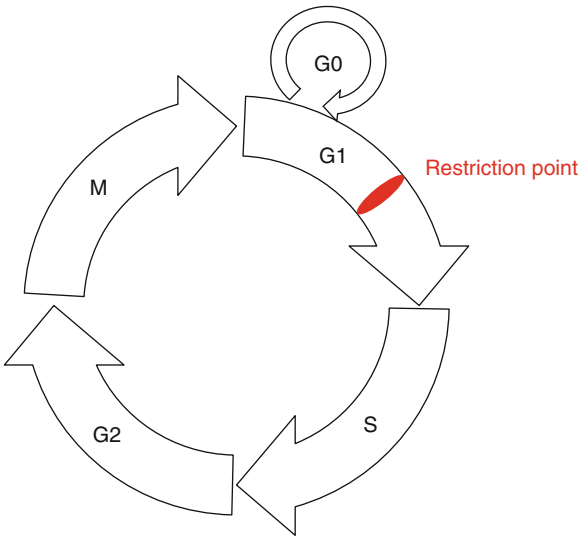
Definition

The restriction point (R-point), similar to the "Start" point in yeast (▶ Cell Cycle, Budding Yeast), is a G1 phase checkpoint (▶ Checkpoints) at which mammalian cells commit to proliferation (▶ Mitosis; ▶ Cell Cycle of Mammalian Cells) and become independent of growth signals for the completion of the cell cycle (Fig.1).

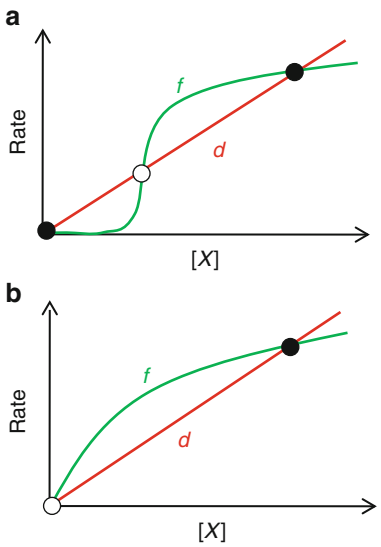
Characteristics

Discovery of the R-point

The key concept of the R-point is the irreversible commitment of a cell to proliferation (▶ Cell Cycle Dynamics, Irreversibility). The R-point was first proposed by Pardee in 1974 (Pardee 1974), when he observed a single, unique arrest point in the G1 phase



Cell Cycle Transition, Principles of Restriction Point, Fig. 1 The cell cycle (Cell cycle, phases of the) and the R-point



Cell Cycle Transition, Principles of Restriction Point, Fig. 2 Rates of synthesis (f(X)) and degradation (d(X)) determine the dynamics of the system described in Equation 1. Suppose d(X) is always first order in terms of X and is sufficiently small. (**a**) If f has ultrasensitive dependence on [X], the system is bistable with two stable steady states (*filled circles*) separated by an unstable one (*open circle*). (**b**) Otherwise, the system is monostable with one unstable steady state and a stable one. If d(X) is too large, the system is monostable for both cases, with a single trivial steady state ([X] = 0)

of the cell cycle associated with different starvation conditions. That is, cells arrested at quiescence by different starvation regimens resumed DNA synthesis at apparently the same time point. With sequential applications of different starvation regimens, cells did not escape from any previous arrest condition (indicating overlapping arrest points). Pardee and colleagues further demonstrated that the R-point is located several hours before the S-phase. Passing this critical time point, cells commit to proliferation and continue to transit the cell cycle even when the exogenous growth condition was removed (Yen and Pardee 1978). The transition between proliferative and quiescent states was further analyzed by Zetterberg and colleague (Zetterberg and Larsson 1985), who measured the timing of the irreversible commitment to proliferation in individual cells by time-lapse cinematograph, and found the R-point occurs at a remarkably constant time after mitosis in exponentially growing cells.

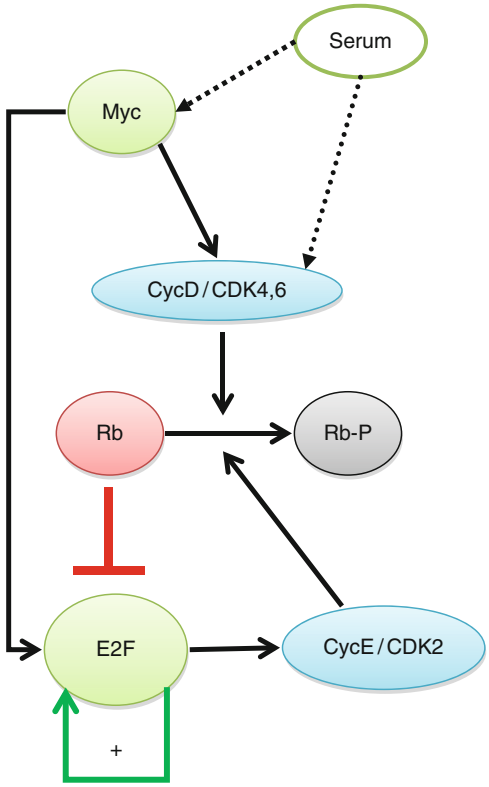
R-point Governed by a Bistable Switch

The R-point has two main characteristics: first, an initial high threshold on growth-stimulating signals, which ensures a tight control of cell growth; second, a low-maintenance mechanism, which ensures that once committed, cell proliferation will be completed even in the absence of sustained growth signals. Theoretical studies have suggested potential mechanisms underlying the R-point control of cell cycle entry (Thron 1997; Aguda and Tang 1999; Novak and Tyson 2004). While these studies differed in molecular details, their consensus is that the R-point is regulated by a bistable switch (▶ Bistability), which can generate all-or-none and hysteretic (history-dependent) response.

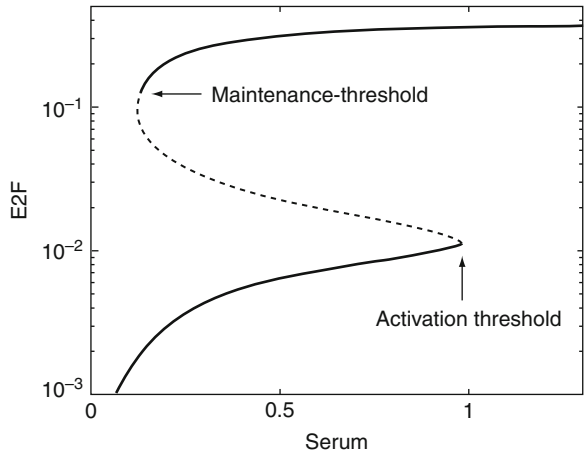
A bistable switch can result from one or more positive feedback loops with sufficient nonlinearity (Ferrell 1996). For example, consider that X regulates its own synthesis via a positive feedback loop,

$$\frac{dX}{dt} = f(X) - d(X), \tag{1}$$

where f and d denote the synthesis and degradation rates of X, respectively. When the synthesis of X is ultrasensitive (e.g., due to cooperativity, multistep, zero-order, or inhibitor ultrasensitivity (Ferrell 1996)), the $f(X)$ curve and the $d(X)$ curve can cross at three steady states ($\frac{dx}{dt} = 0$, Fig. 2a). This will generate two stable steady states (and thus, a bistable system) separated by an unstable one. When the system at the



Cell Cycle Transition, Principles of Restriction Point, Fig. 3 A simplified Rb-E2F network. It contains two positive feedback loops: (1) auto-catalysis of E2F, and (2) mutual-inhibition feedback involving E2F, CycE/Cdk2, and Rb



Cell Cycle Transition, Principles of Restriction Point, Fig. 4 Hysteresis in E2F activation. E2F follows two different trajectories when switching from OFF to ON and ON to OFF. The "activation threshold" is where E2F switches from OFF to ON, and the "maintenance threshold" is where E2F switches from ON to OFF. The region between the two thresholds is a "bistable region" where E2F can be either at ON or OFF state, depending on the history of the system

unstable steady state is subjected to minor perturbations, the system will move away from the unstable steady state towards the stable ones. Without ultrasensitivity in the synthesis of X, the system is always monostable. For example, it can have an unstable steady state and a stable one (Fig. 2b).

Molecular Interactions at the G1/S Transition: The Rb-E2F Switch

The molecular events leading to the traverse of the R-point involve a sophisticated network of molecules and interactions (defined as the Rb-E2F network, ▶ Cell Cycle Transition, Detailed Regulation of Restriction Point) (Calzone et al. 2008). Here we present a highly simplified model that focuses on the cell cycle entry events (Yao et al. 2008), as shown in Fig. 3.

In this simplified model, the node Rb represents the entire pocket protein family (RB, p107, and p130), and the node E2F represents all E2F activators (E2F1, E2F2, and E2F3a). E2F is a transcriptional factor that can activate a large battery of genes encoding activities involved in DNA replication and cell cycle progression. In quiescent cells, E2F protein is bound and repressed by tumor suppressor Rb, and transcription from the E2F promoter is also inhibited by Rb. Upon sufficient growth stimulation, Myc and cyclin D (CycD) are induced. CycD associates with and activates Cyclin dependent kinases (Cdk4,6), which phosphorylate Rb and release E2F. Myc directly binds to the E2F promoter, facilitating E2F binding to nearby sites. E2F then activates its own transcription, forming a positive feedback loop. In addition, E2F also activates Cyclin E, which forms a complex with Cdk2 to further phosphorylate Rb and relieve its repression of E2F, constituting another positive feedback loop. Theoretical studies suggested that the positive feedback mediated by E2F and the ultrasensitivity created by the Rb repression of E2F could potentially generate a bistable switch (Fig. 4, ▶ Cell Cycle Model Analysis, Bifurcation Theory) (Thron 1997).

Recent experimental analysis has demonstrated that the Rb-E2F network indeed functions as a bistable switch and generates bimodal and history-dependent E2F activities (Yao et al. 2008). Mammalian cells in

quiescence remain in the OFF state of E2F when deprived of growth stimulation. Following sufficient growth stimulation, these cells eventually activate E2F and commit to proliferation. Afterward, E2F stays at the ON state to drive the completion of the cell cycle, even after the growth stimuli are removed. For a given serum concentration, the transition timing has been found to be highly variable, which can be described with a stochastic model (Lee et al. 2010). A similar bistable switch mechanism was identified in yeast for the control of the Start point (Charvin et al. 2010).

Cross-References

- ▶ Bistability
- ▶ Cell Cycle Checkpoints
- ▶ Cell Cycle Dynamics, Irreversibility
- ▶ Cell Cycle Model Analysis, Bifurcation Theory
- ▶ Cell Cycle of Mammalian Cells
- ▶ Cell Cycle Transition, Detailed Regulation of Restriction Point
- ▶ Cell Cycle, Budding Yeast
- ▶ Mitosis

References

Aguda BD, Tang Y (1999) The kinetic origins of the restriction point in the mammalian cell cycle. Cell Prolif 32:321–335

Calzone L, Gelay A, Zinovyev A, Radvanyi F, Barillot E (2008) A comprehensive modular map of molecular interactions in RB/E2F pathway. Mol Syst Biol 4:173

Charvin G, Oikonomou C, Siggia ED, Cross FR (2010) Origin of irreversibility of cell cycle start in budding yeast. PLoS Biol 8(1):e1000284

Ferrell JE (1996) Tripping the switch fantastic: how a protein kinase cascade can convert graded inputs into switch-like outputs. Trends Biochem Sci 21:460–466

Lee T, Yao G, Bennett D, Nevins J, You L (2010) Stochastic E2F activation and reconciliation of phenomenological cell-cycle models. PLoS Biol 8:9

Novak M, Tyson JJ (2004) A model for restriction point control of the mammalian cell cycle. J Theor Biol 230:563–579

Pardee AB (1974) Restriction point for control of normal animal-cell proliferation. Proc Natl Acad Sci USA 71:1286–1290

Thron CD (1997) Bistable biochemical switching and the control of the events of the cell cycle. Oncogene 15:317–325

Yao G, Lee TJ, Mori S, Nevins JR, You LC (2008) A bistable Rb-E2F switch underlies the restriction point. Nat Cell Biol 10:476–U255

Yen A, Pardee AB (1978) Exponential 3t3-cells escape in mid-G1 from their high serum requirement. Exp Cell Res 116:103–113

Zetterberg A, Larsson O (1985) Kinetic-analysis of regulatory events in G1 leading to proliferation or quiescence of swiss 3t3 cells. Proc Natl Acad Sci USA 82:5365–5369

Cell Cycle Transitions, G2/M

Judit Zámborszky
The Microsoft Research – University of Trento Centre for Computational and Systems Biology, CoSBi, Trento, Italy

Synonyms

Entry into mitosis; G2/M transition; Mitotic entry

Definition

G2/M transition (or mitotic entry) is the progression from G2-phase to M-phase (also referred to as "mitosis") of the mitotic cell division cycle.

Characteristics

Eukaryotic cells govern a repetitive series of events that result in self-reproduction (Morgan 2006). The cell cycle (▶ Cell Cycle) consists of four different stages: G1-, S-, G2-, and M-phase by which a growing cell replicates all its components and divides into two nearly identical daughter cells (▶ Cell Cycle, Physiology). The alteration between cell cycle phases is sophisticatedly controlled, ensuring appropriate timing and order of sequential action (G1-phase, G1/S transition, DNA replication, G2-phase, G2/M transition, mitosis, and mitotic exit (▶ Cell Cycle Transitions, Mitotic Exit)). Transitions are tightly monitored by "surveillance mechanisms" (checkpoints (▶ Cell Cycle Checkpoints)) (Kastan and Bartek 2004). Checkpoints are control mechanisms governing crucial irreversible transitions of the cell division cycle (▶ Cell Cycle Dynamics, Irreversibility). Monitoring the environmental conditions (nutrient, hormones, etc.) and the stage of the DNA (if there are lesions or

chromosome aberrations) or check of the size of the cell (if it is large enough for the process of the cycle) are important regulations for the maintenance of genomic stability. Chromosome breaks or rearrangements are known to decrease survival of an individual cell, thus proper checkpoint control is crucial for cancer prevention in multicellular organisms (Kastan and Bartek 2004) (▶ Cell Cycle, Cancer Cell Cycle and Oncogene Addiction).

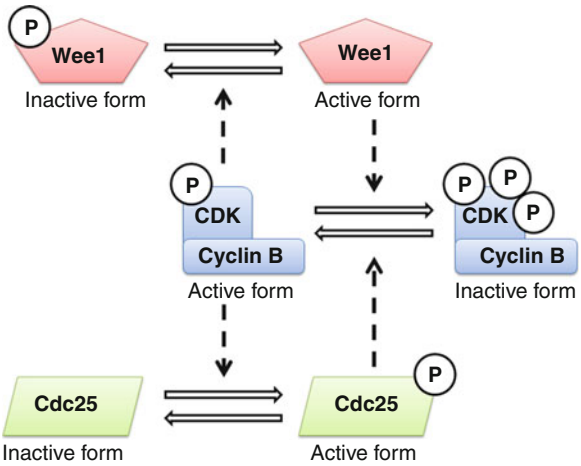
The cell cycle transitions are controlled through cell cycle proteins' activities (▶ Cyclins and Cyclin-dependent Kinases). Cyclin-dependent kinases (CDKs) are responsible for most basic cell cycle processes and their role is crucial in cell cycle progression (Morgan 1995). CDKs' activity depends on (1) their regulatory cyclin subunits (those are specific to the different cell cycle phases), (2) their phosphorylation state, and (3) the presence of proteins forming inhibitory complexes with CDKs.

The G2/M transition is triggered by the increased activity of CDK in combination with B-type cyclins; and the active CDK/Cyclin B complex has several substrates that are responsible for mitotic events (▶ Mitosis) afterward. CDK activity in G2-phase depends on phosphorylation or dephosphorylation reactions (▶ Post-translational Modifications) of specific tyrosine and threonine residues.

First of all, phosphorylation of CDK on a conserved threonine residue (T161 in mammals) by cdk-activating kinase (CAK) is crucial for its kinase activity (Kaldis et al. 1996) (Activator).

Second, phosphorylation of CDK (Cdc2 in mammals) within the active site of the kinase (at tyrosine 15 (Y15) and threonine 14 (T14) in mammals) prevents CDK activity and mitosis (▶ Cdk Inhibitors). This modification is carried out by inhibitory protein kinases of the Wee1-class (Wee1, Mik1, Myt1 in mammals; Swe1 in budding yeast; all referred as "Wee1" in the latter text) (Coleman and Dunphy 1994). Thus, during the G2-phase, the presence of active Wee1 inactivates the CDK/Cyclin B complex and blocks the cell cycle entry into M-phase while cells are growing and preparing for mitosis.

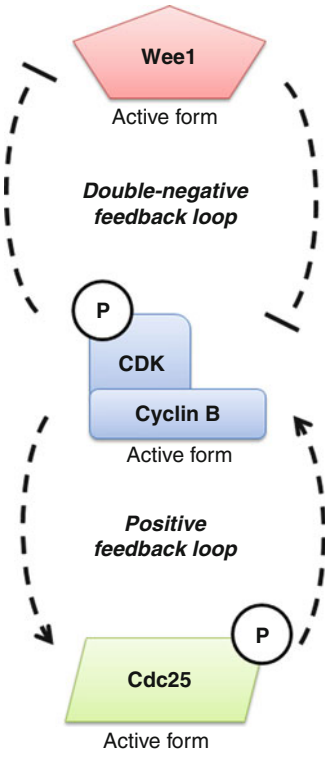
However, when cells are ready to divide, a phosphatase of the Cdc25-class removes the inhibitory phosphate groups from the inactive CDK that results in activation of the complex (Coleman and Dunphy 1994). This modification facilitates the G2/M transition as the active CDK/Cyclin B becomes able to beat the inhibitory effect of Wee1.



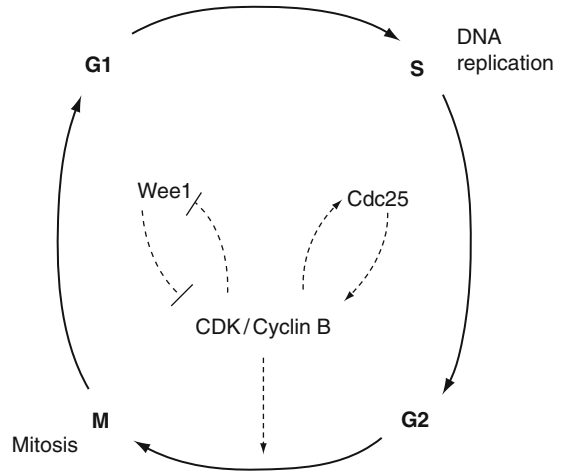
Cell Cycle Transitions, G2/M, Fig. 1 Regulatory mechanism of the G2/M transition. The active forms of the enzymes (kinases and a phosphatase) are able to phosphorylate their target (dashed arrows). The active CDK/Cyclin B phosphorylates (thus inhibits and activates) Wee1 and Cdc25. Active Wee1 (and Cdc25) modify the CDK/Cyclin B complex by adding (and removing) phosphate groups to (from) the active (and inactive) form

Regulation of CDK is more complex: CDK feeds back (▶ Feedback Regulation) by phosphorylating both Wee1 and Cdc25 (Solomon et al.1990, Hoffmann et al. 1993) (Fig. 1). The phosphorylation of Wee1 by CDK inactivates Wee1, creating a mutual inhibitory link between the two kinases. One should think about the relation between CDK and Wee1 as they are enemies inhibiting (phosphorylating) each other, providing a double-negative feedback regulation (also called as antagonism (▶ Negative Feedback)). This molecular network structure has been shown to be able to form a bistable system (▶ Bistability, ▶ Cell Cycle Dynamics, Bistability and Oscillations) with either a state of low CDK and high Wee1 activity (in G2-phase) or a high CDK and low Wee1 activity (in mitosis). The latter stage is appropriate for triggering mitotic entry as the active CDK dimer is required for the control of several crucial molecular mechanisms of M-phase (mitosis (▶ Mitosis)).

The phosphorylation of Cdc25 phosphatase by CDK presents a mutual activation as both enzymes "help" each other: CDK activates its phosphatase Cdc25 which in return removes the inhibitory phosphate groups from CDK. This is remarkably a different



Cell Cycle Transitions, G2/M, Fig. 2 Representation of two interconnected positive feedback loops regulating entry into mitosis during cell cycle. Dashed arrows symbolize activations and dashed lines with an end -| assist for inhibitory reactions



Cell Cycle Transitions, G2/M, Fig. 3 Simplified diagram of the cell cycle and the regulation of the G2/M transition. Dashed arrows represent activations and dashed lines with an end -| assist for inhibitory reactions

type of positive feedback regulation than the CDK-Wee1 link (Fig. 2).

In sum (Fig. 3), during the G2-phase of the cell cycle, CDK activity is kept low with inhibitory phosphorylation events carried out by the Wee1 kinase. As the activity of CDK/Cyclin B increases in time, it activates its helper phosphatase Cdc25 which in return removes the inhibitory phosphate groups from CDK. The sudden increase of CDK activity triggers inactivation of Wee1 kinase by its phosphorylation. The active CDK generates mitotic events (e.g., chromosome condensation, nuclear envelope breakdown, chromatid segregation, assembly of mitotic spindle (▶ Mitosis, ▶ Cell Cycle, Physiology)) and cells enter the M-phase.

The two positive feedback loops (▶ Feedback Regulation, ▶ Positive Feedback) described previously act synergistically and provide an abrupt increase in the activity of CDK making the G2/M transition an irreversible, switch-like event (▶ Cell Cycle Dynamics, Irreversibility, ▶ Cell Cycle Dynamics, Bistability and Oscillations, ▶ Cell Cycle Model Analysis, Bifurcation Theory). Bistability (and hysteretic transitions) of the Wee1-CDK-Cdc25 control system in frog egg cells has been experimentally tested in Jill Sible's lab (Sha et al. 2003) and in Jim Ferrell's lab (Pomerening et al. 2003).

The switches of cell cycle phases are triggered by transient signals that emerge when the checkpoint conditions are satisfied. The irreversibility of these transitions ensures that when the generated signal disappears, cells do not step back into their previous stage resulting in aberration of cells (▶ Cell Cycle Dynamics, Irreversibility). For instance, if cells escape the checkpoint in G2-phase, they might end up with nonidentical daughter cells with extra or less DNAs than the mother cell, rolling over the stability of the genome that causes fatal errors (Kastan and Bartek 2004).

Upon DNA damage, the checkpoint proteins are often recruited to DNA lesions by repair complexes that generate signals to activate the checkpoint response pathway (▶ Cell Cycle Arrest After DNA Damage). Transducers transmit and amplify the signal to downstream targets switching on the DNA-repair apparatus and blocking the cell cycle machinery (Kastan and Bartek 2004). In case of activated

Cell Cycle Transitions, G2/M, Table 1 Different names used for functional homologues of the regulatory elements in the G2/M transition network

Generic name	Function	Mammals	Budding yeast	Fission yeast	Xenopus laevis
CDK	Cyclin-dependent kinase	Cdc2	Cdc28	Cdc2	Cdc2
Cyclin B	Regulatory subunit of CDKs in mitosis	CycB	Clb1,2	Cdc13	CycB
Wee1	CDK/Cyclin B inhibitory kinase	Wee1	Swe1	Wee1, Mik1	Wee1, Myt1
Cdc25	CDK/Cyclin B activatory phosphatase	Cdc25	Mih1	Cdc25	Cdc25
CAK	Cyclin-dependent kinase activating kinase	CAK	Cak1	Csk1	MO15

checkpoint response in G2-phase, Cdc25 activation is downregulated resulting in G2 arrest as CDK is not able to overcome and eliminate its "enemy," the Wee1 kinase. However, when DNA repair is complete, among several actions, Wee1 is degraded which leads to activation of CDK that allows G2/M transition.

There are two important exceptions from the common regulatory way of G2/M transition. During oogenesis, the egg cell grows without dividing as its G2 checkpoint is blocked by Wee1 activation or Cdc25 inhibition. At the same time, the fertilized egg undergoes a series of rapid mitotic cycles without growth due to its removed checkpoints (▶ Cell Cycle of Early Frog Embryos).

Research of the mechanisms underlying G2/M transition is crucial as cells with defective checkpoints have progeny with damaged or lost chromosomes, which can result in loss of viability or abnormal growth. Checkpoint defects cause genetic instability that can result in tumors of multicellular organisms. The basic concept of G2/M transition has been discovered with the active help of theoretical work of systems biology (▶ Cell Cycle Modeling, Differential Equation, ▶ Cell Cycle Model Analysis, Bifurcation Theory). Among several researchers, the contribution of the groups of Béla Novák and John J Tyson always had a great importance to our knowledge (Tyson et al. 2001). With the assistance of biological modeling and the usage of dynamical systems theory in cell cycle research, we became able to understand the underlying logic of interconnected feedback loops of the cell cycle providing crucial phenomena, such as bistability, hysteresis, or irreversible transition. The differences in the network between organisms and the variety of molecular elements playing role in the G2/M transition raise further questions to be answered in the future (Table 1).

Cross-References

- ▶ Activator
- ▶ Bistability
- ▶ Cdk Inhibitors
- ▶ Cell Cycle
- ▶ Cell Cycle Arrest After DNA Damage
- ▶ Cell Cycle Checkpoints
- ▶ Cell Cycle Dynamics, Bistability and Oscillations
- ▶ Cell Cycle Dynamics, Irreversibility
- ▶ Cell Cycle Model Analysis, Bifurcation Theory
- ▶ Cell Cycle of Early Frog Embryos
- ▶ Cell Cycle Transitions, Mitotic Exit
- ▶ Cell Cycle, Cancer Cell Cycle and Oncogene Addiction
- ▶ Cell Cycle, Physiology
- ▶ Cyclins and Cyclin-Dependent Kinases
- ▶ Feedback Regulation
- ▶ Mitosis
- ▶ Negative Feedback
- ▶ Positive Feedback
- ▶ Post-translational Modifications

References

Coleman TR, Dunphy WG (1994) Cdc2 regulatory factors. Curr Opin Cell Biol 6(6):877–882

Hoffmann I, Clarke PR, Marcote MJ, Karsenti E, Draetta G (1993) Phosphorylation and activation of human cdc25-C by cdc2-cyclin B and its involvement in the self-amplification of MPF at mitosis. EMBO J 12:53–63

Kaldis P, Sutton A, Solomon MJ (1996) The Cdk-activating kinase (CAK) from budding yeast. Cell 86(4):553–564

Kastan MB, Bartek J (2004) Cell-cycle checkpoint and cancer. Nature 432(7015):316–323

Morgan DO (1995) Principles of CDK regulation. Nature 374:131–134

Morgan DO (2006) The cell cycle: principles of control. New Science, London

Pomerening JR, Sontag ED, Ferrell JE Jr (2003) Building a cell cycle oscillator: hysteresis and bistability in the activation of Cdc2. Nat Cell Biol 5:346–351
Sha W, Moore J, Chen K, Lassaletta AD, Yi CS, Tyson JJ, Sible JC (2003) Hysteresis drives cell-cycle transitions in Xenopus laevis egg extracts. Proc Natl Acad Sci USA 100:975–980
Solomon MJ, Glotzer M, Lee TH, Philippe M, Kirschner MW (1990) Cyclin activation of p34cdc2. Cell 63:1013–1024
Tyson JJ, Chen K, Novák B (2001) Network dynamics and cell physiology. Nat Rev Mol Cell Biol 2:908–916

Cell Cycle Transitions, Mitotic Exit

P. K. Vinod and Béla Novák
Department of Biochemistry, Oxford Centre for Integrative Systems Biology, University of Oxford, Oxford, UK

Definition

Mitotic exit refers to the completion of mitosis with dephosphorylation of cyclin-dependent kinase (Cdk) substrates triggered in response to cyclins destruction and phosphatase(s) regulation.

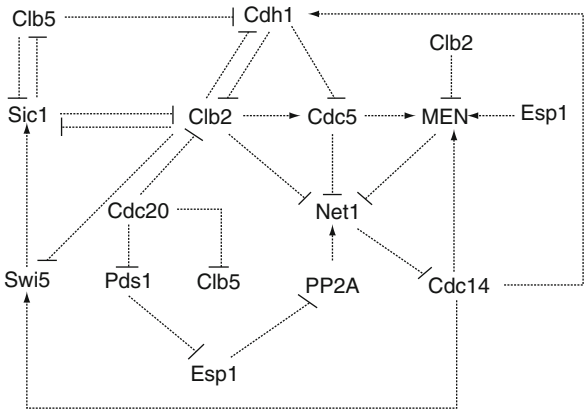
Characteristics

Cell cycle events are tightly regulated such that DNA replication (S phase) is always followed by mitosis (M phase). In eukaryotic cells, such a strict alteration of S and M phases is largely dependent on the periodic activation of cyclin-dependent protein-kinases (Cdks) activated by their specific cyclins partners (Morgan 2007). The underlying molecular machinery comprising of systems-level feedbacks ensures periodic rise and drop in Cdk activities that drive the cell cycle progression.

In eukaryotic cells, mitosis is brought about by intense phosphorylation of cellular proteins initiated by the activation of Cdk1/Cyclin-B complex. During early stages of mitosis (metaphase), replicated chromosomes are still held together by cohesin complexes, and they align (biorient) on the bipolar mitotic spindle with the sister chromatids attaching to the opposite poles. During the following phase of mitosis (anaphase), cohesins holding sister-chromatids together are cleaved by a specific protease (separase) and sister-chromatids are pulled apart by the elongating spindle. Anaphase requires the activation of the Anaphase Promoting Complex (or Cyclosome, APC/C) by its co-activator Cdc20 (also called Fizzy in some organisms) (Morgan 1999). APC/C^{Cdc20} is an ubiquitin-ligase that targets the separase inhibitor, securin and the Cdk1 activator, Cyclin-B (CycB) for proteasome-dependent degradation. Securin degradation relieves the inhibition of separase and triggers anaphase. In addition, CycB degradation inactivates Cdk1 and thereby reduces the rate of protein phosphorylation, which helps in mitotic exit. Since returning to interphase depends on dephosphorylation of the mitotic phospho-proteins, CycB degradation alone is insufficient for mitotic exit which requires the action of a protein-phosphatase (Queralt and Uhlmann 2008). In principle, the phosphatase could have a constant activity and can dephosphorylate mitotic phospho-proteins once its activity overcomes the falling protein-kinase activity. However in budding yeast, where the mitotic exit process is known best, the mitotic exit phosphatase (Cdc14) is regulated and it gets activated at the end of mitosis. In the following, we describe the details of the budding yeast mitotic exit process.

In budding yeast, APC/C^{Cdc20}-mediated degradation of mitotic CycBs (called Clbs) is incomplete and Cdk1 downregulation requires the activation of either another APC/C activator (Cdh1) or a stoichiometric Cdk inhibitor (Sic1). APC/C^{Cdh1} targets CycB to proteasome-dependent degradation similar to APC/C^{Cdc20}, while Sic1 binds to and inhibits the Cdk1/CycB complexes. However, Cdk1/CycB complexes downregulate both Cdh1 and Sic1 thereby creating an antagonistic relationship between Cdk1/CycB and these Cdk1 regulators (Fig. 1). Cdk1/CycB phosphorylated Cdh1 cannot bind to APC/C; therefore, it is inactive while Cdk1/CycB complex inhibits the synthesis and promotes the degradation of Sic1: The Swi5 transcription factor of Sic1 is inactivated by Cdk1/CycB and phosphorylated form of Sic1 gets rapidly ubiquitinated and degraded.

Antagonism between Cdk1/CycB and its negative regulator (Cdh1 and Sic1) forms the underlying principle of cell cycle regulation in budding yeast, which makes the control system bistable, with alternative steady states of high and low Cdk1 activities (Chen et al. 2004). At mitotic exit, APC/C^{Cdh1} is directly



Cell Cycle Transitions, Mitotic Exit, Fig. 1 Interaction network for the mitotic exit in budding yeast. Cdc20 initiates the network by degradation of Clb2, Clb5, and securin (Pds1). This trigger the downregulation of Cdk1 activities and upregulation of a phosphatase, Cdc14 leading to the activation of Cdh1 and Sic1. Cdh1/Sic1 exists in antagonist relationship with Cdk1 and helps in the mitotic exit by further reducing the Cdk1 activity. Cdc14 release from the nucleolus is crucial for the mitotic exit, which is inhibited by Net1 until anaphase. Cdc14 inhibition is relieved through Net1 phosphorylation controlled by Clb2, Cdc5, MEN (mitotic exit network), and separase (Esp1). Esp1 activates Cdc14 release through its non-proteolytic (PP2A downregulation) and proteolytic (MEN activation) functions

activated by Cdc14-mediated dephosphorylation, whereas the level of Sic1 gets upregulated by Cdc14-dependent activation of its transcription factor, Swi5 and turning off Sic1 degradation. In addition, Cdc14 also dephosphorylates other Cdk1 substrates at the end of mitosis and thereby promotes mitotic exit directly as well.

Cdc14 is kept inactive in the nucleolus during most of the cell cycle by its inhibitor, Net1. As cells enter into anaphase, Cdc14 is activated and released from nucleolus by two regulatory networks, Cdc-Fourteen Early Anaphase Release (FEAR) network and Mitotic Exit Network (MEN) (D'Amours and Amon 2004; Stegmeier and Amon 2004). Both of these networks bring about phosphorylation of Net1 which allows Cdc14 release. The FEAR network–mediated Cdc14 release is important for proper anaphase progression but it is dispensable for mitotic exit. On other hand, MEN is required to sustain the released Cdc14 in nucleus and to disperse it into cytoplasm during later stages of mitosis. MEN is essential for mitotic exit, therefore loss of its function arrest cells in telophase with relatively high Cdk1 activity. Phosphorylated Cdh1 and Swi5 are both concentrated in the cytoplasm and hence their dephosphorylation and activation requires MEN-dependent cytoplasmic release of Cdc14.

Cdk1 and polo-kinase (Cdc5) are responsible for Net1 phosphorylation in the FEAR network, and the FEAR also includes separase (Esp1), Slk19, and Spo12 (D'Amours and Amon 2004). Phosphorylated Net1 gets dephosphorylated by $PP2A^{Cdc55}$ and by released Cdc14 itself, the later mechanism creates a negative feedback in the network. Separase (Esp1) promotes Net1 phosphorylation and thereby Cdc14 release by inhibiting $PP2A^{Cdc55}$ but this action does not require the proteolytic function of separase (Queralt et al. 2006). Active $PP2A^{Cdc55}$ keeps Net1 dephosphorylated and Cdc14 inactive until the degradation of securin (Pds1 in budding yeast) by APC/C^{Cdc20} at anaphase, which activates separase (Esp1 in budding yeast). However, FEAR-dependent Cdc14 release is also limited by APC/C^{Cdc20}-dependent CycB degradation, which partially inactivates Cdk1. Since this makes Cdc14 release transient, the MEN is required to sustain Net1 phosphorylation and Cdc14 release at later stages of mitosis. In addition to that, the MEN also helps in the cytoplasmic accumulation of Cdc14 by restricting its nuclear import.

Activation of MEN is complex and it is linked to the entry of a spindle-pole-body (SPB) into the bud. This control allows the cells to exit mitosis only after proper positioning of anaphase spindle (spindle orientation checkpoint) (Bardin and Amon 2001; Stegmeier and Amon 2004). Movement of SPB into bud depends on the cleavage of sister-chromatid cohesins which is catalyzed by the proteolytic activity of separase. The MEN component, Tem1 is a small G-protein localized at the SPB and it is negatively regulated by GTPase activating protein (GAP) complex, Bfa1/Bub2. Tem1 activation depends on a putative GDP:GTP exchange factor, Lte1, present at the bud cortex. In this way, Tem1 activation depends on sister-chromatid separation and anaphase spindle elongation which brings Tem1 and its activator, Lte1 in proximity. Tem1 triggers the activation of two MEN kinases, Cdc15 and Mob1- Dbf2, which are localized at the daughter SPB. Both Cdc15 and Mob1 are inactivated by Cdk1/CycB-mediated phosphorylation, and they undergo

dephosphorylation in Cdc14-dependent manner during mitotic exit. Therefore, MEN activation and Cdc14 release comprise a positive feedback loop by which they activate each other. The early release of Cdc14 by FEAR network provides an initial activation for MEN which helps to turn on the positive feedback loop (Queralt et al. 2006). This coordination ensures that the Cdc14-MEN positive feedback depends on the activation of separase and thereby on the execution of anaphase. In the absence of FEAR network function, activation of the Cdc14-MEN positive feedback is delayed and hence cells exit from mitosis only after a time delay.

It can be seen that during mitotic exit budding yeast cells control the Cdk1/Cdc14 activity ratio to flip the Cdk1-Cdh1/Sic1 bistable switch from high to low Cdk1 activity. APC/C^{Cdc20} functions as a critical node by controlling both the kinase (Cdk1) and phosphatase (Cdc14) branches of the network required for mitotic exit (Fig. 1). APC/C^{Cdc20} activation decreases the Cdk1/Cdc14 ratio by decreasing the Cdk1 and increasing the Cdc14 activities. APC/C^{Cdc20}-dependent degradation of its substrates (the major mitotic and S phase CycBs, Clb2 and Clb5 respectively as well as securin) is the trigger required to flip the bistable switch (Shirayama et al. 1999).

Once budding cells exited from mitosis in a Cdc14-dependent manner, Cdc14 activity is not required anymore as it can interfere with the entry into next cell cycle. Therefore, the MEN is inactivated and Cdc14 gets re-sequestered into the nucleolus. This is mainly achieved by APC/C^{Cdh1}-dependent degradation of Cdc5, a key component of MEN. This creates the following negative feedback: Cdc5 → Cdc14 → Cdh1 —| Cdc5 in the regulatory network which can produce free running oscillations (so-called Cdc14 endocycles) in the presence of nondegradable Clb2 (Lu and Cross 2010). But Clb2 degradation entrains these free running Cdc14 endocycles with the periodic activation and inactivation of Cdk1. Thus, Cdk1 inactivation during mitotic exit is required for restricting endocycles besides initiating cytokinesis and promoting entry into G1 phase of the next cycle.

Although Cdc14 is essential in budding yeast, it is dispensable for mitotic exit in several other eukaryotes. This could be because in budding yeast Cdc14 activation is required for Cdk1 downregulation (through Cdh1/Swi5 dephosphorylation) besides to dephosphorylate several other Cdk1 substrates. However, in other organisms the Cdk1 inactivation is achieved by APC/C^{Cdc20}-dependent degradation of CycB only, thereby exit from mitosis can be driven by a constitutive phosphatase. In fission yeast, Clp1 (Cdc14 homologue) phosphatase is not crucial for mitotic exit but it regulates G2/M transition. Clp1 and Cdk1 exist in antagonistic relationship, which helps to upregulate the phosphatase with Cdk1 downregulation by APC/C^{cdc20} and provide switch like behavior to mitotic exit. Similar to MEN, fission yeast has septation initiation network (SIN), but it only has a role in cytokinesis and not in mitotic exit (Bardin and Amon 2001).

Hence, the mitotic exit control in the cell cycle regulation depends on how the Cdk downregulation is coordinated with phosphatase(s) regulation. There is true lack of knowledge about the phosphatases and their regulation with regard to mitotic exit in several other organisms. As seen in budding yeast, Cdc14 upregulation helps to flip the bistable switch Cdk1-Cdh1 toward lower Cdk1 activity. A positive feedback loop on the MEN helps to accelerate the Cdc14 release and exit. It has to be seen whether such an upregulated phosphatase control is required to drive mitotic exit in other eukaryotes.

Cross-References

- ▶ Cell Cycle, Budding Yeast
- ▶ Cell Cycle Checkpoints
- ▶ Cell Cycle Dynamics, Irreversibility
- ▶ Cell Cycle, Fission Yeast
- ▶ Cell Cycle of Mammalian Cells
- ▶ Cell Cycle Transitions, G2/M
- ▶ Cyclins and Cyclin-dependent Kinases
- ▶ Endoreplication
- ▶ Mitosis

References

Bardin AJ, Amon A (2001) Men and sin: what's the difference? Nat Rev Mol Cell Biol 2(11):815–826

Chen KC, Calzone L, Csikasz-Nagy A, Cross FR, Novak B, Tyson JJ (2004) Integrative analysis of cell cycle control in budding yeast. Mol Biol Cell 15(8):3841–3862

D'Amours D, Amon A (2004) At the interface between signaling and executing anaphase-Cdc14 and the FEAR network. Genes Dev 18(21):2581–2595
Lu Y, Cross FR (2010) Periodic cyclin-Cdk activity entrains an autonomous Cdc14 release oscillator. Cell 141(2):268–279
Morgan DO (1999) Regulation of the APC and the exit from mitosis. Nat Cell Biol 1(2):E47–E53
Morgan DO (2007) The cell cycle. Principles of control. New Science Press Ltd, London
Queralt E, Uhlmann F (2008) Cdk-counteracting phosphatases unlock mitotic exit. Curr Opin Cell Biol 20(6):661–668
Queralt E, Lehane C, Novak B, Uhlmann F (2006) Downregulation of PP2A (Cdc-95) phosphatase by separase initiates mitotic exit in budding yeast. Cell 125(4):719–732
Shirayama M, Toth A, Galova M, Nasmyth K (1999) APC (Cdc20) promotes exit from mitosis by destroying the anaphase inhibitor Pds1 and cyclin Clb5. Nature 402(6758):203–207
Stegmeier F, Amon A (2004) Closing mitosis: the functions of the Cdc14 phosphatase and its regulation. Annu Rev Genet 38:203–232

Cell Cycle, Archaea

Chris J. Norbury
Sir William Dunn School of Pathology,
University of Oxford, Oxford, UK

Definition

The archaea, the "third domain of life" (as distinct from bacteria and eukaryotes), have some of the general features of bacteria, notably single, circular chromosomes, but the biochemistry of the archaeal cell cycle is much more closely related to that of eukaryotes than it is to that of bacteria.

Cross-References

► Cell Cycle, Physiology

References

Barry ER, Bell SD (2006) DNA replication in the archaea. Microbiol Mol Biol Rev 70:876–887
Bernander R (2007) The cell cycle of Sulfolobus. Mol Microbiol 66:557–562

Cell Cycle, Biology

Roberta Alfieri and Luciano Milanesi
Institute for Biomedical Technologies – CNR (Consiglio Nazionale delle Ricerche), Segrate, Milan, Italy

Synonyms

Cell division cycle; Cell proliferation process

Definition

The cell cycle is a complex and crucial event for the life of every organism (Mitchison 1971; Nurse 2000). It consists of a series of ordered events, which regulate the cellular growth and the division depending on external stimuli. The cell cycle can also be defined as the space which elapses from the formation of a cell as result of mitosis and its complete division into two daughter cells. Cell division can occur in two different ways, depending on the nature of the cells themselves: A somatic cell undergoes a cellular division named mitosis while a germ cell undergoes a cellular division named meiosis (Jorgensen and Tyers 2004).

Cell division consists of two consecutive processes, mainly characterized by DNA replication and segregation of replicated chromosomes into two separate cells. Originally, cell division was divided into two stages: mitosis (M), i.e., the process of nuclear division; and interphase, the interlude between two M phases. Stages of mitosis include prophase, metaphase, anaphase, and telophase. The interphase cells simply grow in size, but different techniques revealed that the interphase includes G1, S, and G2 phases (Norbury and Nurse 1992; Blow and Tanaka 2005). Replication of DNA occurs in a specific part of the interphase called S phase.

Cross-References

► Cell Cycle Database

References

Blow JJ, Tanaka TU (2005) The chromosome cycle: coordinating replication and segregation. EMBO Rep 6(11): 1028–1034
Jorgensen P, Tyers M (2004) How cells coordinate growth and division. Curr Biol 14:1014–1027
Mitchison JM (1971) The biology of the cell cycle. Cambridge University Press, Cambridge
Norbury C, Nurse P (1992) Animal cell cycles and their control. Annu Rev Biochem 61:441–470
Nurse P (2000) A long twentieth century of the cell cycle and beyond. Cell 100(1):71–78

Cell Cycle, Budding Yeast

John J. Tyson[1], Katherine C. Chen[1] and Béla Novák[2]
[1]Department of Biological Sciences, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA
[2]Department of Biochemistry, Oxford Centre for Integrative Systems Biology, University of Oxford, Oxford, UK

Synonyms

Saccharomyces cerevisiae

Definition

The cell cycle of budding yeast has become a hallmark problem of molecular systems biology for a number of reasons. (1) The cycle of DNA replication, mitosis, and cell division is crucial to all aspects of biological growth, development, and reproduction. (2) The genes and pathways for cell cycle control seem to be very similar in all eukaryotes, including mammals. (3) Budding yeast is exceptionally good for genetic analysis of cell cycle controls because it can proliferate as haploid cells, and its genetic makeup can be easily altered by standard tools of molecular genetics. For this reason, a great deal is known about the molecular machinery regulating the events of the budding yeast cell cycle. (4) The machinery is very complicated, and its properties cannot be worked out reliably by intuitive reasoning alone. (5) Comprehensive mathematical models of the budding yeast cell cycle have been built and tested against a wide range of experimental observations. (6) The models make interesting predictions that have been confirmed, in large part, by experimental tests.

Characteristics

Physiology

Budding yeast gets its name from its unusual style of asymmetric division into a large mother cell and a small daughter cell (Pringle and Hartwell 1981). After a G1 period, the budding yeast cell initiates a new bud at about the same time that it enters S phase (DNA synthesis). Also, at this time, the yeast cell replicates its spindle pole bodies and begins preparations for mitosis. These cotemporaneous events of the budding yeast cell cycle are referred to as START. The bud first emerges from the cell in a burst of polarized growth but quickly switches to isotropic growth, to form an expanding spherical protrusion. Most of the net cell growth after this time goes into the bud. After DNA synthesis is finished, an intranuclear mitotic spindle is built and the replicated chromosomes are aligned at the metaphase plate. Simultaneously, the G2/M nucleus migrates to the neck between the mother and bud compartments and orients itself with one pole of the mitotic spindle in the mother cell and the other pole in the bud. During anaphase, the replicated chromosomes are partitioned into two groups: one group is pushed into the mother cell and the other into the bud. The stretched nucleus divides in two, and the cell separates at the bud neck to produce mother and daughter cells. The daughter cell generally has a considerably longer G1 period than the mother cell. It must grow to a certain threshold size before it can initiate a new round of budding, DNA replication, and division.

Progression through the budding yeast cell cycle is governed by a series of checkpoints. For instance, we have already mentioned that daughter cells must grow to a critical size before they can execute START. Also, DNA damage can delay execution of START until the damage is repaired. These cells will not execute the metaphase/anaphase transition until all

chromosomes are fully replicated and properly aligned on the mitotic spindle (the "spindle assembly checkpoint" or SAC). Budding yeast cells have a "morphogenetic checkpoint" which blocks mitosis if the bud is not properly formed, and a spindle alignment checkpoint that blocks cell division if the daughter cell nucleus is not properly pushed into the bud compartment.

Molecular Cell Biology

Genetic analysis of the molecular machinery controlling the major events of the cell cycle was begun in budding yeast by Leland Hartwell, who identified dozens of genes that induced the "cell division cycle" phenotype. By Hartwell's definition, a *CDC* gene is a gene that, when mutated, arrests all cells in a specific stage of the cell cycle. For example, *cdc7* mutant cells do not replicate their DNA or divide, but bud emergence and growth continue as normal. Mutant *cdc* genes are isolated as temperature-sensitive alleles that on replicate plates form colonies at 25°C but not at the restrictive temperature (36°C). Presumptive mutant cells on the high temperature plate must be examined microscopically to determine that the cells have uniformly arrested at a particular morphological stage of the cell cycle. Hartwell's first set of *CDC* genes defined two parallel pathways (Hartwell et al. 1974): one pathway relating to bud emergence, isotropic growth, and nuclear migration, and the other relating to DNA replication and mitosis. These two pathways diverge at START and come together finally at mitotic exit. The gene with the earliest effect, whose mutation blocked cells before the START transition, was *CDC28*.

Hartwell's work set off a landslide of activity identifying *CDC* genes in a variety of organisms, finding other genes that interacted with *CDC* genes, and later cloning these genes and characterizing their protein products and the interactions among these proteins (for review, see Morgan 2007). Worthy of note was Paul Nurse's discovery of a fission yeast gene, $cdc2^{+}$, which seemed to encode the same protein as *CDC28* in budding yeast. Nurse went on to show, quite unexpectedly, that a human gene, later called *CDK1*, encodes a protein that can rescue $cdc2^{-}$ mutant fission yeast cells. In 1989 a consortium of geneticists and biochemists demonstrated that MPF (mitosis promoting factor in frog eggs, a master controller for the progression of the cell cycle) is a dimer of a protein kinase, encoded by the *CDC28*/$cdc2^{+}$/*CDK1* gene, and a regulatory subunit called cyclin B which had been identified years earlier by Tim Hunt. Because the cyclin subunit is required for activity of the protein kinase, the kinase is called cyclin-dependent kinase, or CDK for short. The discovery of the composition of MPF was the beginning of the modern era of molecular biology of cell cycle control in eukaryotic cells. In 2001 the Nobel Prize in Physiology or Medicine was shared by Hartwell, Nurse, and Hunt.

In the 1990s a host of cell cycle genes were identified in budding yeast, including:

CLN1,2: encoding G1 cyclins important for START and bud emergence

CLB1-6: encoding B-type cyclins essential for DNA synthesis and mitosis

CDC20: encoding a ubiquitin-ligase component essential for sister chromatid separation and cyclin degradation in anaphase

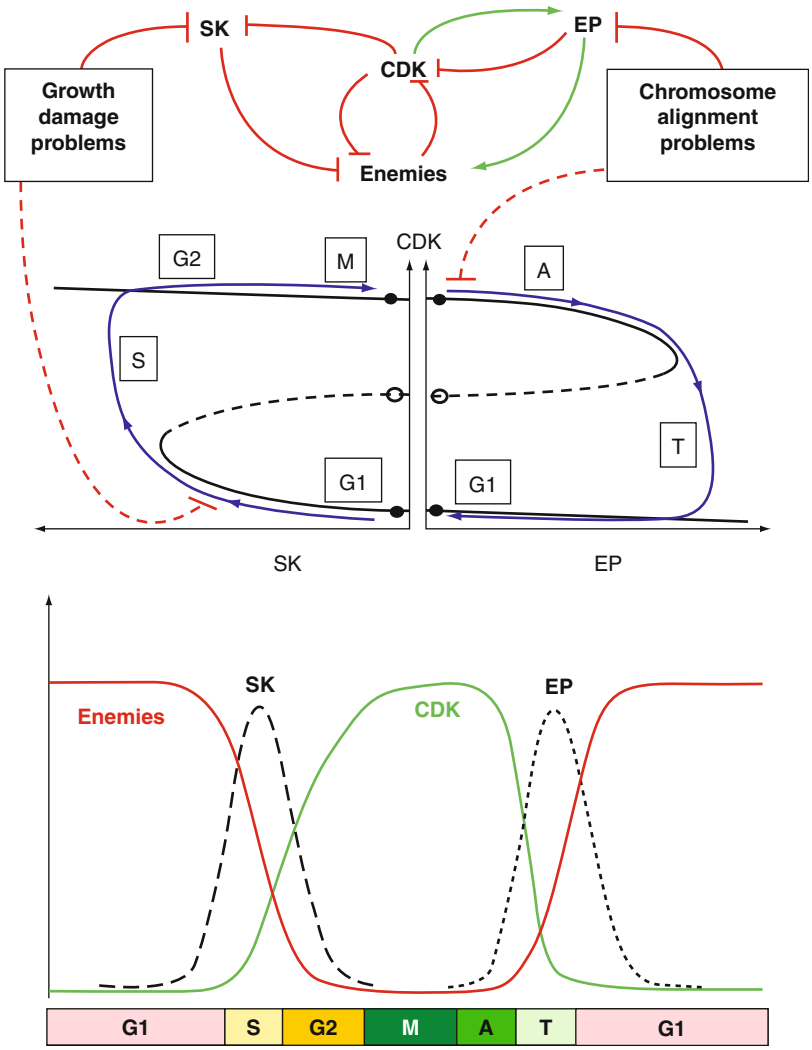
CDC14: encoding a phosphatase essential for mitotic exit

CDH1, SIC1: encoding proteins that stabilize G1 phase of the cell cycle

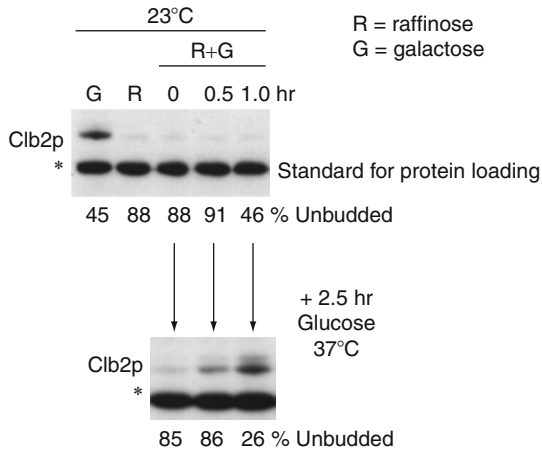
The Heart of the Budding Yeast Cell Cycle

Kim Nasmyth (1996) proposed to think of the budding yeast cell cycle as two alternative, self-maintaining states: *G1* (unreplicated chromosomes, growing but not committed to proliferate), and *S/G2/M* (committed to chromosome replication and mitosis). START is the transition from *G1* to *S/G2/M*, and mitotic exit is the transition from *S/G2/M* back to *G1*. The molecular correlates of these two states are *G1*: high activity of Cdh1 and Sic1, no kinase activity associated with Clb1-6; and *S/G2/M*: low activity of Cdh1 and Sic1, high Clb-dependent kinase activity. Finally, Nasmyth recognized that a possible reason for these alternative, self-maintaining states was the antagonism between Cdh1 and Sic1 on one side (destroying the activity of the Clb-dependent kinase complexes) and the B-type cyclins, Clb1-6, on the other side (inactivating Cdh1 and Sic1).

Nasmyth stopped short of identifying his idea with a bistable control system flipping back and forth between two stable steady states as cell-cycle-progression signals pushed the control system past



Cell Cycle, Budding Yeast, Fig. 1 The heart of the budding yeast cell cycle. Adapted from Tyson and Novak (2008). (*Top center*) Cyclin-dependent kinase (CDK) and its enemies (Cdh1 and Sic1) are involved in an antagonistic (double negative) feedback loop, which can persist in either of two stable steady states: a G1-like state with low CDK activity and prominent enemy forces, and an M-like steady state with high CDK activity and the enemies in retreat. These two stable steady states are represented by black circles along the CDK axis (*middle center*). The pairs of circles side-by-side are meant to represent the same steady state under conditions when both the starter kinase (SK) and exit phosphatase (EP) are close to zero. The two stable steady states are separated by an unstable steady state (*open circle*). A newborn cell in G1 (*middle left*) can be induced to enter S/G2/M by activation of SK (Cln-dependent kinase), which phosphorylates and weakens Cdh1 and Sic1, allowing CDK activity to rise and trigger S phase (follow the *blue arrow* for the cell's passage through the cell cycle). Among other duties, CDK downregulates SK, but the bistable switch remains in the upper state. Notice that, as SK activity increases, the stable G1-like steady state merges with and is annihilated by the unstable steady state at a saddle-node bifurcation. Exit from mitosis (*middle right*) is induced by EP (Cdc14 phosphatase), which activates the proteins that destroy CDK activity. When the upper steady state merges with the unstable steady state, the cell returns to the G1 state. Because EP activity depends on CDK, after CDK falls, EP activity also disappears, but now the enemies have the upper hand. The two blue arrows together form a "hysteresis loop" that switches the CDK-control system from OFF to ON and back again during each cell cycle. (*Bottom*) Progression around the hysteresis loop corresponds to periodic changes in the activities of CDK, enemies, SK, and EP. At START, a burst of SK activity flips the CDK switch ON, and at EXIT, a burst of EP activity flips the CDK switch OFF. (Top left and right) Checkpoints function by inhibiting either SK or EP. Growth requirements and damage repair processes typically halt the cell cycle in G1 phase by preventing up-regulation of SK activity. Misalignment of chromosomes on the spindle blocks exit from mitosis by preventing activation of EP



Cell Cycle, Budding Yeast, Fig. 2 Experimental confirmation of alternative stable states with low or high Clb2-dependent kinase activity. Adapted from Cross et al. (2002). To test the prediction that a stable *G1* state of low Clb-dependent kinase activity coexists with a stable *S/G2/M* state of high Clb-dependent kinase activity when SK = EP = 0, Fred Cross and colleagues measured Clb2 abundance and cell-cycle phase in a mutant strain of budding yeast, *cln1Δ cln2Δ cln3Δ GAL-CLN3 cdc14^{ts}*. When grown in raffinose or glucose, the *GAL* promoter is repressed, and the cells are unable to make any Cln proteins, so they have no Cln-dependent kinase activity (SK = 0 in Fig. 1). When grown at 37°C, this strain has no Cdc14 activity (EP = 0 in Fig. 1). In raffinose or glucose at 37°C, this strain has neither Cln3 nor Cdc14 activity, i.e., it is in the "neutral" position in Fig. 1. In galactose-rich medium (*GAL* promoter active) at 23°C, the cells have both SK and EP, so they are able to progress around the cell cycle (the blue hysteresis loop in Fig. 1). Under these conditions, the cells proliferate asynchronously, as indicated (*upper left*) by the facts that 45% of the cells are unbudded (in G1) and the remainder of the cells are in S/G2/M with significant amounts of Clb2 protein. When this asynchronous population of cells is shifted to raffinose medium at 23°C for 6 h (i.e., deprived of SK), the cells arrest primarily in G1 phase (88% unbuddded) with undetectable levels of Clb2 (lane R). This is the *G1* steady state in Fig. 1. Next, the culture is shifted to galactose + raffinose medium at 23°C to induce production of Cln3. At the end of the indicated periods of time, two aliquots of the culture are removed. One aliquot is analyzed immediately for bud index and Clb2 content (lanes 3, 4, 5 of *upper gel*); the other aliquot is shifted to glucose medium at 37°C for 2.5 h and then analyzed as before (*lower gel*). During the variable period of exposure to galactose, the cells produce increasing amounts of Cln3 protein (i.e., increasing pulses of SK in Fig. 1). Cells exposed to a small pulse of Cln3 (0 or 0.5 h in galactose) remained in the stable *G1* state, as indicated by lack of buds and little or no Clb2 protein (lanes 1 and 2 of *lower gel*). Cells that got a larger pulse of Cln3 (1 h in galactose) switched to the stable *S/G2/M* state (mostly budded cells with abundant Clb2 – lane 3 of *lower gel*) and stayed there because EP was disabled at the higher temperature. The cells in all three lanes of the lower gel are exposed to the same conditions (SK = EP = 0), but the cells in lanes 1 and 2 are in the *G1* state, whereas the cells in lane 3 are in the *S/G2/M* state; i.e., the system is bistable

saddle-node bifurcations. But this connection became the basis of a series of mathematical models of the budding yeast cell cycle control system developed by Novak, Tyson, and coworkers (Chen et al. 2004; Tyson and Novak 2008) and by Alberghina et al. (2009). A brief overview of these models is presented in Fig. 1, and the interested reader is referred to the original literature for details.

Experimental Tests

Cross and collaborators (Cross et al. 2002) have confirmed experimentally a number of predictions of the Chen model, including the idea of two alternative stable steady states (see Fig. 2).

Cross-References

- ▶ Cell Cycle Dynamics, Bistability and Oscillations
- ▶ Cell Cycle Dynamics, Irreversibility
- ▶ Cell Cycle Model Analysis, Bifurcation Theory
- ▶ Cell Cycle Modeling, Differential Equation
- ▶ Cell Cycle of Early Frog Embryos
- ▶ Cell Cycle, Fission Yeast

Web Resources

Model webpage: http://mpf.biol.vt.edu/research/budding_yeast_model/pp/

References

Alberghina L, Coccetti P, Orlandi I (2009) Systems biology of the cell cycle of *Saccharomyces cerevisiae*: from network mining to system-level properties. Biotechnol Adv 27:960–978

Chen KC, Calzone L, Csikasz-Nagy A, Cross FR, Novak B, Tyson JJ (2004) Integrative analysis of cell cycle control in budding yeast. Mol Biol Cell 15:3841–3862

Cross FR, Archambault V, Miller M, Klovstad M (2002) Testing a mathematical model for the yeast cell cycle. Mol Biol Cell 13:52–70

Hartwell LH, Culotti J, Pringle JR, Reid BJ (1974) Genetic control of the cell division cycle in yeast. Science 183:46–51

Morgan DO (2007) The cell cycle: principles of control. New Science, London

Nasmyth K (1996) At the heart of the budding yeast cell cycle. Trends Genet 12:405–412

Pringle JR, Hartwell LH (1981) The *Saccharomyces cerevisiae* cell cycle. In: Strathern JN, Jones EW, Broach JR (eds)

The molecular biology of the yeast *Saccharomyces*. Cold Spring Harbor Laboratory, Cold Spring Harbor, pp 97–142
Tyson JJ, Novak B (2008) Temporal organization of the cell cycle. Curr Biol 18:R759–R768

Cell Cycle, Cancer Cell Cycle and Oncogene Addiction

Vito Quaranta[1], Darren Tyson[2] and Peter Frick[2]
[1]The Vanderbilt-Ingram Cancer Center, Nashville, TN, USA
[2]Vanderbilt University, Nashville, TN, USA

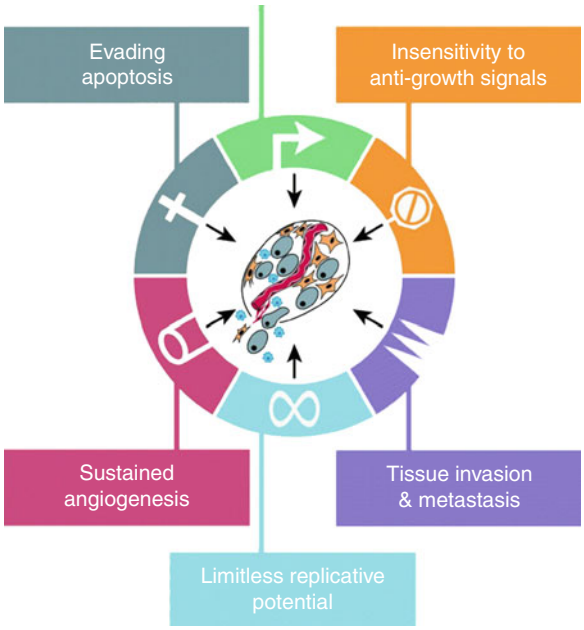
Synonyms

Cancer cell cycle; Oncogene addiction

Definition

Cancer Cell Cycle

Cancer is a disease of cell proliferation, whereby cancer cells progressively and inexorably lose normal cell cycle control. As a consequence of tumor growth, the burden on patients increases and tumors eventually spread and metastasize. Cancers evolve through a multistep progression of accumulating somatic genetic changes (Vogelstein and Kinzler 1993) accompanied with epigenetic abnormalities (Jones and Baylin 2002). These alterations lead to an increasing departure from regular cellular function and confer progressively tumorigenic phenotypes. Cancers arise out of normal tissues, and it is generally thought that they originate from a single cell. This cell loses the intrinsic cellular processes that prevent abnormal cell division and acquires a conserved set of traits – hallmarks – characteristic of all cancers (Hanahan and Weinberg 2000). Of these hallmarks, three relate to the ▶ cell cycle: self-sufficiency in proliferative signals, insensitivity to anti-proliferative signals, and limitless replicative potential. Normal cells require environmental cues to divide, may receive signals to stop dividing, and have a limited number of replicative cycles; the cancer cell cycle, by definition, contains alterations that bypass these regulatory mechanisms. Any cancer therapy must therefore address the uncontrolled cell proliferation driven by the dysregulated cell cycle (Fig. 1).



Cell Cycle, Cancer Cell Cycle and Oncogene Addiction, Fig. 1 A schematic of the classic six hallmarks of cancer taken from the original article (Hanahan and Weinberg 2000), courtesy of Elsevier Limited

Characteristics

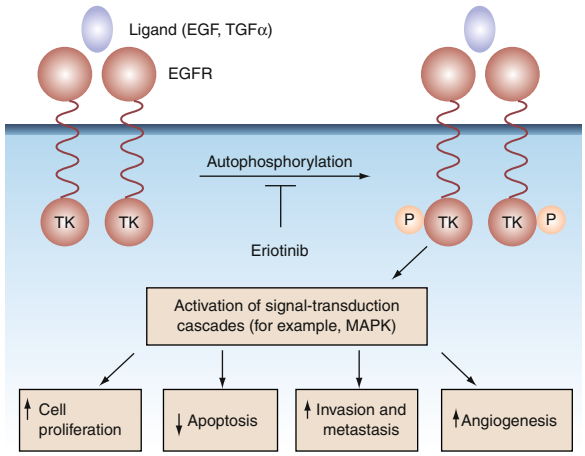
Oncogene Addiction

While a minimalistic phenotypic description can describe all cancers, cancer types are heterogeneous with respect to diverse molecular and clinical descriptions. A variety of molecular alterations can synergistically act to dysregulate the cell cycle. In the past decade, it has become apparent that subset of cancers displays a profound dependence on one or a few oncogenes – so-called oncogene-addicted cancers (Weinstein 2002). While cancers typically incur considerable genetic damage, these oncogene-addicted cancers rely heavily on the activity of few oncogene products. The induced signaling from these genes supports the tumor phenotype and is necessary for its maintenance. Inhibition or deletion of these proteins leads to a reversal of the tumorigenic phenotype.

Recognition of the oncogene-addiction phenomenon has opened the way to an extremely promising therapeutic opportunity: dramatic clinical responses and minimal patient toxicity (Sawyers 2003). Transgenic mouse models and cell lines validate this prospect, but ultimately, patient tumor regressions (such as Chronic Myeloid Leukemia patients treated with a BCR-ABL inhibitor, Imatinib) constitute the most convincing proofs. In the context of the cancer cell cycle, the deactivation of a single oncogene in these addicted cancers may overcome the ability of a cancer cell to bypass normal cellular regulation. However, it should be noted that the molecular basis for oncogene addiction is not fully understood.

Oncogene Addiction and the Cell Cycle

The relationship between oncogene addiction and the cell cycle is perhaps best illustrated by an example, activating mutations in the epidermal growth factor receptor (EGFR) driving cell cycle progression. The EGFR is a transmembrane receptor tyrosine kinase that, upon ligand binding, propagates downstream signaling to elicit primarily pro-survival and proliferative responses. A subset of lung cancer patients has a conserved set of EGFR mutations causing ligand-independent signaling; these mutations predict favorable patient responses to specific EGFR inhibitors, e.g., erlotinib, gefitinib (Pao and Chmielecki 2010). Studies in cultured cancer cell lines with addictive EGFR mutations suggest that EGFR inhibition leads to quiescence and apoptosis. Molecular antagonism of EGFR in these cases shows how therapeutics may effectively target the altered cancer cell cycle. More accurately, it is the inhibition of biochemical signaling downstream of EGFR that mediates cellular responses to the drug, i.e., *pathway* inactivation. Drug-induced alternative receptor signaling, e.g., MET amplification, illustrates this point, because it can rescue drug-mediated pathway inhibition in these EGFR-reliant cells even in the presence of successful EGFR inhibition. Therefore, our example shows an important paradigm: Cell cycle pathway regulation correlates with phenotypic cellular changes. Notably, many proteins implicated in oncogene addiction have established roles in cell cycle regulation – EGFR, HER2, MYC, BCR-ABL, Cyclin D, RAS, WNT, etc. (Fig. 2) (Weinstein and Joe 2008).



Cell Cycle, Cancer Cell Cycle and Oncogene Addiction, Fig. 2 EGFR and erlotinib in cancer signaling. Ligand binding induces dimerization (*left*), subsequent phosphorylation and initiation of downstream signaling (*right*). In cancer, EGFR signaling can drive the tumorigenic phenotype. Clinically used inhibitors such as Erlotinib prevent the phosphorylation of the intracellular substrate and can prevent EGFR signaling (Figure taken from (Minna and Dowell 2005), courtesy of Nature Publishing Group)

Limitations

Oncogene addiction, however, comprises only a fraction of all cancers. Most cancers also have deregulation of other pathways along with altered molecular feedback mechanisms, thus reducing the opportunity for single-agent therapy. Even in oncogene-addicted cancers, tumor heterogeneity and genetic instability prevent lasting treatment responses in patients. Still, the pathway inhibition concept suggests that proper inhibition of tumor cell signaling may effectively overcome deregulation of the cancer cell cycle. Indeed, the late I. Bernard Weinstein highlighted the need for rational drug combinations and the promise of systems biology approaches to understand complex cancer signaling (Weinstein and Joe 2008). At the cellular level, ▶ cancer systems biology may also aid cancer therapeutics by lending insight into the dynamics underlying cancer treatment (Abbott and Michor 2006).

Cross-References

▶ Cancer Systems Biology
▶ Cell Cycle

References

Abbott LH, Michor F (2006) Mathematical models of targeted cancer therapy. Br J Cancer 95(9):1136–1141

Hanahan D, Weinberg RA (2000) The hallmarks of cancer. Cell 100(1):57–70

Jones PA, Baylin SB (2002) The fundamental role of epigenetic events in cancer. Nat Rev Genet 3(6):415–428

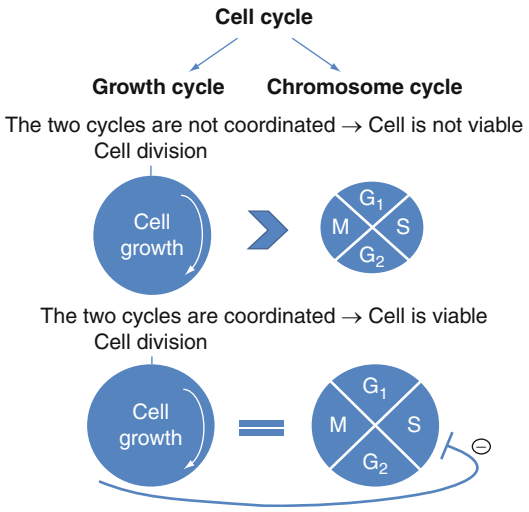
Minna JD, Dowell J (2005) Erlotinib hydrochloride. Nat Rev Drug Discov Suppl 4:S14–S15

Pao W, Chmielecki J (2010) Rational, biologically based treatment of EGFR-mutant non-small-cell lung cancer. Nat Rev Cancer 10(11):760–774

Sawyers CL (2003) Opportunities and challenges in the development of kinase inhibitor therapy for cancer. Genes Dev 17(24):2998–3010

Vogelstein B, Kinzler KW (1993) The multistep nature of cancer. Trends Genet 9(4):138–141

Weinstein IB (2002) Cancer. Addiction to oncogenes – the Achilles heal of cancer. Science 297(5578):63–64



Cell Cycle, Cell Size Regulation, Fig. 1 Checkpoints ensure size homeostasis via negative signal(s) from the growth cycle to the chromosome cycle

Cell Cycle, Cell Size Regulation

Ákos Sveiczer[1] and Anna Rácz-Mónus[2]
[1]Department of Applied Biotechnology and Food Science, Budapest University of Technology and Economics, Budapest, Hungary

Synonyms

Size checkpoint; Size control

Definition

Cell size regulation during the cell cycle is a mechanism coordinating growth and division. It ensures that cell size distribution of a population remains constant in consecutive generations.

Characteristics

Size homeostasis is a general characteristic feature of populations of unicellular organisms, at least in steady-state conditions. It means that an "average" cell must double its size and all its components between two successive divisions. At the individual cellular level, size control mechanisms operate, which ensure that at specific checkpoints the cell cycle progression is blocked, unless a critical size has been reached. Such ▶ cell cycle checkpoints exist both in G1 and G2 phases, which regulate the onset of S and M phases, respectively (Alberts et al. 2009). However, in a specific cell type, both size checkpoints do not necessarily work efficiently: one might dominate, while the other becomes cryptic. In the absence of both size controls, cells would be smaller and smaller from cycle to cycle and finally probably lose viability (Fantes and Nurse 1981; Sveiczer et al. 1996). The reason is that growth (cytoplasmic) cycle was normally longer than chromosome (nuclear) cycle; therefore the cell would not double its components and its size before cytokinesis (Fig. 1). By contrast, size control extends the duration of either G1 or G2 phase by blocking the next phase transition, ensuring a compensation mechanism, that is, the larger the cell at birth is, the shorter its cycle will be. With other words, size homeostasis is achieved via negative signal(s), which are generated by the growth cycle and affect progression of the chromosome cycle. This entry summarizes how size checkpoints act in the most important model organisms studied so far, namely, budding yeast, fission yeast, early embryonic cells, and finally mammalian cells. In every case, we will briefly discuss the methods used in size control studies and the conclusions drawn.

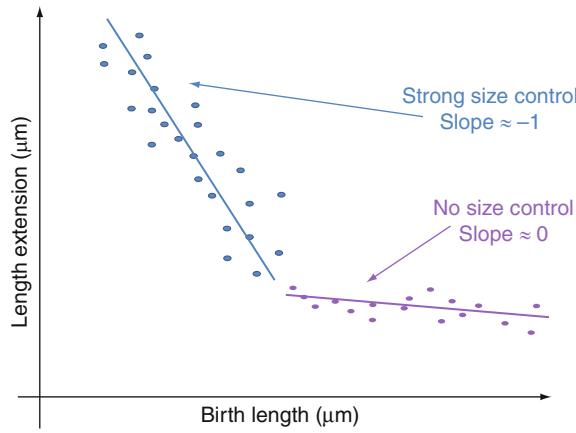
At the end, we will give a generalized view of size regulation in eukaryotes from aspects of systems biology. Another important side of size control is its modulation by changes in the environment (shift-up and shift-down experiments); however, due to space limitations this point will not be discussed here.

Cell Size Regulation in Budding Yeast

Unicellular models like the budding yeast *Saccharomyces cerevisiae* are practically able to grow and divide infinitely, if the required nutrients (C-source, N-source, minerals, etc.) are available in the environment and the physical factors (pH, temperature, water activity, etc.) also support proliferation. Although some cells age and even die, the whole population may increase exponentially for many generations if the milieu does not change. Coupling growth and division happens in late G1, controlling the G1/S transition, and the control point is called the Start event in this organism (Morgan 2007). The mechanism of the intracellular size measurement is mainly obscure. An indirect indicator of cell size is probably the protein synthesis rate, which is proportional to the number of ribosomes and also to cell mass. Budding yeast has only one type of cyclin-dependent kinase (Cdk), called Cdc28; however several different cyclins (Clns and Clbs) are activated at different cell cycle stages, moreover, the different cyclin-Cdc28 complexes have different substrate specificities, leading to an irreversible progression through the whole cycle (▶ Cell Cycle, Budding Yeast). Cln3 is the first cyclin emerging in G1, and this protein is highly unstable. Therefore, Cln3 is a good candidate to be a size-sensing protein, whose concentration reflects translation rate in the cell (Morgan 2007). This hypothesis is also supported by the fact that cells overproducing Cln3 are abnormally small. Probably the most relevant effect of the Cln3-Cdc28 kinase complex is that it directly phosphorylates a nuclear Start inhibitor, the Whi5 protein. Phosphorylated Whi5 is exported from the nucleus, allowing a specific transcription program to start, leading (among many other proteins) to Cln1 and Cln2 production, and finally to initiation of ▶ DNA replication (Di Talia et al. 2007).

Cell Size Regulation in Fission Yeast

Its regular cylindrical shape, tip growth with constant diameter, and symmetric division make the fission yeast *Schizosaccharomyces pombe* an attractive model in size control studies. Cell length, which is proportional to volume and mass, is an easily measurable variable, if cells have been grown under a thermostated photomicroscope on an agar surface. After measuring many cells from time-lapse films, and plotting the total length extension versus birth length (Fig. 2), a strong negative correlation indicates the existence of a size checkpoint (Fantes 1977). The closer the slope of the regression line to −1 is, the stronger the size control is. In abnormally oversized temperature-sensitive *cdc* mutants generated by a block and release experiment, however, size control is abolished, as indicated by the lack of negative correlation. This size checkpoint, in contrast to budding yeast, regulates the onset of ▶ mitosis, acting somewhere in the middle of G2 phase (Sveiczer et al. 1996). In small *wee1* mutants, this checkpoint no more operates; however, a strong G1/S size regulation is turned on here, which mechanism is cryptic in normal-sized wild type cells. The main cyclin-dependent kinase complex in fission yeast is the Cdc13-Cdc2 dimer, also known as the M phase promoting factor (MPF). Its activity in G2 is mainly regulated by a kinase-phosphatase pair, Wee1 and Cdc25, phosphorylating and dephosphorylating the inhibitory Tyr-15 site of Cdc2, respectively (▶ Cell Cycle, Fission Yeast). Similarly to budding yeast, a translational control seems to be involved in size sensing, in the case



Cell Cycle, Cell Size Regulation, Fig. 2 The existence of a size checkpoint in fission yeast is indicated by a strong negative correlation between total length extension and birth length. In extremely oversized cells, size control is abolished

of fission yeast both on the levels of Cdc13 and Cdc25 (Morgan 2007). Analyzing time-lapse films of several cell cycle mutants of *S. pombe*, however, indicates that its mitotic size control acts mainly through the Wee1 protein, while the cryptic Start size checkpoint operates via the cyclin-dependent kinase inhibitor, Rum1 (Sveiczer et al. 1996).

Cell Size Regulation in Early Embryos

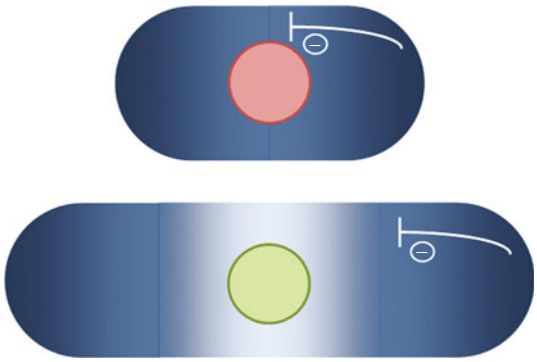
In early embryos of multicellular animals, cell growth and division are uncoupled, that is, size homeostasis is not maintained by any regulatory mechanisms. The most frequently studied models are embryos of the fly *Drosophila* and of the frog *Xenopus* (▶ Cell Cycle of Early Frog Embryos). First, giant eggs are produced by the females, in which the cell cycle is blocked in G2, but growth is continued, stockpiling huge amounts of maternal proteins and mRNA, required later for embryogenesis (Morgan 2007). After maturation and fertilization, rapid and synchronous cycles occur practically without any growth, consisting of only S and M phases, but neither G1 nor G2. Size checkpoint is lost here, or more precisely, the large size makes it cryptic. During about a dozen of these size-independent cycles, cell size dramatically decreases. At around the midblastula transition, however, when cell size reaches a critical level from above, size control starts to operate. Cycle time increases as G1 and G2 phases are involved, and cell divisions are no more synchronous during the following developmental stages. In experiments with *Xenopus*, plotting cycle time as a function of cell radius at birth, similar characteristics were found to those ones discussed above for fission yeast. In cells before the midblastula transition, cycle time was found to be totally independent of birth size; by contrast, after this stage, a strong negative correlation arose between them (Wang et al. 2000). The main cellular parameter controlling these cycles was found to be the nucleocytoplasmic ratio, that is, the DNA content of the cell divided by its cell mass.

Cell Size Regulation in Mammalian Cells

Mammalian cells, similarly to yeasts, require nutrients for proliferation, but they also need growth factors generated normally by other cells of the same organism. With other words, division of these cells is under strict tissue control; in the absence of growth factors the cells are in a phase called G0, and they neither grow nor divide in this stage (Morgan 2007). The effect of growth factors results in passing the G0/G1 transition, when the cell starts to proliferate (▶ Cell Cycle of Mammalian Cells). Their most important checkpoint is the Start control in late G1, similarly to that of budding yeast; however, it is rather called the restriction point (▶ Cell Cycle Transition, Detailed Regulation of Restriction Point) in animal cells. The main molecular regulators of the restriction point are E2F transcription factors required for G1/S progression, an inhibitor of E2Fs (the retinoblastoma protein or RB), and different ▶ cyclins and cyclin-dependent kinases. To get rid of problems of tissue control, in experiments cultured animal cells are generally involved, supplied artificially with growth factors. A further problem is the irregular shape of cells, causing uncertainties in size determination. One possibility is to measure total protein content or protein synthesis rate in the cells, or the other one is to measure cell volume by an electronic cell counter. A very long debate has been continued on to decide whether cell size affects or not the restriction point in animal cells (Morgan 2007). The answer to this question may alter in different cell types; however, a seminal paper suggests us that probably size control mechanisms generally operate in higher eukaryotes (Dolznig et al. 2004), which conclusion has recently been confirmed (Tzur et al. 2009).

Systems Biology of Size Control

How a cell measures its size during the cell cycle is not clear, however, any size-sensing mechanism must work via the biochemical cell cycle engine, driving any cell from birth to division. Therefore, cyclins, Cdks, their inhibitors, and other regulators are generally thought to be involved in size regulation. As discussed above, translational control on highly unstable proteins is hypothesized to have a general role here. So, in mathematical models of the cell cycle consisting of ordinary differential equations on the biochemical reactions of the most important proteins, cyclins are generally assumed to accumulate in the nucleus proportional to cell mass. Such models were shown to describe correctly the size checkpoints in different cell types; for example in the case of fission yeast, even a stochastic version of a former deterministic model was successfully developed (Sveiczer et al. 2001). However, a novel way of size sensing was recently discovered in rod-shaped cells, like in fission



Cell Cycle, Cell Size Regulation, Fig. 3 Rod-shaped cells may measure their size by generating a spatial gradient of a cell cycle inhibitor protein at the cellular cortex. In small cells (top), the inhibitory effect from the tips is large in the middle, which blocks the nuclear cycle. In large cells (bottom), the inhibitory effect from the tips is little in the middle, which releases the block

yeast and even in some bacteria (*Bacillus subtilis*). In these cells, a spatial gradient of some proteins is formed at the cortex, having a maximal local concentration at the cell tips, and a continuous decrease in the direction of the middle (Fig. 3). These proteins inhibit cell cycle progression in the middle of the cell, when the cell is small. As the cell becomes larger and larger, the protein in the middle dilutes, and at a critical cell length, it loses its efficient inhibitory effect (Moseley and Nurse 2010). How far this mechanism is really involved in size regulation and how universal this way might be are interesting questions for the future to be studied.

Acknowledgment Research in our group is supported by the Hungarian Scientific Research Fund (OTKA K-76229) and by the New Széchenyi Plan (Project ID: TÁMOP-4.2.1/B-09/1/KMR-2010-0002).

Cross-References

- ▶ Cell Cycle Checkpoints
- ▶ Cell Cycle of Early Frog Embryos
- ▶ Cell Cycle of Mammalian Cells
- ▶ Cell Cycle Transition, Principles of Restriction Point
- ▶ Cell Cycle, Budding Yeast
- ▶ Cell Cycle, Fission Yeast
- ▶ Cyclins and Cyclin-dependent Kinases
- ▶ DNA Replication
- ▶ Mitosis

References

Alberts B, Bray D, Hopkin K, Johnson A, Lewis J, Raff M, Roberts K, Walter P (2009) Essential cell biology, 3rd edn. Garland Science, New York and London

Di Talia S, Skotheim JM, Bean JM, Siggia ED, Cross FR (2007) The effects of molecular noise and size control on variability in the budding yeast cell cycle. Nature 448: 947–951

Dolznig H, Grebien F, Sauer T, Beug H, Müllner EW (2004) Evidence for a size-sensing mechanism in animal cells. Nat Cell Biol 6:899–905

Fantes PA (1977) Control of cell size and cycle time in *Schizosaccharomyces pombe*. J Cell Sci 24:51–67

Fantes PA, Nurse P (1981) Division timing: controls, models and mechanisms. In: John PCL (ed) The cell cycle. Cambridge University Press, Cambridge, pp 11–33

Morgan DO (2007) The cell cycle. Principles of control. New Science Press, London

Moseley JB, Nurse P (2010) Cell division intersects with cell geometry. Cell 142:189–193

Sveiczer A, Novak B, Mitchison JM (1996) The size control of fission yeast revisited. J Cell Sci 109:2947–2957

Sveiczer A, Tyson JJ, Novak B (2001) A stochastic molecular model of the fission yeast cell cycle: role of the nucleocytoplasmic ratio in cycle time regulation. Biophys Chem 92:1–15

Tzur A, Kafri R, LeBleu VS, Lahav G, Kirschner MW (2009) Cell growth and size homeostasis in proliferating animal cells. Science 325:167–171

Wang P, Hayden S, Masui Y (2000) Transition of the blastomere cell cycle from cell size-independent to size-dependent control at the midblastula stage in *Xenopus laevis*. J Exp Zool 287:128–144

Cell Cycle, Coupled with Circadian Clock

Christian I. Hong
Department of Molecular and Cellular Physiology, University of Cincinnati College of Medicine, Cincinnati, OH, USA

Definition

Cell cycle and ▶ circadian rhythms carry out specific functions to regulate cell divisions and provide temporal controls, respectively. These two distinct mechanisms interact with each other via molecular

coupling factors such as WEE1 and CHK2. WEE1 is directly regulated by a heterodimeric circadian transcription factor, BMAL1/CLOCK. On the other hand, a cell cycle checkpoint regulator, CHK2, phosphorylates one of the core clock components, mPER1, and leads to its subsequent degradation upon DNA damage. This bidirectional coupling between the cell cycle and the circadian clock is conserved from filamentous fungi, *Neurospora crassa*, to mammals.

Characteristics

In most eukaryotic organisms, networks of cell cycle and circadian rhythms coexist and work in harmony to create optimum conditions for cells to grow and adapt to surrounding environment. The underlying molecular mechanisms of cell cycle are conserved from yeast to mammals (Nurse 2000). This robust control system is equipped with multiple checkpoints for controlled growth and cell divisions. Cell cycle produces an oscillatory phenomenon of repeated growth and division. The period of this oscillation, however, varies with external conditions such as nutrient and temperature. Cell cycle system is optimized for growth and division, but not for time keeping. Circadian rhythms keep track of time and provide temporal regulations in most eukaryotic organisms with a period of about 24 h (Dunlap 1999). In contrast to the period of cell division cycle, the period of circadian clock is relatively insensitive to external conditions such as nutrients or temperature. These two molecular mechanisms for distinct functions are connected by molecular coupling components, which are addressed below.

Molecular mechanisms of cell cycle progress and their dynamical properties are identical in all eukaryotes, and extensively studied in various ways including mathematical modeling (Csikasz-Nagy et al. 2006). Key regulatory components are conserved across different species. In mammals, four different complexes of Cyclin-dependent-kinases (CDKs) and their regulatory Cyclin partners (Cdc2/CycB, Cdk2/CycA, Cdk2/CycE, and Cdk4/CycD) orchestrate cell cycle progressions. Their appearances, activities, and disappearances are regulated by multiple components such as inhibitors (Rb, p27Kip1), transcription factors (c-Myc, E2F, Mcm), and degradation factors (p55Cdc/APC, Cdh1/APC). For this entry, it is important to note two cell cycle components: (1) an inhibitory kinase, WEE1, that inactivates the activity of Cdc2 and determines the entry into the mitosis, and (2) a cell cycle checkpoint regulator, CHK2, that is induced upon DNA damage to relay proper signals for DNA damage responses.

In mammals, circadian rhythms are present in various cell types including liver, fibroblasts, etc. The circadian clock is robust and stable at a single cell level. In fibroblasts, however, cells do not communicate with each other, and they are asynchronous as a population. The master clock resides in the suprachiasmatic nucleus (SCN), which is situated in the hypothalamus. These neuronal cells do communicate and synchronize with each other, receive input signals, and relay output signals to other cells. Both SCN neurons and peripheral tissues have identical molecular components that constitute a 24-h oscillator. A conserved PAS domain containing heterodimeric transcription factors, BMAL1/CLOCK or BMAL1/NPAS2, activates transcriptions of core clock components such as *mPers* (*mPer1* and *mPer2*) and *mCrys* (*mCry1* and *mCry2*) (Ukai and Ueda 2010). Translated products from these transcripts, mPERs and mCRYs, translocate into the nucleus and negatively regulate their own transcription factors creating a time-delayed negative feedback loop. In a simplified picture, this feedback circuit generates a stable oscillator. The circadian clock system, however, is much more complex with other interlocked feedback loops with components such as RevErb-α. Furthermore, transcriptional, post-transcriptional, post-translational, and nuclear import/export regulations play important roles in the circadian clock dynamics adding multiple layers of controls.

Cell cycle and circadian rhythms are coupled together despite their discrete functions. The circadian clock–gated cell division cycles are observed in various organisms from cyanobacteria to mammals (Yang et al. 2010; Sahar and Sassone-Corsi 2009). From 1950s, scientists observed cell divisions that only occur during particular times of the circadian cycle (Sweeney and Hastings 1958). However, it was only recently that scientists identified molecular coupling factors between the cell cycle and the circadian clock (Sahar and Sassone-Corsi 2009). A circadian transcription factor, BMAL1/CLOCK, directly binds to *Wee1*'s E-box elements and activates the transcription of *Wee1*. Both the abundance and the

activity of WEE1 oscillate and are high during the evening. These, in turn, determine the duration of G2-phase. Cells divide in late evening when WEE1 level decreases in order to allow cells to enter into the M-phase.

WEE1 undergoes complex regulation, because cell cycle–controlled WEE1 is in concert with another regulation that is periodically imposed by the circadian clock. The period of cell division cycle varies as temperature and/or nutrient conditions change while the period of circadian clock remains relatively constant. This fact sets up a stage for *in silico* studies for the cell cycle and the circadian clock as coupled oscillators. Computational modeling and analysis of this particular coupling via WEE1 revealed: (1) phase-lock and synchronization of cell division cycle with circadian rhythms when the mass doubling time (MDT, or the period of cell cycle) is close to 24 h, (2) quantized cell division cycles when the MDT is different than 24 h, and (3) circadian influenced cell size control (Zamborszky et al. 2007). If the MDT is close to 24 h, the circadian-regulated WEE1 synchronizes cell cycle to the phase of circadian rhythms regardless of initial phase of cell cycle. However, this coupling results in multi-modal cell cycle distributions or quantized cell cycle lengths if the MDT deviates from 24 h, because the circadian clock influences different phases of cell cycle in each generation due to their period differences. The circadian clock may lengthen the MDT via WEE1 depending on its influences on particular cell cycle phase. This periodic break on cell cycle by the circadian clock enforces cell size control by regulating the duration of growth at the G2 phase. These multi-modal distributions are recently reported in cyanobacteria cell cycle system under the influence of the circadian clock with single cell measurements (Yang et al. 2010).

The above seemingly unidirectional coupling was recently shown to be conditionally bidirectional. Circadian rhythms show unique phase shifts upon DNA damage, which involves cell cycle checkpoint regulator, CHK2. DNA damage induces DNA damage responses that slow down cell cycle, arrest cell cycle, or lead to programmed cell death, apoptosis. In mammalian system, ionizing radiation causes double-strand breaks that relays signal transduction cascades via ATM and CHK2. Both ATM and CHK2 interact with mPER1, and CHK2 phosphorylates mPER1, which leads to its subsequent degradations (Sahar and Sassone-Corsi 2009). In other words, CHK2-induced phosphorylation triggers premature degradation of mPER1, which creates phase shift in the circadian clock. The DNA damage–induced phase shifts are unique because DNA damage predominantly creates phase advances, but not much of delays. This unique phase response was investigated with computational modeling and revealed possible molecular criteria for such phenotype: (1) An autocatalytic positive feedback mechanism is required in addition to the time-delayed negative feedback loop and (2) CHK2 phosphorylates and triggers subsequent degradations of mPERs that are not bound to their transcription factors, BMAL1/CLOCK (Hong et al. 2009). This *in silico* study provides testable hypotheses for the detailed mechanism of DNA damage–induced phase shifts of circadian rhythms, which is important in understanding why DNA damage signaling cascade influences the circadian clock. The implications of such an interaction are hypothesized that cell cycle machinery utilizes circadian rhythms to activate WEE1 by prematurely degrading mPERs, which subsequently releases negative feedback on BMAL1/CLOCK. The BMAL1/CLOCK-activated WEE1 will delay cell cycle progress upon DNA damage, and provide more time for DNA repair.

There are more clues of coupling between the cell cycle and the circadian clock. The transcript of *c-Myc* oscillates with a circadian period, and BMAL1/NPAS2 suppresses its transcription (Sahar and Sassone-Corsi 2009). Its molecular details, however, remain unexplored. Additional cell cycle components such as *Cyclin D1*, *Gadd45α*, and *Mdm-2* show rhythmic transcripts, but their detailed regulations are unknown (Sahar and Sassone-Corsi 2009). The rhythmic *Cyclin D1* transcripts suggest a role of the circadian clock at G1-to-S transition in addition to G2-to-M transition via WEE1. Furthermore, rhythmic *Gadd45α* and *Mdm-2* transcripts suggest potential roles of the circadian clock in DNA damage responses. It is possible that there are other molecular coupling components. These may be clock-controlled genes (CCGs) like *Wee1*, conditionally affecting the circadian clock like CHK2, or novel parts of the circadian clock. Bidirectional communications of these two oscillators require further investigations.

In contrast to previous data, a recent effort to study this coupling indicated that this connection might depend on cell types (Yeom et al. 2010). Yeom and colleagues reported the circadian-independent cell divisions in rat-1 fibroblasts. They showed that about 13-h cell division cycles at 37°C had no phase correlations with the 24-h circadian rhythms when they followed 3–4 generations of cell divisions with bioluminescence assay. Their data, however, also indicate variability in cell cycle lengths, which may reflect quantized cell cycles as shown in cyanobacteria studies (Yang et al. 2010). For accurate conclusions, larger data set of cell cycle lengths is necessary for careful study of coupling between the cell cycle and the circadian clock.

Cross-References

- ▶ Cell Cycle
- ▶ Cell Cycle Transitions, G2/M
- ▶ Circadian Rhythm

References

Csikasz-Nagy A, Battogtokh D, Chen KC, Novak B, Tyson JJ (2006) Analysis of a generic model of eukaryotic cell-cycle regulation. Biophys J 90:4361–4379

Dunlap JC (1999) Molecular bases for circadian clocks. Cell 96:271–290

Hong CI, Zamborszky J, Csikasz-Nagy A (2009) Minimum criteria for DNA damage-induced phase advances in circadian rhythms. PLoS Comput Biol 5(5):e1000384

Nurse P (2000) A long twentieth century of the cell cycle and beyond. Cell 100:71–78

Sahar S, Sassone-Corsi P (2009) Metabolism and cancer: the circadian clock connection. Nat Rev Cancer 9(12):886–896

Sweeney BM, Hastings JW (1958) Rhythmic cell division in populations of Gonyaulax polyedra. J Protozool 5:217–224

Ukai H, Ueda HR (2010) Systems biology of mammalian circadian clock. Annu Rev Physiol 72:579–603

Yang Q, Pando BF, Dong G, Golden SS, van Oudenaarden A (2010) Circadian gating of the cell cycle revealed in single cyanobacterial cells. Science 327(5972):1522–1526

Yeom M, Pendergast JS, Ohmiya Y, Yamazaki S (2010) Circadian-independent cell mitosis in immortalized fibroblasts. Proc Natl Acad Sci USA 107(21):9665–9670

Zamborszky J, Hong CI, Csikasz-Nagy A (2007) Computational analysis of mammalian cell division gated by a circadian clock: quantized cell cycles and cell size control. J Biol Rhythms 22(6):542–553

Cell Cycle, Fission Yeast

Ákos Sveiczer and Anna Horváth
Department of Applied Biotechnology and Food Science, Budapest University of Technology and Economics, Budapest, Hungary

Synonyms

Cell division cycle; Mitotic cycle; Schizosaccharomyces pombe

Definition

Fission yeast is an attractive unicellular model organism in eukaryotic cell cycle research from several aspects (morphogenesis, size regulation, checkpoints, phase transitions, chromosome structure and dynamics, meiosis, mathematical modeling, etc.). Fission yeasts form four species in the genus *Schizosaccharomyces* (Ascomycota, Archiascomycetes), among which the best known one is *S. pombe*.

Characteristics

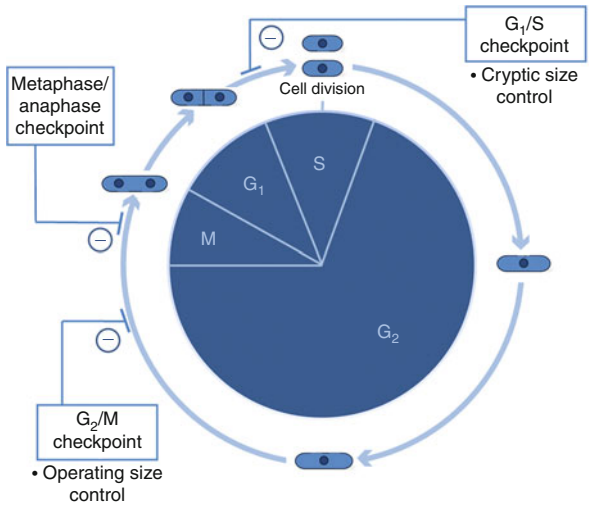
Cell Biology and Physiology of the Fission Yeast Cell Cycle

The fission yeast *Schizosaccharomyces pombe* has very frequently been applied since the 1950s as a model organism in cell biology, microbial physiology (▶ Cell Cycle, Physiology), and genetics, and later also in molecular biology, genomics, and systems biology (Mitchison 1990; MacNeill 2002; Sveiczer et al. 2004). As a species, *S. pombe* belongs to class Schizosaccharomycetes, subphylum Archiascomycotina, and phylum Ascomycota. The genus *Schizosaccharomyces* consists of four species, namely *S. pombe*, *S. japonicus*, *S. octosporus*, and *S. cryophilus*. All are considered to be fission yeasts; however, *S. pombe* was the first one to be discovered and it is also the most known and important. Fission yeast is a haploid unicellular eukaryote with rod-shaped cells of almost constant diameter (~3.5 μm). Cell length at birth is ~7–8 μm, which extends to

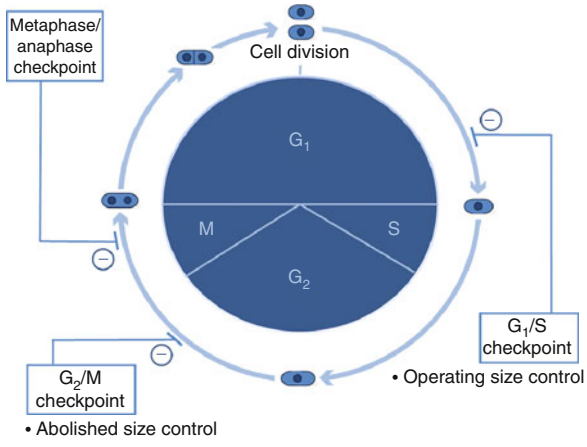
~13–14 μm till the end of the mitotic cell cycle by tip growth (MacNeill and Fantes 1995). Although starvation blocks cell proliferation and induces either sexual differentiation (mating, meiosis, and sporulation) or dormancy, all these processes are not discussed in this essay.

Fission yeast divides symmetrically, producing two nearly identical progenies. As the sister cells behave similarly in the next cycle, artificially generated synchronicity in a population remains sufficient for several generations. Moreover, cell length correlates with age; and all these characteristics make *S. pombe* an attractive model organism in cell cycle research. The generation time of its populations is generally ~2–4 h, depending on temperature and the media used for growth. The fission yeast cell cycle (Fig. 1) contains short G1, S, and M phases, while G2 is very long (~70% of total cycle time), because size control operates in G2 in wild-type cells (Mitchison 1989). After ▶ mitosis, the cells cannot divide immediately, because septation requires ~30–40 min. During this septation period the cells have two nuclei, which are mainly in their (next) G1 phase. However, as G1 is very short, the cells are actually replicating their DNA (▶ DNA Replication) parallel with ▶ cytokinesis.

The first cell division cycle (*cdc*) mutants of fission yeast were isolated in the 1970s by classic genetic methods. Loss-of-function point mutants were generated, which could grow normally at the permissive temperature (25°C), but stopped proliferation at the restrictive temperature (35°C). The terminal phenotype was usually a highly elongated cell, blocked at a specific event in the cell cycle, called the transition point of the mutated gene. The position of the transition point discriminated the 25 identified *cdc* genes as G1/S, S, G2, G2/M, and septation genes. The temperature-sensitive mutants also enabled a good synchronization technique; a simple temperature block-and-release experiment produced induction synchrony in the population (MacNeill and Fantes 1995). Among these *cdc* mutants, some grew normally, but their size were about half that of wild type cells. Isolated in Edinburgh (Scotland), they were soon renamed *wee* mutants, since wee means small in ancient Scottish language. Later it became clear that mutations in only two genes (*wee1* and *wee2*) could cause wee phenotype; however, *wee1*$^-$ mutants were found to be loss-of-function point mutants of the *wee1* gene, while *wee2*$^-$ mutants were found to be

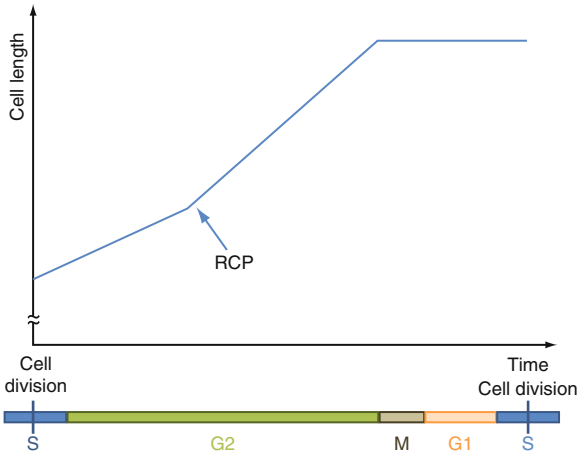


Cell Cycle, Fission Yeast, Fig. 1 The general features of the fission yeast cell cycle (wild type)



Cell Cycle, Fission Yeast, Fig. 2 The general features of the fission yeast cell cycle (*wee1*$^-$ mutant)

gain-of-function point mutants of the *cdc2* gene (the *wee2* and *cdc2* genes were identical, and named *cdc2* afterward). Analyzing the cell cycle of *wee1* mutants revealed that G1 phase was much extended at the expense of G2 (Fig. 2) (Mitchison 1989). A general conclusion was drawn that fission yeast has two size control mechanisms during the mitotic cycle (▶ Cell Cycle, Cell Size Regulation). In wild type cells the G2 size checkpoint operates, while the G1 one is cryptic because of large cell size. By contrast, in *wee* mutants the G2 size control is abolished, cell length decreases,



Cell Cycle, Fission Yeast, Fig. 3 Length growth profile during the cell cycle of fission yeast

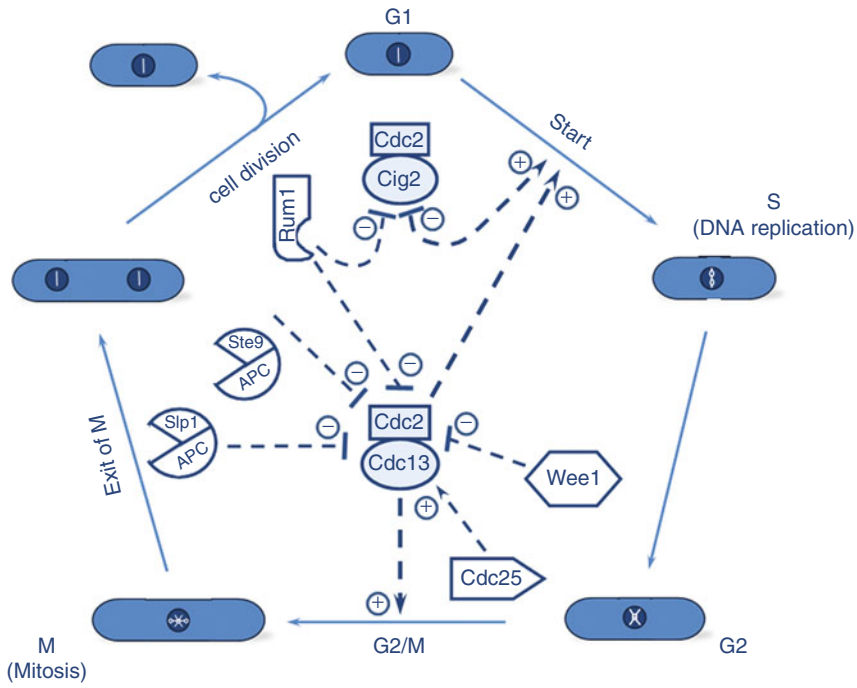
and the G1 mechanism maintains size homeostasis at a reduced level (Fig. 1, Fig. 2).

The time profile of cellular growth during the fission yeast cell cycle has also been an extensively studied question for several decades. Although debates still arise, a more or less generally accepted view is that length growth starts at birth and lasts until about mitotic onset, when the cell starts to prepare for cytokinesis. So, the last ~25% of the cycle without visible growth is called the constant volume phase, meanwhile the first ~75% is the growth phase. The latter one is made up of two linear segments, separated by a rate change point (RCP) (Fig. 3), i.e., length growth in fission yeast is not exponential (Mitchison and Nurse 1985). The reason of this RCP is also controversial. In wild type cells, tip growth is unipolar at the beginning of the cycle and it is bipolar later, therefore changing the growth manner in mid G2 phase was thought to cause a change in growth rate. However, such a general conclusion has been challenged when different cell cycle mutants were also studied.

Molecular Biology of the Fission Yeast Cell Cycle

In the 1980s and 1990s novel sophisticated methods made possible that geneticists could generate several novel types of fission yeast mutants, for example, gene deletion, gene over-expressing, and multiple mutants. At that time, molecular biologists also discovered the biochemical functions of the most important genes and proteins. It became clear that these proteins formed a regulatory network, named the cell cycle engine or control system, which ensured that discrete events of the cycle followed each other in the correct order. Moreover, this control network was found to be evolutionarily conserved (Morgan 2007). Fission yeast was a crucial model organism in discovering the engine's mechanism, because its regulatory network was probably the simplest among the known recent eukaryotes, and even simple mutations easily led to characteristic phenotypes in this haploid species. For example, deleting the mitotic cyclin gene *cdc13* caused ▶ endoreplication cycles (repeated S phases without intervening M phases) (Hayles et al. 1994); meanwhile deleting the replication license factor gene *cdc18* led to a mitotic catastrophe (a lethal execution of mitotic and cytokinetic events from the haploid G1 state) (Nishitani and Nurse 1995). Furthermore, deleting the mitotic activator phosphatase gene *cdc25* in the temperature sensitive *wee1–50* background generated quantized cycles at the restrictive temperature (Sveiczer et al. 1996). By the end of the second millennium, a "wiring diagram" for the fission yeast cell cycle engine was made (Fig. 4), which was based on the above mentioned experiments and could be studied by methods of mathematical modeling (▶ Cell Cycle Modeling, Differential Equation) (Sveiczer et al. 2004). Since sequencing of the whole *S. pombe* genome has been finished (MacNeill 2002), these models probably contained the major players of the network; however, novelties in the regulatory mechanisms still arise. Below we briefly summarize the conception on how this wiring diagram drives fission yeast cells through the cycle, from birth to the next division.

In the center of Fig. 4 is the Cdc2/Cdc13 dimer protein complex (Sveiczer et al. 2004). Cdc2 is the regulatory kinase (cyclin-dependent kinase or Cdk) subunit, which phosphorylates its specific substrate proteins at specific Ser or Thr residues (▶ Cyclins and Cyclin-dependent Kinases). Cdc13 is the regulatory (cyclin) subunit of the dimer, having no enzymatic activity by itself, but its presence is essential for the Cdc2 kinase activity. In S and early G2 phases, the dimer forms even at small cell size, however, it is still inactive, because the Tyr-15 residue of the Cdc2 subunit is phosphorylated by the Wee1 kinase, a mitotic inhibitor. In late G2 when the cell is large enough, by contrast, this inhibitory phosphate group is removed by the Cdc25 phosphoprotein phosphatase, a mitotic activator (MacNeill and Fantes 1995).



Cell Cycle, Fission Yeast, Fig. 4 The cell cycle control system of fission yeast (based on Sveiczer et al. 2004)

The Cdc2/Cdc13 dimer gets fully activated, and its functions help condense chromosomes and remodel microtubules in fission yeast cells. As a consequence, the cells pass the G2/M transition (▶ Cell Cycle Transitions, G2/M), therefore this Cdk/cyclin is also known as the M-phase promoting factor, or MPF. In anaphase the MPF activity abruptly starts to decrease, as the anaphase promoting complex (APC), with the help of its auxiliary subunit Slp1, ubiquitinates Cdc13 cyclin, marking it for degradation by the proteasome (Morgan 2007). Losing Cdk activity leads to exit of M phase, i.e., the cell gets into the next G1 phase in a binucleated state (▶ Cell Cycle Transitions, Mitotic Exit). During early G1, MPF activity is very low for two reasons. First, APC (with the help of another auxiliary protein Ste9) continuously ubiquitinates newly synthesized Cdc13 proteins. Second, Cdc2/Cdc13 dimers are bound to a stoichiometric inhibitor of Cdk, the Rum1 protein (▶ Cdk Inhibitors). In late G1, another cyclin (Cig2) is formed, also making dimers with Cdc2. This complex is called the starter kinase, because it helps the cell pass through the G1/S (or Start) transition, albeit with the help of the remaining MPF activity. (Note that Cdc13 is the only known essential cyclin in fission yeast, i.e., the onset of DNA replication does not require Cig2.) This transition is under size control in small *wee* mutants, but it is cryptic in wild type cells (see above). The length of G1 influences when cell division occurs (Mitchison 1989): either in G1 (*wee* mutant) or rather in S phase (wild type) (for simplicity, Fig. 4 shows a case, where cytokinesis comes about in G1). As far as DNA replication starts, the cell returns to the very point from which our cell cycle journey started in the previous generation.

Acknowledgment Research in our group is supported by the Hungarian Scientific Research Fund (OTKA K-76229) and by the New Széchenyi Plan (Project ID: TÁMOP-4.2.1/B-09/1/KMR-2010-0002).

Cross-References

▶ CDK Inhibitors
▶ Cell Cycle Modeling, Differential Equation

- ▶ Cell Cycle Transitions, G2/M
- ▶ Cell Cycle Transitions, Mitotic Exit
- ▶ Cell Cycle, Cell Size Regulation
- ▶ Cell Cycle, Physiology
- ▶ Cyclins and Cyclin-dependent Kinases
- ▶ Cytokinesis
- ▶ DNA Replication
- ▶ Endoreplication
- ▶ Mitosis

References

Hayles J, Fisher D, Woolard A, Nurse P (1994) Temporal order of S-phase and mitosis in fission yeast is determined by the state of the $p34^{cdc2}$/mitotic B cyclin complex. Cell 78:813–822

MacNeill SA (2002) Genome sequencing: and then there were six. Curr Biol 12:R294–R296

MacNeill SA, Fantes PA (1995) Controlling entry into mitosis in fission yeast. In: Hutchison C, Glover DM (eds) Cell cycle control. Oxford University Press, Oxford, pp 63–105

Mitchison JM (1989) Cell cycle growth and periodicities. In: Nasim A, Young P, Johnson BF (eds) Molecular biology of the fission yeast. Academic, New York, pp 205–242

Mitchison JM (1990) The fission yeast, *Schizosaccharomyces pombe*. Bioessays 12:189–191

Mitchison JM, Nurse P (1985) Growth in cell length in the fission yeast *Schizosaccharomyces pombe*. J Cell Sci 75:357–376

Morgan DO (2007) The cell cycle. Principles of control. New Science, London

Nishitani H, Nurse P (1995) $p65^{cdc18}$ plays a major role controlling the initiation of DNA replication in fission yeast. Cell 83:397–405

Sveiczer A, Novak B, Mitchison JM (1996) The size control of fission yeast revisited. J Cell Sci 109:2947–2957

Sveiczer A, Tyson JJ, Novak B (2004) Modelling the fission yeast cell cycle. Brief Funct Genom Proteom 2:298–307

Cell Cycle, Physiology

Chris J. Norbury
Sir William Dunn School of Pathology,
University of Oxford, Oxford, UK

Definition

The physiology of the ▶ cell cycle describes the processes that underlie and regulate the cycle under normal physiological (as opposed to pathological) conditions. Chromosomal replication (▶ DNA Replication) and ordered segregation of the replicated chromosomes (in eukaryotes by ▶ mitosis or ▶ meiosis) are fundamental aspects of cell cycle physiology that are common to all proliferating cells. Other events such as cellular growth and division (▶ Cytokinesis) are features of many cell cycles, but are not universal. Each of these processes can be regulated by physiological cues from within and outside the cell. Elucidation of cell cycle physiology at the systems level therefore requires both an understanding of the universal chromosomal cycle and consideration of the ways in which other processes are coordinated with these chromosomal events.

Characteristics

Coordination of cell cycle events with other biological processes such as macromolecular synthesis, cell growth, and differentiation is a key aspect of the physiology of all organisms (Alberts et al. 2010; Morgan 2007). This is particularly evident in multicellular organisms, where the development and maintenance of complex structures, tissues, and organs are only possible through timely and highly regulated cell proliferation, differentiation, and, in many cases, ▶ apoptosis. Understanding biological systems of this complexity constitutes a monumental challenge. There is therefore much to be gained by studying comparatively simple and genetically well-defined multicellular organisms such as *Drosophila melanogaster* (▶ Model Organism), by studying the cell cycles of early cleaving embryos, where there is no appreciable cell growth or differentiation (▶ Cell Cycle of Early Frog Embryos), and by studying cell cycle regulation in unicellular models (▶ Cell Cycle, Archaea; ▶ Cell Cycle, Budding Yeast; ▶ Cell Cycle, Fission Yeast; ▶ Cell Cycle, Prokaryotes), including cell lines derived from multicellular organisms (▶ Cell Cycle of Mammalian Cells).

Consideration of what constitutes "normal" physiological conditions can be an important aspect of cell cycle studies. While the conditions experienced by a cell within a multicellular organism can often be assumed to be physiologically relevant, it is usually

difficult to reproduce an authentic "niche" for any given cell type grown in cell culture. For example, proliferation and maintenance of mammalian stem cell cultures depends on engagement of specific combinations of cell surface receptor proteins that, in the authentic stem cell niche, would interact with components of the extracellular matrix, soluble ▶ mitogens and/or neighboring cells. Cell culture studies using defined medium and purified mitogens have allowed the dissection of pathways linking these signaling processes to cell cycle progression, most notably for mammalian fibroblasts (▶ Cell Cycle Transition, Detailed Regulation of Restriction Point). This type of study has given rise to the widely held view that non-proliferating cells are physiologically arrested in a protracted G1-like state (often referred to as G0, though the latter may be identical to G1 at the biochemical level). It is nonetheless important to bear in mind that, under normal physiological conditions, numerous non-proliferating cell types are arrested after ▶ DNA replication, in a G2-like state.

Another important aspect of mammalian cell culture is that standard growth media are routinely supplemented with serum mitogens, often at high levels that would not be encountered under physiological conditions in the tissue of origin. For many microbes, unlike mammalian cells, proliferation in culture is limited only by nutrient availability. Similar considerations apply here, as the nutrient-rich environment experienced in the laboratory may bear little relation to that in which the microbe grows in the wild. Cell kinetic, biochemical and molecular biological parameters measured under non-physiological conditions may well give a misleading picture of cell cycle regulation at the systems level.

Cross-References

- ▶ Apoptosis
- ▶ Cell Cycle
- ▶ Cell Cycle of Early Frog Embryos
- ▶ Cell Cycle of Mammalian Cells
- ▶ Cell Cycle, Archaea
- ▶ Cell Cycle, Budding Yeast
- ▶ Cell Cycle, Fission Yeast
- ▶ Cell Cycle, Prokaryotes
- ▶ Cytokinesis
- ▶ DNA Replication
- ▶ Meiosis
- ▶ Mitogens
- ▶ Mitosis
- ▶ Model Organism

References

Alberts B, Bray D, Hopkin K, Johnson A, Lewis J, Raff M, Roberts K, Walter P (2010) Essential cell biology. Garland Science, New York

Morgan DO (2007) The cell cycle. Oxford University Press, Oxford

Cell Cycle, Prokaryotes

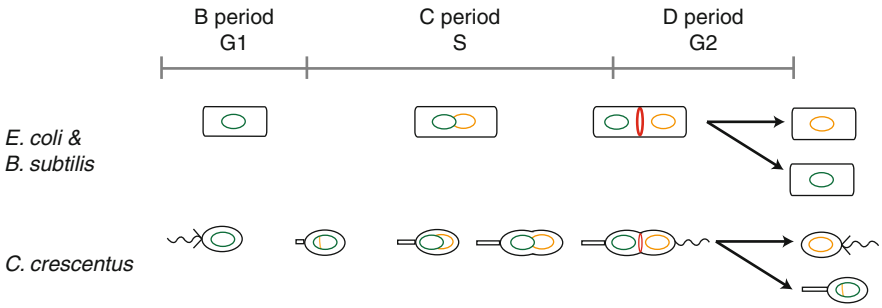
Emanuele G. Biondi
CR1-CNRS-Laboratory of Biochemistry and Integrated Structural Biology, Institut de Recherche Interdisciplinaire - IRI CNRS USR3078, Villeneuve d'Ascq, France

Synonyms

Bacterial cell cycle; Cell cycle progression in bacteria

Definition

The bacterial cell cycle is defined as a series of events that ensure the coordination of three processes in order to duplicate a cell: DNA replication, cell growth, and cell division. The cycle is composed by an initial G1 phase followed by the initiation of DNA replication (S phase), which is usually paralleled by cell growth; then, when replication is completed, chromosomes are positioned to opposite compartments of the predivisional cell (G2 phase); eventually cell division takes place by the polymerization of specific proteins that create a septum, followed by daughter cell separation. In several bacterial species, the two cells can be functionally and morphologically different, such as a spore of *Bacillus subtilis* during the sporulation process or the swarmer cell formation during the cell cycle in *Caulobacter crescentus* (see the section "Characteristics" below).



Cell Cycle, Prokaryotes, Fig. 1 Morphological transitions of cells during the cell cycle in *E. coli* and *B. subtilis* (*upper part*) and *C. crescentus* (*lower part*). Chromosomes are symbolized as yellow and green circles. Cell division machinery is the red circle in the midcell of both models

Characteristics

Bacteria have evolved different systems for the regulation of cell cycle, probably due to different ecological and evolutionary constraints. Differently from eukaryotes, bacteria have no nucleus and usually possess a single circular chromosome which replicates starting from a single origin of replication. Initiation of DNA replication and cell division are processes conserved in all bacteria as illustrated in the next section.

Cell Cycle in Bacteria: *Escherichia coli* and *B. subtilis*

In *E. coli* and *B. subtilis*, cell cycle can be divided (as illustrated in Fig. 1) in three phases (Wang and Levin 2009): (1) the B Period (Birth) between a cell division and the initiation of DNA replication (G1); (2) the C phase, when the chromosome is replicated (S); and (3) the D period, between the end of replication of DNA and cell division (G2).

The doubling time of the cell in these species (defined as the time required to double the mass of the cell) depends on temperature and nutrient availability; more nutrients and optimal growth temperature speed up the doubling time. However, the periods C and D are always constant. In particular conditions of high nutrient concentration, the growth rate is faster than doubling time. In a "slow" cell cycle, replication of DNA takes place from a single origin of replication, named *oriC*. From this single origin, DNA replication proceeds in two directions, ending at the opposite region. In the "fast" replication rate, a new round of replication from *oriC* starts before the previous one has terminated and this special case is called "multifork" replication.

The initiation of DNA replication depends on the activity of the AAA+ family of ATPases member, DnaA, which binds at a specific site of *oriC*, causing an AT-rich region to open and allowing the replisome to start its DNA polymerization activity (Zakrzewska-Czerwińska et al. 2007).

Several mechanisms prevent new rounds of replication right after the replisome has started: in *E. coli*, during DNA elongation, Hda converts the active DnaA-ATP to the less active DnaA-ADP, while in *B. subtilis* YabA blocks DnaA activity by chaining DnaA to DnaN (the beta clamp, which ensures DNA polymerase processivity).

When DNA replication is terminated and the condensed chromosomes (nucleoids) are segregated in opposite compartments of the predivisional cell, bacterial cells need to position the cell division apparatus in order to generate cells possessing a sufficient mass and a complete chromosomal content. Those parameters are monitored by different mechanisms that can vary across bacteria and a clear model has not been elucidated yet.

The initial step of cell division is coordinated by FtsZ, a tubulin-like GTPase, which is able to polymerase, using GTP. FtsZ is able to form a ring-like structure (Z-ring) at midcell and this circular structure serves as scaffold for the formation of a septum, able to constrict at the end of cell cycle. The position of the FtsZ polymers and timing of septum formation depends on different parameters such as glucose availability in *B. subtilis* (by the activity of the UgtP effector protein that senses the nutrient status and is able to inhibit FtsZ polymerization) but largely on DNA replication and chromosome segregation.

In fact, cell division depends on the position of the two *oriC*s which are able to inhibit septum formation; several proteins such as MinC and MinD, which are localized at the poles, together with the DNA-associated proteins SlmA in *E. coli* and Noc in *B. subtilis* are ensuring that cell division can take place only in the midcell (Margolin 2006).

Cell Cycle Regulation in *C. crescentus*

C. crescentus, a gram negative bacterium of the alpha subdivision of proteobacteria, has recently become a model organism for bacterial cell cycle studies (Ryan and Shapiro 2003). In this organism each cell division is asymmetric producing always two functionally and morphologically different cells, the replicating "stalked" cell type and the vegetative "swarmer" type (Fig. 1). After each initiation of DNA replication, the replication fork is kept blocked so that the *Caulobacter* cell cycle can follow a pattern of once-and-only-once replication per division (G1, S, and G2 phases are temporally distinguished).

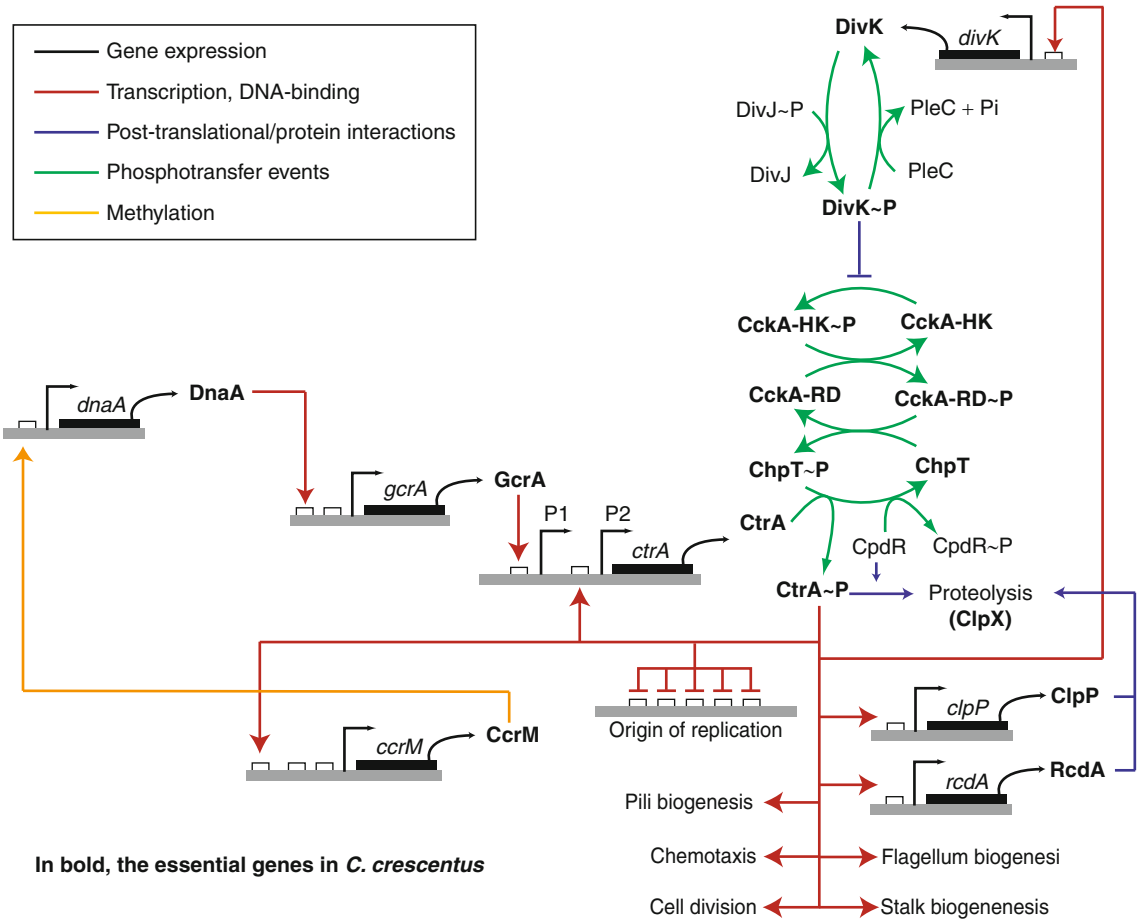
Many factors are known to regulate cell cycle progression and most of them are members of the family of *two-component system* proteins, comprised of histidine kinases, histidine phosphotransferases, and their response regulator substrates. Among those proteins CtrA is the master regulator of the *Caulobacter* cell cycle; it is an essential response regulator whose activity as a transcription factor varies as a function of the cell cycle. CtrA controls various functions during cell cycle progression by activating or repressing expression of genes involved in cell division (*ftsZ*), flagellum biogenesis genes, stalk biogenesis regulatory genes, pili biogenesis genes, and chemotaxis genes. CtrA also blocks the initiation of DNA replication through binding of the replication origin, *oriC*.

CtrA activity and stability varies during the cell cycle. Oscillation of CtrA levels, peaking at the predivisional stage before cell division, is achieved by different mechanisms: transcription, proteolysis, and phosphorylation control. The result of this oscillation is that CtrA is present and phosphorylated in swarmer cells, then is degraded at the G1/S transition and then transcribed again and phosphorylated in the predivisional stage. Eventually, at cell division, CtrA ends up in the swarmer compartments and it is absent in the stalked cell that can immediately starts a new round of replication.

All cell cycle regulation factors are schematized in Fig. 2 where we illustrate the multilevel regulation of the *Caulobacter* cell cycle. Two main oscillators are working during cell cycle progression: (1) the transcriptional and epigenetic circuit (CtrA-DnaA-GcrA-CcrM); (2) the phosphorylation/proteolysis and transcription circuit (CckA-CtrA-DivK). The latter also involves coordination of CtrA proteolysis through the regulation of DivK activity.

Transcription of *ctrA* is controlled by circuit composed by *DnaA* and GcrA, and the DNA methyltransferase CcrM. DnaA, which is also involved in initiation of DNA replication, as discussed for *E. coli* and *B. subtilis*, is a key element in cell cycle regulation in *C. crescentus* and it also controls the transcription of about 40 genes involved in nucleotide biogenesis, cell division, and polar morphogenesis. DnaA also activates the transcription of the *gcrA* gene. GcrA controls the *de novo* transcription of *ctrA* after proteolysis and genes involved in DNA metabolism and chromosome segregation, including those encoding DNA gyrase, DNA helicase, DNA primase, and DNA polymerase III. The transcriptional loop of *ctrA* is closed by CcrM. In fact, CtrA activates the transcription of *ccrM*, which encodes for a DNA methyltransferase whose abundance is cell cycle dependent. CcrM is able to activate the *dnaA* promoter region through methylation, closing the positive feedback composed by CtrA, DnaA, and GcrA.

CtrA must be phosphorylated to bind DNA and its phosphorylation depends on cell cycle progression. An essential phosphorelay, composed by the hybrid histidine kinase CckA and the histidine phosphotransferase ChpT, is responsible for CtrA phosphorylation (Biondi et al. 2006). The membrane hybrid histidine kinase CckA is dynamically localized during cell cycle progression acting as a kinase for CtrA when it is localized at the poles, while it acts as a phosphatase in the delocalized form. DivK, which is a single domain response regulator that controls CckA localization, plays an essential role as a positive regulator of cell cycle progression because, when phosphorylated, it indirectly delocalizes CckA



Cell Cycle, Prokaryotes, Fig. 2 The circuit controlling cell cycle progression in *C. crescentus*. Colors of arrows have different colors depending on their interaction type (as indicated in the legend). See text for details

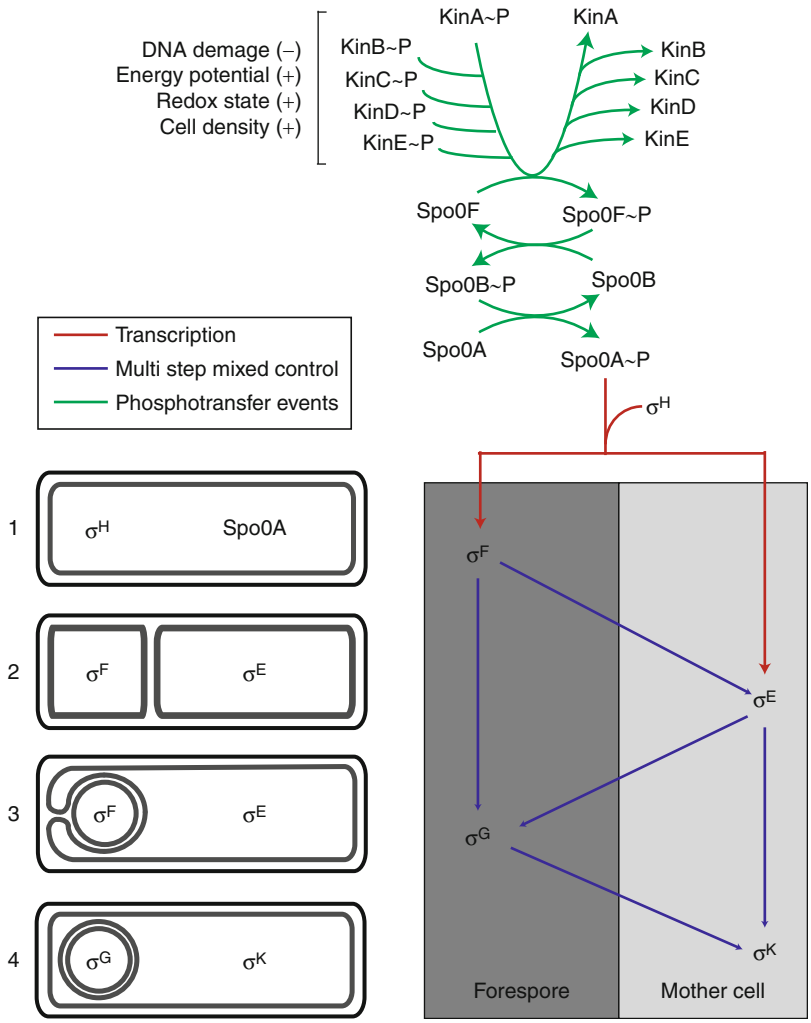
repressing CtrA and thus promotes DNA replication. Two membrane histidine kinases are known to interact with DivK: PleC and DivJ. PleC and DivJ are considered, respectively, the principal phosphatase and kinase of DivK and they are in opposite locations before cell division.

ChpT also transfers the phosphate to a second response regulator named CpdR, which, together with RcdA, is a factor required for CtrA proteolysis mediated by the ClpP-ClpX protease. CtrA is degraded at the stalked pole at the G1/S transition when the origin of replication needs to be cleared and also in the stalked compartment, where initiation of DNA replication occurs immediately after cell division.

Sporulation in *B. subtilis*

B. subtilis, as discussed before, has a regulation of cell cycle progression that shares many features with *E. coli*. However, *B. subtilis* is able in response to starvation to form resistance cell types called spores. This process, called sporulation (Fig. 1), represents a special case of asymmetric cell division and it is discussed separately in this section.

Cells of *B. subtilis* can divide centrally in nutrient rich media leading to identical progeny but, in a specific cell density and nutritional signals conditions, cells can sporulate. A complex phosphorelay system composed by *two-component systems* proteins is controlling the



Cell Cycle, Prokaryotes, Fig. 3 The sporulation in *B. subtilis*. In the upper part, the phosphorelay controlling the activation of Spo0A together with known signals activating (+) or repressing (−) the phosphorylation are shown. After Spo0A activation, stages of spore formation are indicated on the left bottom part of the figure; on the right the compartmentalized cross-activation of Sigma factors is indicated. Arrows have different colors depending on their interaction type (as indicated in the legend). See text for details

phosphorylation of the master regulator of sporulation, Spo0A. Activated Spo0A leads to a polar septation that initially produce a larger mother cell, which eventually lised, and a smaller forespore surrounded by special envelops (Piggot and Hilbert 2004).

In *B. subtilis*, five histidine kinases (KinA, the principal kinase, to E) are controlling the phosphorylation state of Spo0A, through the phosphorylation of Spo0F, which then transfers to the histidine phosphotransferase Spo0B and eventually to the response regulator Spo0A as illustrated in Fig. 3.

Spo0A, which integrates several environmental signals using its kinases sensing apparatus, is then able to start a series of steps leading to the formation of a spore and the decay of the mother cell. This multistep process is carried on by different sigma factors acting in the mother cell or in the forespore. Each sigma factor is indeed able to activate via many other sporulation factors the following sigma factor in a different compartment.

Activated Spo0A and Sigma H are present and functional in the step right before the asymmetric

septation and they together activate many genes including sigma F and sigma E precursors, respectively, acting in the forespore and in the mother cell compartment. In order to have a fully functional Sigma E, Sigma F is required. At this step, Sigma E activates other factors in order to have an active Sigma G in the forespore compartment, which is eventually required to activate Sigma K in the mother cell. Those last two factors are required for full maturation of the spore, which is characterized by a special envelope that is responsible for its special resistance to stresses.

References

Biondi EG, Reisinger SJ, Skerker JM, Arif M, Perchuk BS, Ryan KR, Laub MT (2006) Regulation of the bacterial cell cycle by an integrated genetic circuit. Nature 444(7121):899–904

Margolin W (2006) Bacterial division: another way to box in the ring. Curr Biol 16(20):R881–R884

Piggot PJ, Hilbert DW (2004) Sporulation of *Bacillus subtilis*. Curr Opin Microbiol 7(6):579–586

Ryan KR, Shapiro L (2003) Temporal and spatial regulation in prokaryotic cell cycle progression and development. Annu Rev Biochem 72:367–394

Wang JD, Levin PA (2009) Metabolism, cell growth and the bacterial cell cycle. Nat Rev Microbiol 7(11):822–827

Zakrzewska-Czerwińska J, Jakimowicz D, Zawilak-Pawlik A, Messer W (2007) Regulation of the initiation of chromosomal replication in bacteria. FEMS Microbiol Rev 31(4):378–387

Cell Cycle, Synchronization

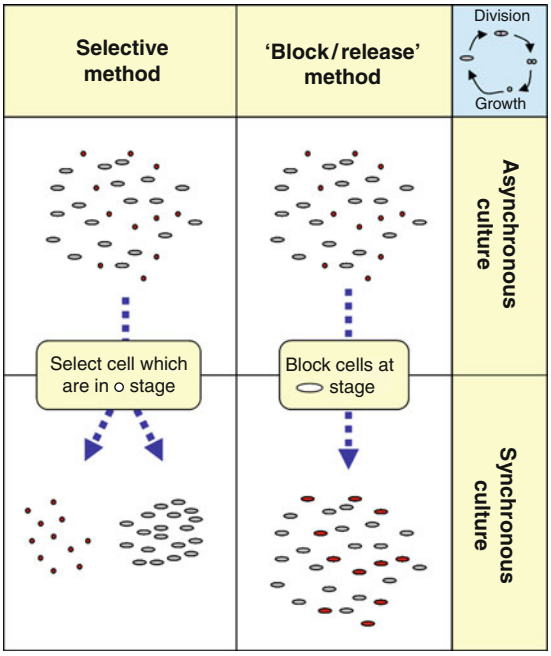
Jürg Bähler and Samuel Marguerat
Department of Genetics, Evolution and Environment and UCL Cancer Institute, University College London, London, UK

Synonyms

Cell synchronization

Definition

This term refers to experimental methods to obtain a population of cells that are in the same cell-cycle stage (Davis et al. 2001). Cell-cycle synchronization allows the analysis of biological processes that are specific to certain cell-cycle stages and that would otherwise be confounded or averaged out in asynchronous populations. Cell-cycle synchronization can be achieved in two general ways (Fig. 1). (1) Using "block and release" approaches where all cells are first arrested in a defined cell-cycle stage ("block"), and then allowed to cycle again in a synchronous manner ("release"). This approach involves adding/removing specific molecules to/from a culture or by using cells with conditional mutant alleles of cell-cycle regulators. (2) Isolating cells that are in a given stage of the cell cycle based on a phenotypic characteristic such as size (centrifugal elutriation) or age ("baby machine").



Cell Cycle, Synchronization, Fig. 1 Schematic representation of the two general classes of cell-cycle synchronization methods applied to an example cell type. The blue inset at the upper right corner describes the morphological changes of these cells during the cell cycle

Cross-References

► Cell Cycle Analysis, Expression Profiling

References

Davis PK, Ho A, Dowdy SF (2001) Biological methods for cell-cycle synchronization of mammalian cells. Biotechniques 30:1322–1326, 1328, 1330–1331

Cell Cycle-regulated Gene Expression

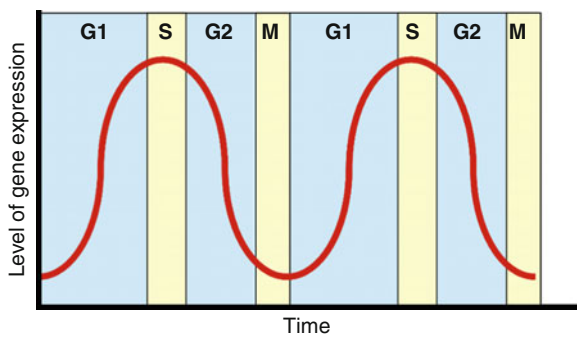
Jürg Bähler and Samuel Marguerat
Department of Genetics, Evolution and Environment and UCL Cancer Institute, University College London, London, UK

Synonyms

Cell cycle–regulated transcription; Periodic gene expression

Definition

This term describes a specific mode of gene regulation where transcript levels peak at a specific cell cycle phase, creating periodic profiles of gene expression across several cycles (Fig. 1). Periodic gene expression is not only found among cell cycle–regulated genes but also among genes regulated during other periodic processes such as the circadian cycle.



Cell Cycle-regulated Gene Expression, Fig. 1 Expression profile of a hypothetical gene with a peak in expression level in the early S-phase

Cross-References

► Cell Cycle Analysis, Expression Profiling

Cell Cycle-regulated Kinase Complexes

► Cyclins and Cyclin-dependent Kinases

Cell Cycle–regulated Transcription

► Cell Cycle-regulated Gene Expression

Cell Death

► Lymphocyte Dynamics and Repertoires, Biological Methods

Cell Division

► Cell Cycle of Mammalian Cells

Cell Division Cycle

► Cell Cycle, Biology
► Cell Cycle, Fission Yeast

Cell Division Database

► Cell Cycle Database

Cell Growth

► Cell Cycle of Mammalian Cells

Cell Labeling

▶ Lymphocyte Dynamics and Repertoires, Biological Methods

Cell Marker

▶ Fluorescent Markers

Cell Membrane Proteins

Mika O. Ruonala
NeuroToponomics Group, Center for Membrane Proteomics Campus Riedberg, Goethe University of Frankfurt, Frankfurt, Germany

Synonyms

Lipid-protein interactions; Transmembrane proteins

Definition

Cellular membranes have a thickness of 7–20 nm and efficiently restrict free passage of nonpolar molecules and cross-membrane communication. To overcome this problem, cells have established a massive membrane protein machinery that regulates most cellular activities ranging from signal transduction to energy balance, respiration, reproduction, and cellular morphology. Thus, the sole existence of multicellular organisms and cells rely on fully functional cell membrane protein system.

Characteristics

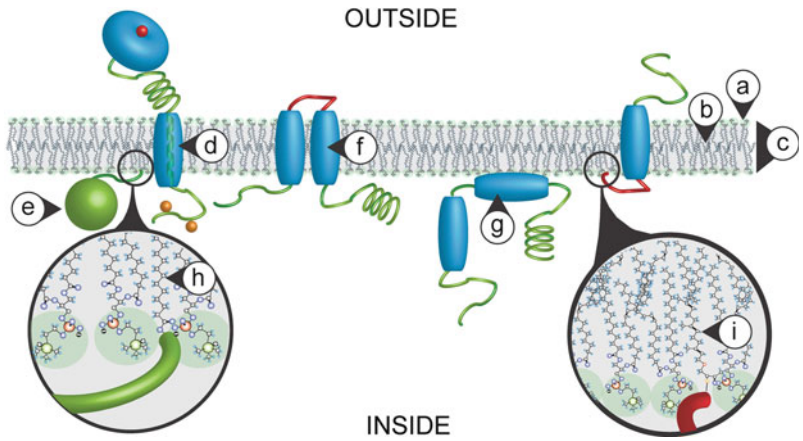
A cell and its individual internal compartments are surrounded by a membrane made of lipids. Lipids have a hydrophilic (water loving) and a hydrophobic (water repelling) fatty acid tail (a and b in Fig. 1, respectively). In an aqueous environment, lipids spontaneously self-organize to form lipid bilayers (Fig. 1c) where the hydrophobic parts are shielded from water between the two outer sheets of hydrophilic head groups. Like lipids, also some proteins have a hydrophobic nature arising from either the presence of a hydrophobic amino acid sequence along the protein amino acid sequence, or an attached lipid molecule. The need to protect hydrophobic parts from an aqueous environment is the major driving force for the membrane attachment of a protein, which is fulfilled by immersing these parts into the hydrophobic lipid bilayer (Fig. 1c).

The most common three-dimensional protein structure traversing the membrane is a $\alpha-helix$ (Fig. 1d): Formed by of a single peptide sequence with amino acids rotating along a common center line, the hydrophobic nonpolar side chains point outside and face the inner core of the lipid bilayer (Alberts 1994). A less frequent membrane-traversing 3D structure is a *β-strand*, assembled of two or more peptide sequences. The hydrophobic side chains of amino acids in a *β-strand* interact with equally hydrophobic head groups of other *β-strand* structures to form a larger membrane-traversing *β-sheet* structure.

Structure-based Classification of Membrane Proteins (Jennings 1989)

Basic classification of membrane proteins to *integral* and *peripheral* membrane proteins (Fig. 1d, e, respectively) follows the nature of membrane association. As the name implies, integral membrane proteins are an *integral* component of the host membrane and thus cannot be removed without simultaneously destroying the lipid membrane, for example, by using detergents. In contrast, peripheral membrane proteins *are not* an integral part of a host membrane, associate with it weakly, and can be separated from the membranes with relative ease.

Integral membrane proteins are categorized to either *bitopic* or *polytopic* integral membrane according to the number of membrane-traversing domains. A bitopic integral membrane protein contains a single membrane-spanning domain (Fig. 1d) whereas a polytopic membrane protein contains two or more membrane-embedded regions (Fig. 1f). Polytopic membrane proteins are finally classified either as *type I* or as *type II* according to the relative orientation of the first and last amino acids of the protein sequence (amino and carboxyl termini, respectively). In contrast to type I membrane protein,



Cell Membrane Proteins, Fig. 1 *Scheme of prominent features of a biological membrane:* Hydrophilic polar head group (*a*) and a hydrophobic oily fatty acid group (*b*) are oriented to form a lipid bilayer (*c*) that seals the fatty acid chains between two layers of polar head groups. A bitopic, type I integral membrane protein (*d*) has parts on both sides of the membrane. A polytopic, type II membrane protein (*f*) has more than one membrane domains, and both ends of the protein reside within the same volume separated by a membrane. Another peripheral membrane protein associated with the membrane via a partially hydrophobic polypeptide chain is shown in (*g*). A peripheral membrane protein (*e*) with attached lipid modification (*h*) is present only on the inner side of the membrane. A lipidated bitopic integral membrane protein is shown in (*i*)

where amino and carboxyl terminals are located at the opposing sides of the separating membrane (Fig. 1d), the peptide chain of type II starts and ends both on the same side of the membrane (Fig. 1f). The membrane incorporation of an integral membrane protein takes place already during protein biosynthesis in the lumen of the *endoplasmic reticulum* (ER) following the detection of a signal peptide by the translation machinery.

Rather than by a membrane-traversing peptide sequence, peripheral membrane proteins associate with the membranes through hydrophobic interactions between the membrane lipids and a hydrophobic polypeptide sequence (Fig. 1g) or via a lipid molecule (Fig. 1e). The attachment of a lipid group to a soluble protein can provide sufficient hydrophobicity to facilitate membrane attachment, and the reversible nature of some lipid modifications can act as a regulatory switch to concentrate a protein from soluble, cytosolic form to a membrane-bound form. A signal sequence, which is recognized by lipid transfer enzymes (such as palmitoyl-, myristoyl-, and prenyl transferases), leads to the covalent attachment of a palmitoyl or a myristoyl fatty acid, an isoprenyl fatty acid lipid group, or a glycosylphosphatidylinositol-(GPI) anchor attached to a cysteine or glycine amino acid. Timing of acylation and the position of the attached fatty acid group are the major differences between the various acylation forms. While myristoylation is an event where a myristoyl fatty acid group (Fig. 1h) is permanently attached exclusively to a glycine amino acid at the beginning of the newly formed polypeptide sequence protein (Fig. 1e), the attachment of a palmitate fatty acid group during palmitoylation to any cysteine residue of a finished protein is a reversible post-translational modification. Isoprenylation is yet another form of post-translational lipid modification where an isoprenyl lipid group, either a farnesyl or geranylgeranyl lipid, is attached to a cysteine amino acid resides at the end of the protein. In general, a lipidated membrane protein can attach itself either on the outer membrane of a cytosolic organelle, or on the intracellular leaflet of the plasma membrane, and thereby executes its function in the cytosolic volume of a cell. Yet another family of lipid-modified proteins is formed by glycosylphosphatidylinositol (GPI)-anchored proteins synthesized directly *on* a GPI-lipid residing on the ER membrane. GPI-anchored membrane proteins are transported to the outer surface of the plasma membrane and since their peptide sequence is exposed toward the extracellular space, they execute their function toward the extracellular matrix, i.e., outside the cells.

Provided that the required targeting signal is present within the peptide sequence, an integral membrane protein can additionally be lipidated in a regulated

manner, for example, with a palmitate fatty acid. Such a lipid molecule provides a membrane protein affinity toward specialized membrane regions, such as caveolae or other cholesterol- and sphingolipid-rich lipid microdomains (also termed as lipid rafts, Ikonen and Simons 1997). Thus, reversible lipidation may act as a switch in signaling efficiency by moving an inactive protein from the bulk membrane inside a specialized signaling platform. An example of such switchable integral membrane protein is the Neuronal Cell Adhesion Molecule (NCAM) that can not only be palmitoylated to intracellular cysteine residues by palmitoyl transferase, but also de-palmitoylated upon need. The location and function of NCAM critically depends on its lipidation status and subsequent residence in cholesterol-rich membrane regions.

Mixed Membrane Attachment Types

Hydrophobic amino acids can be used to build polypeptide chains with one-sided hydrophobicity. Upon orientation of the hydrophobic side of the protein toward a membrane, it can be protected from aqueous environment. This facilitates membrane association and results in a conformation where a part of the protein rather lays "flat" on the plane of the membrane (Fig. 1g) while other parts of the protein can freely move around and possibly also obtain a lipid modification (below). Presence of such a amphiphatic domain in a soluble protein can provide sufficient affinity for membrane attachment; alternatively, via such a domain an integral membrane protein can interact with certain lipid groups on the membrane, and possibly use it for targeting along the membrane plane to regions with specific properties.

Operation Modes of Membrane Proteins

Basically, two action modes exist for membrane proteins, and the type of membrane attachment is indicative for the functional mode. The activity of a monotopic membrane protein is restricted to one side of a membrane (*cis*-mode). For example, a lipidated kinase bound on the cytoplasmic side of the lipid bilayer only phosphorylates other cytoplasmic proteins (soluble or membrane bound). In contrast, a membrane-traversing bi- or polytopic membrane protein is likely to operate on both sides of the membrane in *trans*-mode. One concrete example are the family of receptor tyrosine kinases, e.g., the receptor for epidermal growth factor (EGFR). Here, the binding of epidermal growth factor (EGF) to the extracellular binding pocket of EGFR induces a conformational change to the cytoplasmic tail of the receptor, which then leads to the assembly of signaling complex and subsequent downstream events, such as remodelling of the cellular cytoskeleton. Yet another functional mode for membrane proteins is the selective transport of molecules across the membrane via a channel protein or protein complex. As the hydrophobic lipid bilayer is not readily passable for nonpolar solutes, their transport is mediated by a physical pore or a channel formed by one polytopic membrane protein, or as a complex of several bi- and/or polytopic membrane proteins. Within such a complex, the hydrophobic amino acid side chains orient themselves toward the lipid environment, and the hydrophobic side chains together form the passage. Traffic through the passage can be passive, i.e., the solutes flow from higher concentration toward lower concentration across the membrane; or active, where the transfer against a chemical gradient is coupled to the use of energy, such as ATP. So-called cotransporters couple the chemical concentration gradient to the transport of molecules against a gradient.

Membrane Proteins Move Rapidly in a Regulated Manner

Experimental work has shown that membrane proteins are in constant diffusional movement along the membrane plane and around their own axis. An integral membrane protein can not only freely rotate around itself at high speed but is also able to diffuse across a distance, corresponding to the periphery of a cell at several times per second. Equally, the proteinaceous part of a lipid-modified peripheral membrane protein can freely rotate around its lipid anchor and also along the membrane plane at high speed. However, in cellular plasma membranes, the lateral movement of integral and peripheral membrane proteins is restricted by specialized domains below and upon the membranes. An example of restricted movement of cell membrane proteins can be found from epithelial cell layer lining the digestive system. Here, the surrounding membrane of the epithelial cells is divided by a tight junction to basolateral and apical sides both having a unique membrane protein composition that does not mix with the other side.

Recent data shows that cellular cytoskeleton just beneath the cell plasma membrane forms "picket fences," or "pockets". Where the poles are membrane

proteins interacting with cytoskeletal components, such as actin fibers (Kusumi et al. 2011). The latter, in turn, form the boundaries of the pockets. Once inside such a pocket, a protein is free to move around in lateral direction; however, due to the restricting cytoskeleton, energy is required for an integral membrane protein to jump from one pocket to another one during a so-called hop-diffusion. Thus, the lateral movement along the membrane plane is reduced and controlled by the cellular cytoskeleton.

Another cellular system provides restriction toward lateral movement. Clusters of cholesterol and sphingolipids form lipid microdomains with specific properties, which are significantly different from the bulk of the membrane. Such "lipid rafts" attract both integral and lipid–modified peripheral proteins to form specialized signaling platforms. Interestingly, raft association can be stimulated and modulated upon post-translational attachment of specific lipid molecules, as described above for the NCAM cell adhesion molecule.

References

Alberts B (1994) Molecular biology of the cell, 3rd edn. Garland Publishing, New York

Ikonen E, Simons K (1997) Functional rafts in cell membranes. Nature 387:569

Jennings ML (1989) Topography of membrane proteins. Annu Rev Biochem 58:999

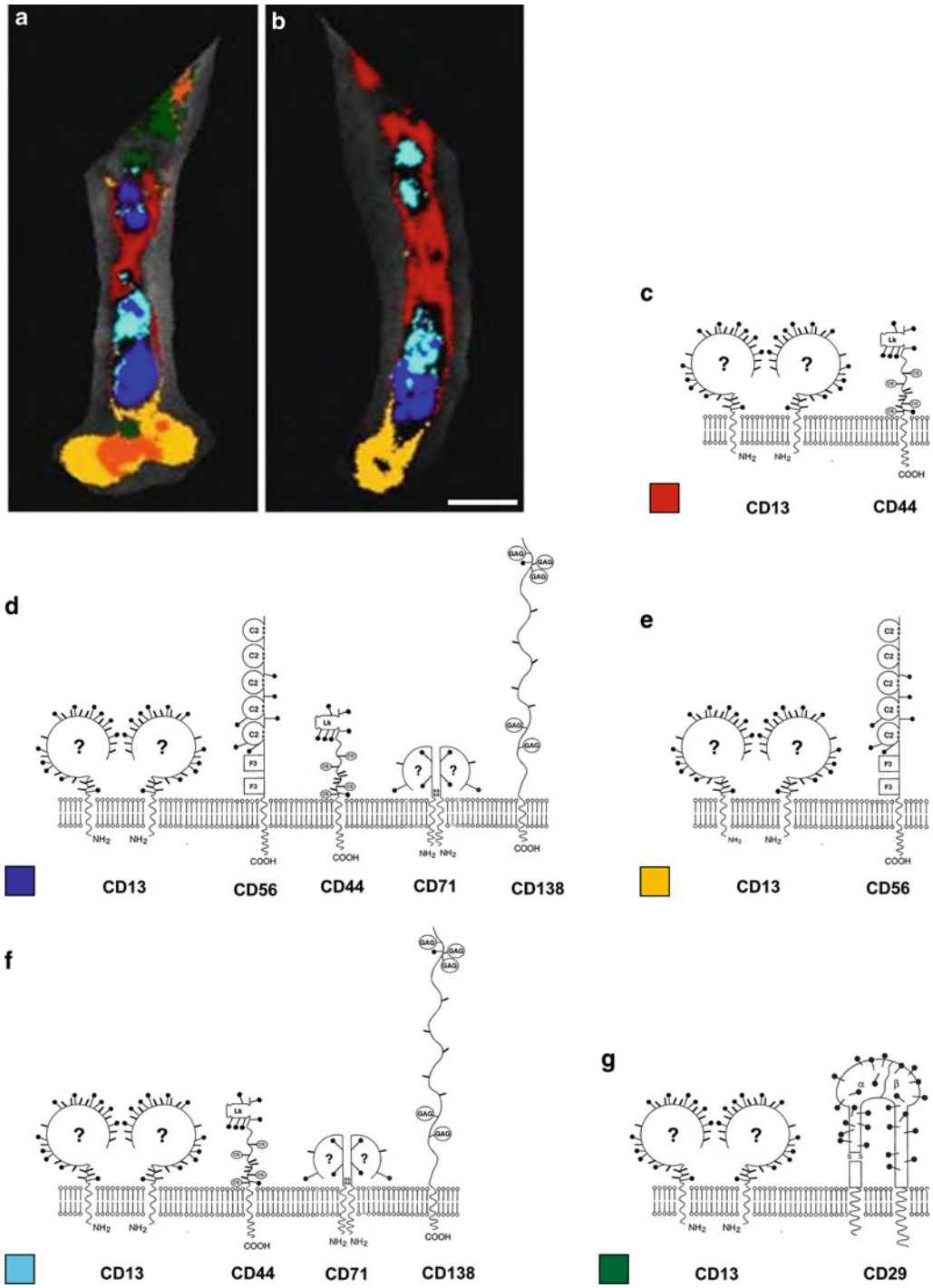
Kusumi A et al (2011) Hierarchical mesoscale domain organization of the plasma membrane. Trends Biochem Sci 36:604

Cell Membrane Toponomics

Anne Gieseler
Molecular Pattern Recognition Research (MPRR) Group, Otto-von-Guericke-University Magdeburg Medical Faculty, Magdeburg, Germany

Definition

Any cell is a collection of highly coordinated subcellular compartments formed by unit membranes. Unit membranes are highly organized as lipid bilayers with a hydrophilic external and internal part and a hydrophobic intermediate part (Lodish et al. 2000). This basic structure allows the cell to integrate an enormous number of distinct proteins in order to organize the myriad of different membrane functionalities (▶ Cell Membrane Proteins). These proteins serve a variety of essential functions, e.g., establishing and controlling cell adhesion, inducing intracellular signaling (by behaving as receptors), maintaining the ion homeostasis (by acting as ion channels), as well as controlling environmental adaptation. A fundamental biological question is how the approx. 8,000 distinct cell surface proteins are spatially organized as supramolecular functional units (toponome) in one and the same cell membrane, and how these units are altered in chronic diseases, such as cancer and Alzheimer's disease. This field of research is referred to as membrane toponomics. It is a subdiscipline in ▶ toponomics related to the functional detection of protein networks of cell membranes in morphologically intact cells or tissue sections by using TIS (▶ TIS Robot) microscopy. It has been shown that quantitative information on the subcellular topology and differential assembly of distinct proteins and their spatial relationships on the cell surface are straightforward to the detection of lead proteins hierarchically controlling cell function: For instance, the systematic mapping of cell surface proteins in rhabdomyosarcoma cells (Schubert 2010) led to understanding on how tumor cells establish multi-protein cluster networks composed of different classes of CD proteins required for cell polarization/migration and tumor morphogenesis (Fig. 1). The CD system is a collection of a variety of distinct molecular classes present on the surface of leucocytes and, in part, also on the surface of many other cell types (so-called *C*luster of *D*ifferentiation (CD) antigens; Zola 2001; Barclay et al. 1995). CD antigens are held to belong to the best characterized molecules of the cell surface membrane. An example of co-mapping of many CD proteins has revealed a tumor cell specific surface toponome (Fig. 1): Underlying experiments have shown that blocking the lead protein CD 13 (present in all multi-protein clusters, or, ▶ CMPs) leads to disassembly of the corresponding protein network and loss of function of the cell to polarize. This underscores the relevance of cell surface toponomics in understanding chronic disease and finding novel drug-target candidates. Analyzing the toponome of cell surface proteins will provide access to the mechanisms underlying self-organization and cell control in health and disease (Schubert et al. 2011; ▶ Clinical Aspects of the Toponome Imaging System (TIS)).



Cell Membrane Toponomics, Fig. 1 Illustration showing a scheme of the molecular protein species combined in various clusters along the cell surface of cell extensions (**a** and **b**) of a rhabdomyosarcoma tumor cell. The sequential arrangement and geometry of the corresponding multi-protein clusters is highly similar in both these extensions. Note that the scheme of the supramolecular structure of the single molecular components within these clusters is shown from (**c**–**g**). The colors "*red*," "*dark blue*," "*yellow*" "*light blue*," "*green*" and correspond to the colors in the toponome maps in (**a**) and (**b**). *Scale bar: 2 μm* (Modified after Schubert 2010)

References

Barclay AN, Brown MH, Law SKA, McKnight AJ, Tomlinson MG, van der Merwe PA (1995) The leucocyte antigen factsbook, 1st edn. Academic, San Diego

Lodish H, Berk A, Zipursky SL et al (2000) Molecular cell biology, 4th edn. W. H. Freeman, New York

Schubert W (2010) On the origin of cell functions encoded in the toponome. J Biotechnol. 15; 149(4):252–259, Review doi: org/10.1016/j.jbiotec.2010.03.009

Schubert W, Gieseler A, Krusche A, Serocka P, Hillert R (2011) Next-generation biomarkers based on 100-parameter functional super-resolution microscopy TIS. N Biotechnol. doi:10.1016/j.nbt.2011.12.004

Zola H (2001) Human leukocyte differentiation antigens as therapeutic targets: the CD molecules and CD antibodies. Expert Opin Biol Ther 3:375–383, Review

Cell Proliferation

▶ Modeling, Cell Division and Proliferation

Cell Proliferation Process

▶ Cell Cycle, Biology

Cell Selection

▶ Lymphocyte Dynamics and Repertoires, Biological Methods

Cell Sorting

Xiaojun Liu
Internal Medicine, The Second hospital of Hebei Medical University, Shijiazhuang, Hebei, China

Definition

Fluorescence-activated cell sorting (FACS): FACS is a special technique based on flow cytometry (▶ Flow Cytometry http://en.wikipedia.org/wiki/Flow_ cytometry), it can make a physical separation of interested cells or particles from a heterogeneous population through measure and select user-defined cell types by illuminating individual cells with a laser and detecting emitted light which can be spectrally separated to show the individual character of the interested cells. After the interested cells are recognized the FACS will produce the fluid into single cell containing droplet and make it electrically charged, then when it passes through an electric field between two high voltage deflection plates of opposite polarities, the droplets contain single cell will be deflected into different containers according to the commander of the user-defined program.

Characteristics

1. Highly efficient: flow cytometer can examine hundreds of cells in a second and decide which cell or particle is interested and charge it to be separated; this is almost the most high-speed sorter now days.
2. High purity: If only the program has been set by the user, FACS can get a pure group of the interested cells since the technique is based on the reaction of antigen and antibody.
3. Accurate number control of sorted cells or particles: FACS can even put a single cell or particle into committed container, so it is easy to control the interested cells number according to the user.

Basic Instruments for Cell Sorting

Cell sorting of FACS needs the help of flow cytometry which includes the following: the fluidics system, the optics and detection, the signal processing, and electrostatic cell sorting.

Fluidics System

In order to be interrogated by the machine's detection system, the cells or the particles injected into the machine must be ordered into a steam of single particle, the fluidics system is designed for this aim. It consists of a central channel/core through which the sample is injected, enclosed by an outer sheath that contains faster flowing fluid. Both of the sample and sheath are driven by an air pump and the air pressure for them will be adjusted by their own pressure regulator, respectively, for the sheath at 4.5 psig and for the

sample at 4.6, 4.8, or 5.0 psig according to need of sample speed. As the sheath fluid moves driven by the air pump, it creates a massive drag effect on the narrowing central chamber. This alters the velocity of the central fluid whose flow front becomes parabolic with greatest velocity at its center and zero velocity at the wall. The effect creates a single file of particles and is called hydrodynamic focusing. Then the particles will become a single stream after they pass this area which is ready for the machine to detect.

Optics and Detection

The excitation optics is consisted of a laser, lenses, and prisms. The laser is the origin of light while the lenses and prisms are aimed to shape and focus the laser beam. After the light beam passes the lenses and prisms it reached a size of 64 × 20 μm light spot. After hydrodynamic focusing the particles will meet the light beam of the laser or the arc lamp in the flow cytometry, then light scattering or fluorescence emission (if the particles is labeled with a fluorochrome) will provide information about the particles properties. The information will be detected by the detections. There are three kinds of detections in the flow cytometry: the forward scatter channel (FSC), the side scatter channel, and fluorescence channel. FSC collects the light that is scattered in the forward direction, typically up to 20 offset from the laser beam's axis. The intensity of the FSC will show the roughly relative size of the particles. It can be used to distinguish between the cellular debris and living cells. SSC collects the part of light that is scattered by the particles at a 90 angle to the excitation line. It gives the information of the content within the particles. The fluorescence (FL) channel collects the fluorescence emited from the fluorescence materials conjugated on the antibody that are special to the cell marker, different FL detect special wavelength light, this is controlled by the optical filter, there different kinds of filter: long pass filter, short pass filter, and band pass filter. They selectively pass certain wavelength light while block the left. Another kind of filter called dichroic mirror it works in a different way, it can pass certain kind of light but block left by deflect it. FL1 normally accepts light with wavelength of 515–545 nm. FL2 564–606 while equal and above 650 nm light will be collected by FL3. Other FLs may be designed to collect special wavelength light according to different manufacture.

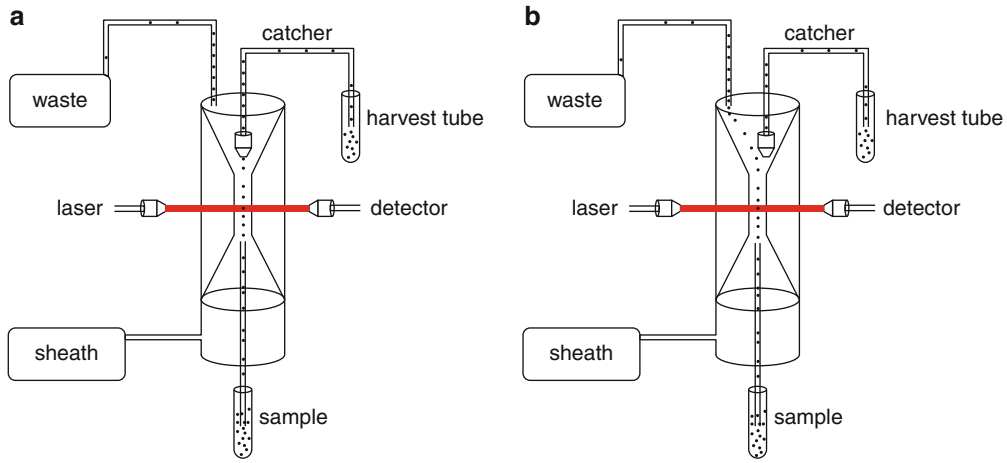
Electronic System

The electronic system is aimed to convert the optical signals to proportional electronic signals (voltage pulses), then after analysis of the voltage pulse height it will convert the signals to digital signals, and at last the computer will analyze the signals and display so that we can get the data we want. After the photons are collected by the photon diode it will turn this signals into electronic signals, while for the FSC signals they were first passed on by a pre amp (E00, E01, E02, E-1). Then they were passed on to the amplifier. The amplifier has two style LIN and LOG, the Lin amplifier will amplify at levels of 1.00–9.99. FSC always choose this style. While for the SSC, FL1, FL2, and FL3 they have different pathways for the amplification course, at the beginning of photon diode there is a PMT power supply (levels 150–999 V) connects with it, and after the current output by the photon diode the detector will not preamplify the signals, but directly pass it on to the amplifier. The signals from SSC FL1 FL2 and FL3 normally are amplified by the LOG style. When the cells or the particles pass through the laser area, signals of light will be turned into electronic signals and then will create a pulse.

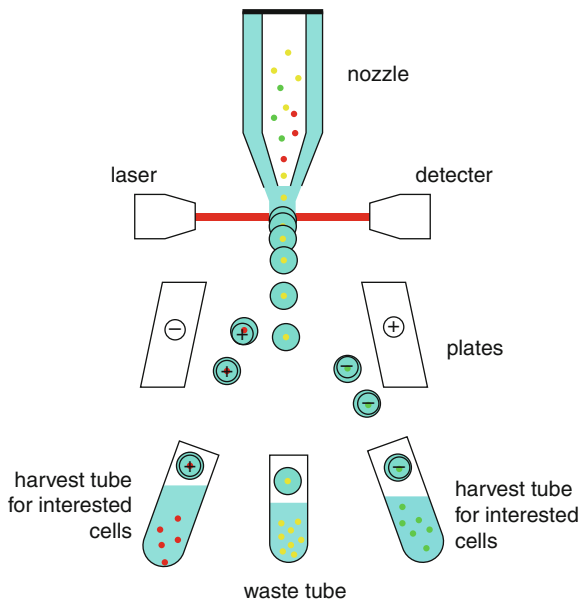
Sorter

Sorter is the final part of the FACS. It receives instructions from the flow cytometer set by the user and divides the cells into interested ones and waste. Different type of sorter may have different manners of harvesting cells, for example, some harvest cells through controlling a catcher; when the flow detects an interested cell it will direct the catcher to move to the fluid stream to catch the cell and then quickly shift aside when it is finished in order to let the waste pass through (Fig. 1). While some other type of sorter will work in a different way they use electric field force to separate the cells through making the cell-contained droplet electronic charged. When the cell-contained droplets pass through the high voltage plate area, the electric field force will drive the interested cells contained and waste droplets into different containers, so that the cells interested can be divided (Fig. 2).

When the cell passes the laser area the signals of the cell will be detected by the detector and then the electronic system of the cytometer will wait for a fixed period of time to allow the interested cell to reach the catcher tube, then the catcher tube will swing into the sample stream to capture the interested



Cell Sorting, Fig. 1 Cell sorter that sorting cells by catcher shifting



Cell Sorting, Fig. 2 Cell sorter that sorting cells by electrostatic field

cell (Fig. 1a). When the cell is judged to be the non-interested cell, the cytometer will switch the capture tube out of the sample steam, and then the cells will pass on to the waste (Fig. 1b).

The FACSVantage SE, a stream-in-air flow cytometer, works in a different way. First, the sample will be hydrodynamically focused so that the particles in the stream will pass the beam of light one after another. Then the character of the particle, like the FSC, SSC, fluorescence, etc., will be detected by the machine and is compared to the sort criteria set on the computer. If the particle character matches the selection criteria, the particle will be charged before the nozzle of fluidics system break it down form the fluid stream which is controlled by the vibration of the nozzle. After charged and broken down from the fluid stream the droplet liquid containing the interested particle will pass through a strong electrostatic field where the droplets will be deflected to the harvest tube or the waste tube (Fig 2).

Channels

Different fluorescence-activated cell sorting machine from different manufactory may have different channels for fluorescence detecting. Normally the more laser source the machine has the more channels detectors will be available on the machine. The followings are some of the channels that contained by the different machine.

FACSCalibur benchtop flow cytometer: FACSCalibur benchtop flow cytometer is a two laser source machine (488 nm blue laser and 635 nm red diode laser). It has six detector channels FSC which receive forward scatter light that pass a filter of 488/10 nm (a wavelength between 488 $\pm$10 nm, which means wavelengths of light that are between 515 and 545 nm). Others are SSC 488/10 nm, FL1 530/30 nm, FL2 585/42 nm, FL3 670LP (long pass which means wavelength over 670 nm can pass the filter and detected by the detector), FL4 661/16 nm.

BD LSR benchtop flow cytometer: BD LSR benchtop flow cytometer is a kind of cytometer that has three lasers, they are 325, 488, and 633 nm laser respectively. So in this kind of machine, the excitation wavelength band is relatively wider than the FACSCalibur which has two lasers, and more detector channels are set in this kind of machine enable the machine to detect more surface marker at the same time. The following are the available detector channels in this kind of machine: FSC 488 nm, SSC 488/25 nm, FL1 530/28 nm, FL2 575/25 nm, FL3 670LP, FL4 510/20 nm, FL5 380LP, FL6 660/13 nm.

With the modern techniques development more laser and detectors can be used at the same time, enable the machine to detect more and more parameters simultaneously so that the machine can get more details of the particles being detected, for example, BD influx system has enabled seven laser paths and supports 24 parameters simultaneously.

Detectable Parameters

Fluorescence-activated cell sorting is based on the technique of flow cytometry so the detectable parameters are the same with flow cytometer and with the development of flow cytometer, the detectable parameters will be enriched gradually. The following are some of the detectable parameters that can be detected:

1. Size and amount of content of cells through FSC and SSC
2. Different component of the cells such as: cell pigments such as chlorophyll, total DNA content, total RNA content, DNA copy number variation (by Flow-FISH), Glutathione, pH, intracellular ionized calcium, magnesium, membrane potential, etc.
3. Original function of cells such as: protein modifications, phospho-proteins, protein expression, and localization
4. Extra component like transgenic products in vivo, particularly the Green fluorescent protein or related fluorescent markers)
5. Antgiens on the cell surface, intracellular, nuclear such as cluster of differentiation (CD), various cytokines, secondary mediators, etc.
6. Function detection such as enzymatic activity, characterizing multidrug resistance (MDR) in cancer cells
7. Cell viability, apoptosis (quantification, measurement of DNA degradation, mitochondrial membrane potential, permeability changes, caspase activity), and membrane fluidity change
8. Different combinations of the detectable markers such as DNA/surface antigen, etc.

Application

The aim of sorting is to separate the interested cell from the sample for further research or clinical use, and if only the special character that is different from other cells can be detected by the flow cytometer the sorter will be able to pick it out. Cell sorting cannot only give you a pure cell population but also can control the population's concise cell number. This character is very important and can enable us to do a lot of things we need. The following are some examples for cell sorting applications:

1. To get pure population of the interested cells for special research or clinical use.
   (a) *Eliminating dead cells for accurate internal staining of cytokines:* Cells are incubated with EMA, produced by Molecular Probes and then EMA enters dead cells (membranes not intact) and intercalates into the DNA. Exposure to light from a fluorescent lamp causes covalent linkage of EMA to DNA. EMA fluorescence remains associated with the dead cells then EMA-stained cells can be gated out before analysis of the FACS data.
   (b) *Sex control of birth:* The yield of flow cytometric sorted X- and Y-chromosome-bearing sperm in a given time period is an important factor in the strategies used for fertilization and the production of sex-preselected offspring. With the help of the flow cytometer sperm can be divided into X-and Y-chromosome-bearing group, then they can be used for fertilizer to control the sex of offspring.
2. To control the concise number of the interested cells for special research or clinical use. For example: the sorter can easily recognize and drop a single interested cell to a well or tube so that single cell research can be done.
3. Get the rare interested cells from vast back ground cells.
   (a) Tumor infiltrating lymphocytes (TILs)
   (b) Residual tumor cells
   (c) Neonatal blood
   (d) Rare event analysis
   (e) Fetal cells in maternal blood

Cross-References

- ▶ Flow Cytometry
- ▶ Fluorescence
- ▶ Forward-scattered Light (FSC)
- ▶ Side-scattered Light (SSC)

References

Herzenberg LA, Parks D, Sahaf B (2002) The history and future of the fluorescence activated cell sorter and flow cytometry: a view from stanford. Clin Chem 48(10):1819–1827

http://en.wikipedia.org/wiki/Flow_cytometry#Fluorescence-activated_cell_sorting

http://www.abdserotec.com/support/electrostatic_cell_sorting-711.html

http://www.cytonome.com/ci/176/Existing_Cell_Sorting_Methods/

http://www.dddmag.com/product-bd-influx-51310.aspx

Kanz L, Bross KJ, Mielke R (1986) Fluorescence-activated sorting of individual cells onto poly-L-lysine-coated slide areas. Cytometry 7:491–494

Manual Part Number:11-11-32-01, BD Biosciences, San Jose

Rens W, Welch GR, Johnson LA (1999) Improved flow cytometric sorting of X- and Y-chromosome bearing sperm: Substantial increase in yield of sexed semen. Mol Reprod Dev 52(1):50–56

11-10784-02 Key operator training manual BD immunocytometry systems 2350 Qume drive San Jose CA 95131–1807

Cell Synchronization

- ▶ Cell Cycle, Synchronization

Cell System Ontology

Euna Jeong, Masao Nagasaki and Satoru Miyano
Institute of Medical Science, University of Tokyo, Tokyo, Japan

Definition

The cell system ontology (CSO) is a system dynamics–centered language for modeling, visualizing, and simulating biological pathways (Jeong et al. 2007a). CSO is a formal ontology that represents computational models as well as visualization of models. CSO can represent various types of biological pathway within a unified framework. As an exchange format, CSO is implemented in a formal ontology (▶ Web Ontology Language (OWL)).

Characteristics

CSO is based on a mathematical model called the hybrid functional Petri nets with extension (Nagasaki et al. 2004). Petri nets have graph-like structures consisting of three elements including place, transition, and arc. In CSO, the Petri net elements are renamed as more intuitive terms: entity, process, and connector, respectively, to bridge the gap between computer science and biology researchers.

Main Capabilities

CSO is a general framework for understanding the behavior of cell systems in an integrated way. The features of CSO are as follows: (1) manipulation of different levels of granularity and abstraction of pathways, for example, metabolic pathways, regulatory pathways, signal transduction pathways, and cell-cell interactions; (2) capture of both quantitative and qualitative aspects of biological function by using Petri net; and (3) encoding of biological pathway data related to visualization and simulation, as well as modeling.

Structure

CSO consists of approximately 200 classes for a comprehensive representation of cell system. There are several classes to represent cell system as a dynamic system.

- **CSMLBase** is the root class for all classes in CSO.
- All data in CSO is structured around **Project** which has slots to represent the comprehensive environment of a pathway model.
- Each project is required to have only one **Model**, which describes pathways via a set of processes.
- **Submodel** can be defined as a subset of a given model. Each submodel contains some selected elements of a model, which may be grouped to convey any meaning.

A model comprises a set of processes connected to entities via connectors, and facts to provide more information related to biological processes. The **ElementBase** class contains fundamental concepts for the model, which has four subclasses: **Entity**,

Process, and **Connector** are Petri net elements and are all related to dynamic simulation, whereas **Fact** represents other properties that cannot be described with Petri net.

- **Entity** describes biological entities (e.g., protein, complex, DNA, and small molecule), cellular compartment, and biological environments (e.g., UV, temperature, and pH).
- **Process** defines interaction among entities, for example, degradation, translation, and phosphorylation.
- **Connector** links entities and processes.
- **Fact** is designed for understanding the pathway functionalities and evaluating the status of dynamic simulation such as concentrations that should satisfy among simulation steps, view elements that do not effect among simulation steps, and biological elements that do not effect among simulation steps.

An entity reflects the concentration of the substance. A process is linked to entities by connectors that are incoming from an entity and outgoing to an entity. A process has a speed that depends on the concentration of the incoming entity. Each connector has its own threshold value. During simulation, the evaluated result of the threshold value decides whether the linked process is activated by consuming concentration of the input entity.

Several classes are also defined in CSO to represent various properties.

- **BiologicalBase** contains biological terms for annotating biological properties. Controlled vocabularies are defined to investigate the reuse of existing structured information from other sources, such as biological event, cell component, cell type, and evidence code.
- **SimulationBase** describes simulation properties specific to each Petri net elements, such as kinetics, thresholds, and variables.
- **ViewBase** is for visualization of model components, including geometric position, graphical shape, image file–related properties that link each element to graphical representation.
- **ChartBase** is for saving the results of model simulation as 2D plots.

Softwares

- Cell Illustrator Online (CIO) is a software platform for systems biology that uses the concept of Petri net for modeling and simulating biological pathways (Nagasaki et al. 2010). Biological pathways in CSO can be created, loaded, edited, saved, and simulated via CIO. The Cell System Markup Language (CSML) is a primary language of CIO based on XML and is fully compatible with CSO.
- Data conversion is possible from several formats, such as CSML to CSO, CSO to CSML, ▶ SBML (http://www.sbml.org/) to CSML, CellML (▶ CellML) to CSML, Transpath to CSML (Nagasaki et al. 2008), BioPAX (http://www.biopax.org/) to CSO (Jeong et al. 2007b) for pathway databases such as KEGG (http://www.kegg.org/), Reactome (http://www.reactome.org/), CellML repository (http://www.cellml.org/), and BioModels (http://www.biomodels.net/). CIO can read and convert these formats into CSML and CSO.
- As a data repository, the curated executable macrophage database MACPAK (http://macpak.csml.org) (Nagasaki et al. 2011) in CSML can be accessed from a web browser. TRANSPATH (Krull et al. 2006) at BIOBASE, focused on signal transduction pathways, is embedded in CIO and can be searched and edited on CIO.
- CSO validator (Jeong et al. 2011a, b) is a semiautomatic validation tool for CSO data, which reduces time spent on checking annotation mistakes and correcting problematic parts in terms of biological meaning and system dynamics.

Cross-References

- ▶ CellML
- ▶ Systems Biology Markup Language (SBML)
- ▶ Web Ontology Language (OWL)

References

Jeong E, Nagasaki M, Saito A, Miyano S (2007a) Cell system ontology: representation for modeling, visualizing, and simulating biological pathways. In Silico Biol 7(55):623–638

Jeong E, Nagasaki M, Miyano S (2007b) Conversion from BioPAX to CSO for system dynamics and visualization of biological pathway. Genome Inform 18:225–236

Jeong E, Nagasaki M, Ikeda E, Sekiya Y, Saito A, Miyano S (2011a) CSO validator: improving manual curation workflow for biological pathways. Bioinformatics 27:2471–2472

Jeong E, Nagasaki M, Ueno K, Miyano S (2011b) Ontology-based instance data validation for high-quality curated biological pathways. BMC Bioinformatics 12(Suppl 1):S8

Krull M, Pistor S, Voss N, Kel A, Reuter I, Kronenberg D, Michael H, Schwarzer K, Potapov A, Choi C, Kel-Margoulis O, Wingender E (2006) TRANSPATH: an information resource for storing and visualizing signaling pathways and their pathological aberrations. Nucleic Acids Res 1(34):D546–D551

Nagasaki M, Doi A, Matsuno H, Miyano S (2004) A versatile petri net based architecture for modeling and simulation of complex biological processes. Genome Inform 15(1): 180–197

Nagasaki M, Saito A, Li C, Jeong E, Miyano S (2008) Systematic reconstruction of TRANSPATH data into Cell System Markup Language. BMC Syst Biol 2:53

Nagasaki M, Saito A, Jeong E, Li C, Kojima K, Ikeda E, Miyano S (2010) Cell illustrator 4.0: a computational platform for systems biology. In Silico Biol 10(2):5–26

Nagasaki M, Saito A, Fujita A, Tremmel G, Ueno K, Ikeda E, Jeong E, Miyano S (2011) Systems biology model repository for macrophage pathway simulation. Bioinformatics 27: 1591–1593

CellML Model Curation

Catherine M. Lloyd and Geoff Nunns
Auckland Bioengineering Institute, University of Auckland, Auckland, New Zealand

Synonyms

Curation; Validation

Definition

► CellML model ► curation refers to the process of validating and annotating a CellML model. In most cases, the creation of a CellML model is a multistep translation process: from computational code to text and rendered mathematical equations and then back to code. Once created, a CellML model generally needs to be curated, since it is rare for a model to reproduce the published results at the first iteration. Typographical errors in the publication, missing parameters, equations, or unit definitions, mean the CellML model is often incomplete, contains errors, or unit inconsistencies. Model ► curation is an ongoing process for which we employ a variety of simulation and validation tools and often seek help from the original model author.

Characteristics

Background

Mathematical models play an essential role in interpreting complex biological processes. However, as these mathematical models themselves become increasingly complex, a need has arisen for defined standards to facilitate model exchange and reuse. ► CellML (Lloyd et al. 2004), and The Systems Biology Markup Language (SBML) (Hucka et al. 2003) are two XML-based markup languages that have been developed in response to such a need. Once a model has been described in either CellML or SBML, it can be uploaded into a centralized, online database, such as the CellML model repository (Lloyd et al. 2008, http://models.cellml.org/) or the BioModels Database (Le Novere et al. 2006, http://www.ebi.ac.uk/biomodels-main/). From there it is freely available to the public for download, simulation, and reuse.

Unless a modeler chooses to develop their mathematical model in CellML from the outset, the creation of a CellML model is a multistep translation process. It begins with a model being developed in a particular computational language, subsequent code conversion into text and rendered equations for publication, and finally translation of the text and equations into CellML. Once it is written, a CellML model has to be curated because each step in the translation process is potentially a source for errors.

Of the ~550 models in the CellML model repository (August 2012), over half have been curated. ► Curation can be carried out by the modeler at the time of CellML model development and/or translation, or by the team of CellML model curators after the model has been uploaded into the CellML model repository.

Curation Workflow

Vaild XML and CellML

The first step in curating a CellML model is to load and validate it with a CellML-compliant tool such as ► OpenCell (http://www.cellml.org/tools/opencell) or Cellular Open Resource, (COR http://cor.physiol.ox.ac.uk/). These modeling environments provide a validation service which highlights general errors and inconsistencies in the CellML model. CellML model validation is carried out by the software in a hierarchical manner. The base XML is checked first, and the tool ensures that it is well defined and

adheres to the XML standard, for example, that each opened XML tag is also closed. The next step tests for syntactic errors as each CellML file must follow general syntax rules: variables have to be defined within components, equations inside MathML elements, and so on.

These two types of code validation are automatically carried out when the CellML file is opened in either ▶ OpenCell or COR, and these errors have to be fixed before the model can be further edited as these types of errors render the model invalid CellML, and the modeling software is unable to work with such files.

The second stage of software validation is a semantic check, which is of main interest for ▶ curation. A common source of semantic errors are connections as any two components are only allowed a single mapping, and the variables mapped must be defined in each component and feature appropriate interfaces ("out" from the component in which the variable is defined and "in" into the component in which the variable is simply used). Another form of semantic errors concerns the mathematical formulation of the model. A model can be underconstrained if there is insufficient information to run a simulation – for example, if a variable is missing a defining equation or an initial condition. Conversely, a model can be overconstrained, for example, if a variable is both defined by an equation and is assigned an initial condition (and the equation is not a differential equation). Furthermore, a model can contain circular dependencies of variables, which in most cases can be detected by the software.

Units Consistency

One of the requirements of a CellML model is that all variables and numbers are assigned a unit – even if that unit declaration is "dimensionless." Further, as part of the validation service in the modeling environments, ▶ OpenCell and COR check for unit consistency. Unit consistency has two parts: across a model and across an equation. An example of the former is the variable "time"; if it is first defined with the unit of "second," "time" should not appear later in the model associated with the unit of "minute" or "hour." For the latter, the tools report unit imbalances in each individual equation. For example, it is incorrect to add together two variables with different units.

Model Simulation

Once the model is known to be valid CellML, with a complete set of equations and parameters and balanced units, the model is run and the simulation output is compared with the results of the original published model. In most cases, this involves setting the parameter values and simulation conditions to reflect those in the publication, and then comparing the simulation output with a figure or data in the article (see Fig. 1).

Obtaining the Original Model Code

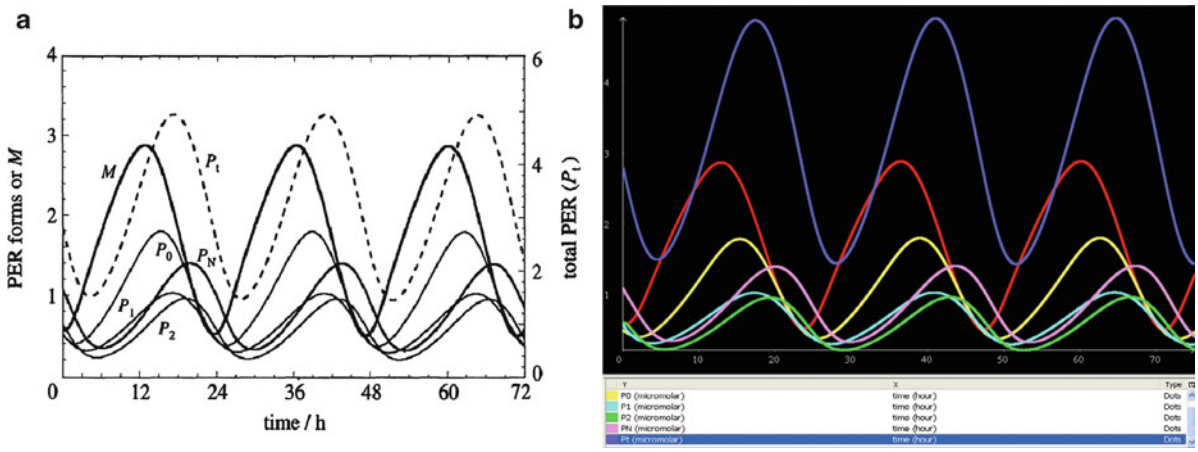
Unfortunately, due to the error-prone nature of the multistep model translation process, it is not uncommon to have a valid CellML file which does not reproduce the published results. In such a situation, the first step is to compare the list of equations and parameters present in the journal article with those in the CellML model to eliminate errors introduced in the creation of a CellML model. This "by eye" check is made easier by the verbose ▶ MathML equations being rendered into condensed, human readable equations (Fig. 2).

If, after this check, the CellML model still will not run to reproduce the published results, it is likely that errors were introduced during conversion of the original model code into the equations and text present in the published paper, and if possible a copy of the original model code is obtained. This primary source has not undergone any translation and is a true representation of the intended form of the published model.

As the research community has become increasingly aware of the CellML project, our chances of obtaining a copy of the original model code from the model authors have improved. However, for some older models we will have to accept that we are unlikely to get a copy of the original code. Further, due to differences in the capabilities of different model description languages, there may be certain functions in the original model which cannot be translated into CellML.

Curation Standards

All the models in the CellML model repository are assigned a ▶ curation rating in the form of a number of "stars" (ranging from zero to three) and a more verbose model status comment, which provides model-specific information such as whether the



CellML Model Curation, Fig. 1 (**a**) Published figure showing the results of the original model (Goldbeter (2005) Proc Biol Sci) and (**b**) the simulation output from the CellML version of the model

```
<math xmlns="http://www.w3.org/1998/Math/MathML">
  <apply><eq/>
    <apply><diff/>
      <bvar><ci>time</ci></bvar>
      <ci>M</ci>
    </apply>
    <apply><minus/>
      <apply><times/>
        <ci> vs </ci>
        <apply><divide/>
          <apply><power/>
            <ci> KI </ci>
            <ci> n </ci>
          </apply>
          <apply><plus/>
            <apply><power/>
              <ci> KI </ci>
              <ci> n </ci>
            </apply>
            <apply><power/>
              <ci> PN </ci>
              <ci> n </ci>
            </apply>
          </apply>
        </apply>
      </apply>
      <apply><times/>
        <ci> vm </ci>
        <apply><divide/>
          <ci> M </ci>
          <apply><plus/>
            <ci> Km </ci>
            <ci> M </ci>
          </apply>
        </apply>
      </apply>
    </apply>
  </apply>
</math>
```

$$\frac{dM}{dtime} = \frac{vs \cdot KI^n}{KI^n + PN^n} - \frac{vm \cdot M}{Km + M}$$

CellML Model Curation, Fig. 2 A comparison of the verbose raw MathML equation and the same equation rendered in COR

CellML model was translated from the published paper or from the original model code.

There are four defined ▶ curation levels:

- Level 0: Not curated.
- Level 1: The CellML model description is consistent with the mathematics in the original published paper.
- Level 2: The CellML model has been checked for:
  - Typographical errors
  - Consistency of units
  - Completeness, such that all parameters and initial conditions have been defined
  - Overconstraints, such that the model does not contain equations or initial values which are either redundant or inconsistent
  - Consistency of simulation output with the published results, such that running the model in an appropriate simulation environment reproduces the results in the original paper
- Level 3: The model has been checked to the extent that it is known to satisfy physical constraints such as conservation of mass, momentum, charge, etc. This level of ▶ curation needs to be conducted by a specialized domain expert, who is not the original model author.

However, flaws have been identified in this ▶ curation rating system. The system is not hierarchical – that is, a model assigned level 2 does not necessarily meet the criteria for level 1. In fact, the two levels are often mutually exclusive because in order to get the CellML model to replicate the published results (level 2) we have to adjust the published model description (level 1). Further, we have observed that the current "star system" can lead to uncertainty as to how to interpret the ▶ curation status of a model. For example, it is often a misconception that unless a model has been assigned three stars, it is not working. The reality is that domain experts are usually very busy individuals who do not have the time to spend on checking other people's models. Also, since most of the CellML models have been derived from published papers, we assume that the model has already undergone peer review and the reviewers were satisfied.

To solve this problem of misinterpretation, we are proposing to replace the current star system with a more descriptive set of ▶ curation flags. Initially, these will be based on the ▶ MIRIAM standard: the Minimum Information Requested in the Annotation of Biochemical Models (Le Novere et al. 2005). This is an agreed set of rules for curating quantitative models of biological systems which ensure a consistent standard of curation between models in distinct databases.

Conclusions

We encourage model developers to publish their computational code in a common format, such as CellML, concurrent with the manuscript of their paper. Indeed, certain journals are already encouraging modelers to submit their model code in SBML. However, until code submission becomes an obligatory step getting a model reviewed and published, the processes of model translation and ▶ curation will remain essential.

References

Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, Kitano H, Arkin AP, Bornstein BJ, Bray D, Cornish-Bowden A, Cuellar AA, Dronov S, Gilles ED, Ginkel M, Gor V, Goryanin II, Hedley WJ, Hodgman TC, Hofmeyr JH, Hunter PJ, Juty NS, Kasberger JL, Kremling A, Kummer U, Le Novere N, Loew LM, Lucio D, Mendes P, Minch E, Mjolsness ED, Nakayama Y, Nelson MR, Nielsen PF, Sakurada T, Schaff JC, Shapiro BE, Shimizu TS, Spence HD, Stelling J, Takahashi K, Tomita M, Wagner J, Wang J (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. Bioinformatics 19:524–531

Le Novere N, Bornstein B, Broicher A, Courtot M, Donizelli M, Dharuri H, Li L, Sauro H, Schilstra M, Shapiro B, Snoep JL, Hucka M (2006) BioModels database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. Nucleic Acids Res 34: D689–D691

Le Novere N, Finney A, Hucka M, Bhalla US, Campagne F, Collado-Vides J, Crampin EJ, Halstead M, Klipp E, Mendes P, Nielsen P, Sauro H, Shapiro B, Snoep JL, Spence HD, Wanner BL (2005) Minimum information requested in the annotation of biochemical models (MIRIAM). Nat Biotechnol 23:1509–1515

Lloyd CM, Halstead MD, Nielsen PF (2004) CellML: its future, present and past. Prog Biophys Mol Biol 85:433–450

Lloyd CM, Lawson JR, Hunter PJ, Nielsen PF (2008) The cellML model repository. Bioinformatics 24:2122–2123

CellML Model Repository

Catherine M. Lloyd and Tommy Yu
Auckland Bioengineering Institute, University of Auckland, Auckland, New Zealand

Synonyms

PMR2; Physiome model repository 2

Definition

The CellML model repository is an online database for storing mathematical models of physiological and biological processes. There are currently over 500 models in the repository, which are freely available for users to download, simulate, modify, and upload. The repository is powered by the ► Physiome Model Repository 2 (PMR2) software, which provides a web-based interface to the repository and defines the user workflows, enabling both human and machine interaction with the repository content.

Characteristics

The Purpose of the CellML Model Repository

The CellML model repository (Lloyd et al. 2008, http://models.cellml.org/) is an online database for storing mathematical models encoded in ► CellML. There exist similar efforts, such as the BioModels Database (Le Novere et al. 2006) which focuses on biochemical pathway models, described in peer-reviewed publications, and expressed in the Systems Biology Markup Language (SBML) (Hucka et al. 2003), JWS Online (Olivier and Snoep 2004) which combines a repository of kinetic models with extensive online simulation and analysis facilities, and ModelDB (Hines et al. 2004), which stores published models in the field of computational neuroscience.

In contrast with these other databases, which tend to focus on specific areas, such as systems biology pathway models or computational neuroscience and exclusively contain models which have already been peer reviewed and published, the CellML model repository contains models describing a wide range of biological processes, including signal transduction and metabolic pathways, gene regulation, the cell cycle, electrophysiology, hormone secretion, immunology, and muscle contraction. In addition, it also supports collaborative model development, merging and alteration, and models are not restricted to those which have been published in journals.

History of the CellML Model Repository

When it was created in 2000 the repository served as a place to store model test cases while the descriptive capabilities of the ► CellML language were explored. Today (June 2012), there are over 500 models in the repository and it acts as a valuable resource for researchers.

From the onset of the project, several objectives for the CellML model repository were identified. It was decided that the repository should be designed to:

- Promote the sharing of models
- Be publicly accessible
- Provide a user-friendly interface to access, query, and store models
- Provide a consistent workflow whereby models can be submitted, reviewed, validated, and published
- Facilitate feedback on models
- Provide a mechanism for advertising new models or changes to existing models

There have been three different versions of the CellML model repository. The original repository achieved few of the goals outlined above; although models were freely accessible to the public, it had no mechanism to promote the sharing of models, facilitate feedback, or track the change history of a model. Similarly the second version of the repository, the ► Physiome Model Repository 1 (PMR1), failed to address many of the objectives. Finally, in July 2009, the ► Physiome Model Repository 2 (PMR2) was created, and the design objectives were met.

The Underlying Software

With the introduction of the PMR2 software in July 2009, the CellML model repository became more than just a place to store models. Some of the key features of the current model repository are:

- Facilitated model exchange directly between modelers, in possibly distant locations, without reliance on a central repository
- A detailed change history record for each model

- User access workflows to control privacy when required
- Embedded ▶ workspaces to enable model reuse
- The ability to include extra information, such as the experimental data on which the model is based, to enhance the model descriptions (metadata)

The content of the repository is managed by using Mercurial (http://mercurial.selenic.com/), a ▶ Distributed Version Control System (DVCS). Such a system allows modelers, who may be located on opposite sides of the world, to work together on the same model simultaneously. This system tracks every set of changes made to the model file, by each individual author, thus safeguarding the data from being unintentionally overwritten. As it is also a distributed system, modelers are able to work independent of the CellML model repository, and can share their changes directly with each other until they decide the model is ready to be uploaded into the repository.

The combination of these above features, together with the CellML language and associated metadata specifications, provides a collaborative model development environment which is capable of enhancing communication throughout the modeling community.

Repository Users

Several specific classes of CellML model repository users have been identified. Each of these classes can be assigned a role, and each role has specific needs:

- *Modelers* create models and need a centralized place to store them. The required storage mechanism is a ▶ Distributed Version Control System (DVCS) to allow collaborators to be able to work on a model simultaneously.
- *Curators* check the quality of the models in the repository. They need to know when new models have been added, or old models have been updated, so they can be checked and then published or rejected.
- *General users* are usually interested in investigating the stored models. They want to be able to search the repository for keywords, model categories, or author names, and need an intuitive interface to access and download the models.
- *Administrators* control who has permission to upload material to the model repository, and they require an interface to grant or revoke permission settings for users of the system.
- *Senior Investigators*, or project leaders, often want to present the capabilities of the models to funding bodies or at conferences. They require a web-based intuitive interface through which they can easily showcase their group's research.

Navigating the CellML Model Repository

There are two main methods to locate a particular model in the CellML model repository:

- *Browsing by category*: Every model stored in the repository is tagged with a least one category type. If a user is interested in metabolism, for example, they can click on this category heading on the repository home page, and a list of all the metabolic pathways models in the repository will be displayed.
- *Keyword search*: If a user is interested in all the models developed by a particular author, or if they want to view everything related to a particular topic, such as HIV dynamics, they can carry out a specific search for the author or topic through the repository search facility.

Downloading a Model

There are two principal methods to download a model from the CellML model repository. The chosen method depends on the intentions of the user. A casual user may just want to test the capabilities of a working, fully curated model, to see how it behaves, while a modeler may want to take an existing model from the repository and either improve on it or develop it further. The former user can either click on a link to a pre-created model simulation session file (if one is available for that particular model), or they can download the raw model file itself and subsequently open it in their preferred simulation tool and run the model. By contrast the latter user needs to take advantage of the ▶ Distributed Version Control System (DVCS), they should clone (or copy) the entire model ▶ workspace (or folder) through a Mecurial client onto their own computer. This workflow ensures that a model's modification history is properly recorded, and enables the other advantages as were discussed in the software section above.

Uploading a Model

To upload a model, or any other associated data (such as an image file or experimental data sets), into the CellML model repository, a user requires an account. If the model already exists in the repository and this is

a newer version, the new model has to be committed to the local clones of an existing ▶ workspace with a comment on what has been changed, and then pushed back into the repository. It is then the responsibility of the curators to check the quality of the model before they ▶ expose (or publish) it. Alternatively, if the model being added is a completely new one, the creation of a new ▶ workspace is required. Depending on the permissions granted to the user, they can either create a new ▶ workspace themselves, or they have to ask the curators to create one on their behalf. The same pushing, commenting, exposure, and publication process then takes place.

Private Models and Password Protection

Sometimes model developers may want to utilize the ▶ Distributed Version Control System (DVCS) feature of the CellML model repository, to track the changes made to their model during the development process, without making their model publicly available until it is complete. They can work in a private ▶ workspace and choose not to ▶ expose their model until the very end. Until they are made publicly available, models will not come up in search listings. Alternatively, a modeler may want to provide only a few, select individuals with access to their model, such as collaborators, or reviewers of a journal article. If this is the case the model can be ▶ expose but user access to this exposure can be controlled by the repository administrator granting permission to only a few specific users.

Persistent Model References

Each model entry in the repository has a unique identifier in the form of a unique URL. Even if an older version of a model is superseded by a newer published version, the URL of the old model is retained, with a note that that particular version of the model has been superseded, or "expired." This feature allows a modeler to refer to a specific model version, at a specific point in time, and they can be confident that a link in a journal article, or in correspondence with a colleague, will remain unbroken.

Cross-References

- ▶ BioModels Database: A Repository of Mathematical Models of Biological Processes
- ▶ Distributed Version Control System (DVCS)
- ▶ JWS Online
- ▶ Systems Biology Markup Language (SBML)

References

Hines ML, Morse T, Migliore M, Carnevale NT, Shepherd GM (2004) ModelDB: a database to suport computational neuroscience. J Comput Neurosci 17:7–11

Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, Kitano H, Arkin AP, Bornstein BJ, Bray D, Cornish-Bowden A, Cuellar AA, Dronov S, Gilles ED, Ginkel M, Gor V, Goryanin II, Hedley WJ, Hodgman TC, Hofmeyr JH, Hunter PJ, Juty NS, Kasberger JL, Kremling A, Kummer U, Le Novere N, Loew LM, Lucio D, Mendes P, Minch E, Mjolsness ED, Nakayama Y, Nelson MR, Nielsen PF, Sakurada T, Schaff JC, Shapiro BE, Shimizu TS, Spence HD, Stelling J, Takahashi K, Tomita M, Wagner J, Wang J (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. Bioinformatics 19:524–531

Le Novere N, Bornstein B, Broicher A, Courtot M, Donizelli M, Dharuri H, Li L, Sauro H, Schilstra M, Shapiro B, Snoep JL, Hucka M (2006) BioModels database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. Nucleic Acids Res 34: D689–D691

Lloyd CM, Lawson JR, Hunter PJ, Nielsen PF (2008) The cellML model repository. Bioinformatics 24:2122–2123

Olivier BG, Snoep JL (2004) Web-based kinetic modelling using JWS online. Bioinformatics 13:2134–2144

CellML

Andrew K. Miller[1], Randall D. Britten[1], Catherine M. Lloyd[1], David P. Nickerson[1] and Poul M. Nielsen[2]
[1]Auckland Bioengineering Institute, University of Auckland, Auckland, New Zealand
[2]Department of Engineering Science, University of Auckland, Auckland, New Zealand

Definition

CellML (Hedley et al. 2001; Cuellar et al. 2006) is a format for specifying mathematical models, building upon Extensible Markup Language (XML). It is supported by a range of software tools and databases, and is most commonly used to specify mathematical models of biological systems.

Characteristics

Systems biology researchers often need to exchange mathematical models. One approach is to publish a paper containing the mathematical equations in the model. However, the process of manually converting a model to mathematical equations, and then back into a computer modeling format is both time consuming and error prone. Another approach is to exchange programs for solving the model in a procedural language (such as Fortran or MATLAB). This approach has several major drawbacks: it tends to result in models where the numerical algorithms used are intermixed with the model, it is difficult to compose two models to make a model incorporating features of each model, and it is hard to use the model for any purpose other than the one it was originally coded for.

These problems are avoided by the CellML format (http://www.cellml.org), which acts as a container for mathematical expressions encoded in Mathematical Markup Language (MathML), and so can represent a wide variety of models. CellML tools are generally focused on supporting the class of problems known as systems of differential algebraic equations (DAEs).

CellML has been used to describe a wide range of models relevant to systems biology. At the time of writing, the CellML Model Repository included 500 models (including electrophysiology, gene regulation, metabolism, and signal transduction models (Cooling et al. 2007; Nickerson and Buist 2008; Terkildsen et al. 2008) through to endocrine models). CellML allows models to be built in a reusable and modular fashion, and this has been exploited to build a modular library of synthetic biology models (Cooling et al. 2010).

All mathematical expressions in a CellML model are contained within components. A component is a container that may represent concrete or abstract concepts. For example, a component could represent an ion channel, a compartment in a cell, the surrounding environment of a model, or the collection of adjustable parameters.

Each CellML component may contain variables, which are named values that may or may not change with respect to other variables. Every named value in the mathematics within a component needs to be declared as a variable; variables are used to represent parameters, state variables, and named constants.

In CellML, the physical units of all variables and constant numbers (which can be dimensionless) must be specified. If units are specified, they may either be one of the standard SI base or derived units, or units defined in the model. New units can be defined in terms of other units with specified multipliers and offsets (e.g., a millimeter can be defined as a thousandth of a meter, nanomolar can be defined as a billionth of a mole per liter), and new base units can also be defined.

A variable in one component can be connected to a variable in another component. A connection between two variables means that the two variables represent the same quantity. If the two connected variables are in the same physical units, then the connection is treated as equality. If they are in different, but dimensionally compatible units, then the CellML processing software will automatically apply the appropriate units conversion factor.

In a CellML model, components can be arranged in an encapsulation hierarchy. This can be used so that some components (the encapsulation children) describe the internal details of another component, which provides a higher level interface (the encapsulation parent). The internal details of each encapsulation child component can in turn be described in the next level of the encapsulation hierarchy, and so on to whatever depth is required.

To ensure that there is a distinction between the internal and externally exposed parts of a model, components have two types of interface, the public interface and the private interface. A connection between variables in encapsulation siblings uses the public interface of both components. A connection between an encapsulation parent and an encapsulation child uses the public interface of the encapsulation child and the private interface of the encapsulation parent. Variables can have different visibility on the public and private interfaces. Their visibility is directional, and is set to either "in" or "out," which indicates that the variable is visible on that interface, and may be connected to a variable of opposite directionality. Visibility defaults to "none," meaning that the variable is not visible on that interface.

In large systems biology models, the same mathematical constructs are often used multiple times, both

within a model, and across different models. For example, a particular mathematical formulation for an ion channel might be used to represent both sodium and potassium channels in a cell model, and might be reused in other models. Duplicating the markup for the definition of an ion channel component multiple times would be a repetitive and error-prone approach, and improvements to the component would need to be added manually to all copies. To address this, CellML 1.1 uses a mechanism called importing. Importing allows components and units to be imported by linking to the Uniform Resource Identifier (URI) of the imported model document from within the importing model document. When a component is imported, the entire encapsulation hierarchy underneath it is implicitly imported, including the connections between components in that encapsulation hierarchy, so that complex submodels made up of multiple components can be imported.

Using the structures described so far, CellML allows the mathematics underlying a model to be described. However, for a model to be useful, it is necessary to describe the correspondence between parts of the model and parts of the real-world system being modeled. For this reason, CellML models can include information, known as metadata, about the model and its parts (Beard et al. 2009). This metadata is normally represented within the CellML model in Resource Description Format (RDF). Metadata specifications exist for basic information about a model, such as the details of the publication it is based on, and the chemical species represented by a concentration variable. There are also metadata specifications for stipulating parameters to numerical integrators, and how to graphically display simulation results.

A number of tools and libraries capable of processing CellML models have been developed (Garny et al. 2008); a listing can be found at http://www.cellml.org/tools.

The CellML API (Miller et al. 2010) is a description of a programming interface for working with CellML models. Software developers can use the CellML API when incorporating support for the CellML format into their software. The API includes a core for performing basic model manipulations, as well as a number of optional extension modules. The API comes with an efficient full reference implementation in C++, licensed under Free/Open Source licenses so that it can be freely used. In addition, bindings that allow this implementation to be accessed from other languages are available (e.g., for Python and Java).

Modelers can use CellML as the primary source language for models, while leveraging numerical tools in other environments. This is achieved using an API extension module that allows tools to convert models from CellML into procedural programming languages like C or MATLAB.

In addition, modelers can make use of several popular stiff and non-stiff numerical integrators, using an API module that allows simulations of models to be run with numerical integrators from the GSL, CVODE, and IDA packages.

Another API module can be used to validate that models follow the rules of the CellML specification, as well as good modeling practices. For example, the validation service can generate a warning if the units in a mathematical expression are inconsistent.

There is also a range of tools for use by modelers: one of these is OpenCell, which is based on the CellML API. It is intended to be used by people creating or working with CellML models. Models can be edited and validated, simulations can be run and results can be plotted on graph axes. OpenCell is specifically designed to support working with multiple models simultaneously, and allows the results from different models to be compared graphically on the same set of axes.

OpenCell also supports session files, which are a stored description of the complete OpenCell working environment including the list of open models and information on which graphs to display. Session files can also include embedded figures; these figures can be made interactive so that clicking on part of a figure will bring up a corresponding trace on the graph.

COR (http://cor.physiol.ox.ac.uk) is another popular software application for modelers, focused on the editing and use in simulation of CellML, with similar features to OpenCell. Other end-user tools that do not focus on CellML, but include support for it are JSim, VirtualCell, and insilicoIDE.

An important bioengineering application of CellML is for multiscale organ simulation. Two simulation systems that allow part of the mathematical model to be specified by means of CellML are Chaste and OpenCMISS.

Cross-References

- ▶ CellML Model Repository
- ▶ Systems Biology Markup Language (SBML)

References

Beard DA, Britten R, Cooling MT, Garny A, Halstead MDB, Hunter PJ, Lawson J, Lloyd CM, Marsh J, Miller A et al (2009) CellML metadata standards, associated tools and repositories. Phil Trans R Soc A: Math Phys Eng Sci 367(1895):1845

Cooling M, Hunter P, Crampin EJ (2007) Modeling hypertrophic IP3 transients in the cardiac myocyte. Biophys J 93(10):3421–3433

Cooling MT, Rouilly V, Misirli G, Lawson J, Yu T, Hallinan J, Wipat A (2010) Standard virtual biological parts: a repository of modular modeling components for synthetic biology. Bioinformatics 26(7):925

Cuellar A, Nielsen P, Halstead M, Bullivant D, Nickerson D, Hedley W, Nelson M, Lloyd C (2006) CellML 1.1 Specification. CellML Specification. http://www.cellml.org/specifications/cellml_1.1

Garny A, Nickerson DP, Cooper J, Santos RW, Miller AK, McKeever S, Nielsen PMF, Hunter PJ (2008) CellML and associated tools and techniques. Phil Trans R Soc A: Math Phys Eng Sci 366(1878):3017

Hedley WJ, Nelson MR, Bullivant DP, Nielsen PF (2001) A short introduction to CellML. Phil Trans R Soc Lond A: Math Phys Eng Sci 359(1783):1073

Miller AK, Marsh J, Reeve A, Garny A, Britten R, Halstead M, Cooper J, Nickerson DP, Nielsen PF (2010) An overview of the CellML API and its implementation. BMC Bioinformatics 11(1):178

Nickerson D, Buist M (2008) Practical application of CellML 1.1: the integration of new mechanisms into a human ventricular myocyte model. Prog Biophys Mol Biol 98(1):38–51

Terkildsen JR, Niederer S, Crampin EJ, Hunter P, Smith NP (2008) Using physiome standards to couple cellular functions for rat cardiac excitation-contraction. Exp Physiol 93(7):919–929

Cell-to-Cell Communication

- ▶ Cellular Communication

Cellular Aging

- ▶ Senescence

Cellular Automata

Carlo Cosentino, Basilio Vescio and Francesco Amato
Department of Clinical and Experimental Medicine, Magna Graecia University of Catanzaro, Catanzaro, Italy

Synonyms

Cellular spaces; Cellular structures; Homogeneous structures; Iterative arrays; Tessellation automata; Tessellation structures

Definition

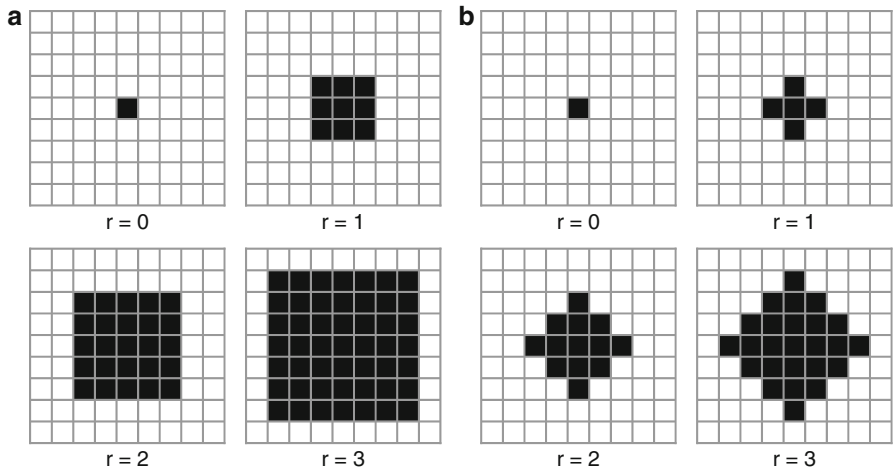
A cellular automaton (pl. cellular automata, abbrev. CA) is a discrete model studied in computability theory, mathematics, physics, complexity science (▶ Complexity), theoretical biology, and microstructure modeling. They are used to build parallel computing (▶ Parallel Computing, Data Parallelism) architectures as well as to model and simulate physical systems. Moreover they can be used to investigate and reproduce the emergence of patterns (▶ Spatiotemporal Pattern Formation), self-replication and self-similarity properties in natural systems.

Characteristics

History

CA were first developed in the 1940s by Stanislaw Ulam, who was investigating the growth of crystals, using a simple lattice network as his model, and by John Von Neumann, who was working on the problem of self-replicating systems (Von Neumann 1966).

In 1946, Norbert Wiener and Arturo Rosenblueth developed a cellular automaton model of excitable media. Their specific motivation was the mathematical description of impulse conduction in cardiac systems. Their original work continues to be cited in modern research publications on cardiac arrhythmia and excitable systems (Wiener and Rosenblueth 1946). In the 1960s, cellular automata were studied as a particular type of dynamical systems and the



Cellular Automata, Fig. 1 (**a**) Moore neighborhood; (**b**) Von Neumann neighborhood

connection with the mathematical field of symbolic dynamics was established for the first time (Hedlund 1969).

In the 1970s, K. Zuse proposed that the physical laws of the universe are discrete by nature, and that the entire universe is the output of a deterministic computation on a giant cellular automaton (Zuse 1982). In those same years, a two-state, two-dimensional cellular automaton named "Game of Life" became very widely known, particularly among the early computing community. The Life CA is able to perform computations, and after much effort, it has been shown that the Game of Life can emulate a universal Turing machine (Rendell 2002).

In the 1980s, Stephen Wolfram started a systematical investigation of a very basic but essentially unknown class of cellular automata, which he called "elementary cellular automata." The unexpected complexity of the behavior of these simple automata led Wolfram to suspect that ▶ complexity in nature may be due to similar mechanisms (Wolfram 2002).

Cellular Automata Structure

A CA consists of a regular grid of elements, named cells, each in one of a finite number of states, such as "On" and "Off" or "0" and "1." The grid can be in any finite number of dimensions.

Neighborhood

For each cell, a set of cells called "neighborhood" is defined, consisting of the cell itself, together with the surrounding cells within a certain distance. In the case of two-dimensional CA, the most used neighborhoods are Moore neighborhood and Von Neumann neighborhood (Fig. 1). More formally, if we refer to a certain cell in the grid as a location identified by discrete coordinates (x, y), being (x_0, y_0) the central cell and being r the neighborhood radius, we can define its Moore neighborhood as

$$N_M(x_0, y_0) = \{(x, y) : \ |x - x_0| \ <= r, |y - y_0| \ <= r\}, \tag{1}$$

while its Von Neumann neighborhood is

$$N_V(x_0, y_0) = \{(x, y) : \ |x - x_0| + |y - y_0| \ <= r\}. \tag{2}$$

Given N, the total number of cells, and being $\alpha = (x, y)$ a generic cell, we can define the neighboring function as (Sipper 1997) $g : N \times N \rightarrow 2\ N \times N$, defined as

$$g(\alpha) = \{\alpha + \delta_1, \alpha + \delta_2, \ldots, \alpha + \delta_n\}, \tag{3}$$

for all $\alpha \in N \times N$, where $\delta i \in N \times N$, $i = 1, \ldots, n$, is fixed.

Rule

Each cell contains a copy of the finite automaton (V, v0, f), where V is the set of cellular states, v0 is a particular state called quiescent state, and f is the transition function $f : V\ n \rightarrow V$. This function is subject to the constraint f(v0, v0, …, v0) = v0, that

is, a cell whose neighborhood is entirely quiescent remains quiescent. The state vt(α) of a cell α at time t is precisely the state of its associated automaton at time t. Each cell α is connected to the n neighbor cells α + δ1, α + δ2, . . ., α + δn. The neighborhood state function ht : I × I → V n is defined by ht(α) = (vt(α), vt(α + δ2), . . ., vt(α + δn)), assuming that δ1 = (0, 0).

Now we can relate the neighborhood state of a cell α at time t to the state of that cell at time t + 1 by f(ht(α)) = vt + 1(α). The function f is referred to as the CA rule and is usually given in the form of a rule table, specifying all possible pairs of the form (ht(α), vt + 1(α)). Such a pair is termed a transition or rule-table entry. An allowable assignment of states to all cells in the space is called a configuration. By applying the transition function to an allowable configuration, a new configuration is generated and reiterating this process gives out a succession of configurations, called evolution. A rule (and its corresponding CA) is said to be totalistic when the value of a cell at time t depends only on the sum of the values of the cells in its neighborhood (possibly including the cell itself) at time t − 1.

Evolution

An initial state (time t = 0) is selected by assigning a state for each cell. A new generation is created (advancing t by 1), according to rule f. For example, the rule might be that the cell is "On" in the next generation if exactly two of the cells in the neighborhood are "On" in the current configuration, otherwise the cell is "Off" in the next generation. In uniform CA, the rule for updating the state of cells is the same for each cell, while it may change in nonuniform CA. Typically, the rule does not change over time, and is applied to the whole grid simultaneously.

The Game of Life

The best-known two-dimensional cellular automaton is the famous Game of Life, devised by John H. Conway in 1970. It consists of a two-dimensional grid of cells, whose state may be dead (0) or alive (1), and Let $\Delta = \{\delta_0, \delta_1, \ldots, \delta_8\} = [-1, 1] \times [-1, 1]$, being $\delta_0 = (0, 0)$, and let be $S = \sum_{k=1}^{8} h^t(\alpha)^{(k)}$ the sum of the states of all the neighbors of cell α. The following rule describes Conway's Game of Life cellular automaton:

$$v^{t+1}(\alpha) = \begin{cases} 0, & S < 2 \vee S > 3 \\ v^t(\alpha), & S = 2 \\ 1, & S = 3 \end{cases} \quad (4)$$

A compact representation of this rule is"B3/S23": This essentially means that, for the next generation, a cell has a new birth (B) if its neighborhood consists of exactly 3 alive cells and survives (S) if there are two or three living cells in its Moore neighborhood (r = 1).

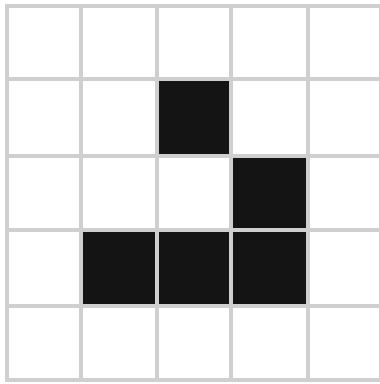
Properties

CA are actually simple computer programs capable of a remarkable range of behaviors. Some CA have been proven to be universal computers. Others exhibit properties familiar from traditional science, such as thermodynamic behavior, continuum behavior, conserved quantities, percolation, sensitive dependence on initial conditions, and others. They have been used as models of traffic, material fracture, crystal growth, biological growth, and various sociological, geological, and ecological phenomena. Another feature of simple programs is that making them more complicated seems to have little effect on their overall ▶ complexity and this argument has been used as evidence that simple programs are enough to capture the essence of almost any complex system (▶ Complex System). Von Neumann showed that some CA can be universal, in the sense that they can perform every computation, just like a universal Turing machine.

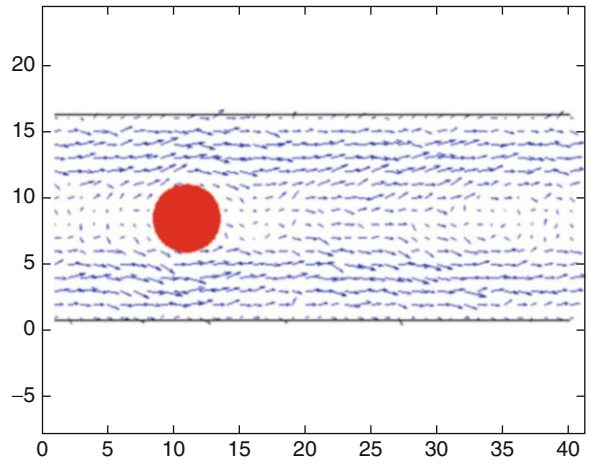
The Game of Life, given a proper initial configuration, can generate gliders (Fig. 2), or patterns that are able to replicate and propagate themselves through the grid. This property has revealed to be useful for performing calculations, and a universal Turing machine has been implemented in Life, thus demonstrating the universality of this CA (Rendell 2002).

Computing Architectures Based on CA

Toffoli and Margolus (1987) have pioneered the idea of building computer architectures based on cellular automata. Conventional computer architectures are optimized for the arithmetic treatment of continuum models. On the other hand, CA can faithfully model continuum systems, such as fluids. Unlike differential equations (▶ Ordinary Differential Equation (ODE)), they can be realized exactly by digital hardware. With a more appropriate architecture, a performance factor of at least 104 can be easily gained in the simulation of



Cellular Automata, Fig. 2 A glider in the Game of Life: This pattern can replicate and propagate itself through the CA grid



Cellular Automata, Fig. 3 Simulation of a flow through a circular cylinder obtained by means of a lattice-gas cellular automaton

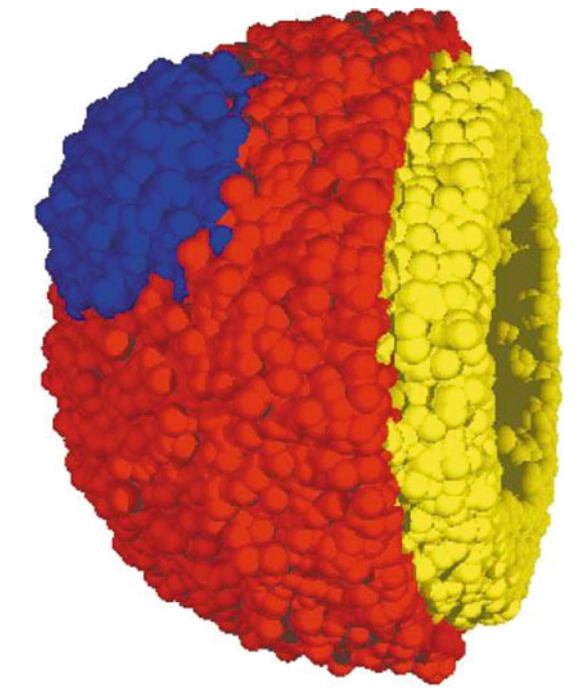
cellular automata, thus allowing to simulate very complex physical phenomena. This is the case of the CAM (Cellular Automata Machine) architectures implemented in hardware by Toffoli and Margolus, successfully used to model and solve fluid dynamics, diffusion, and collective phenomena problems.

Applications

The following are some of the many applications of cellular automata for the simulation of physical systems.

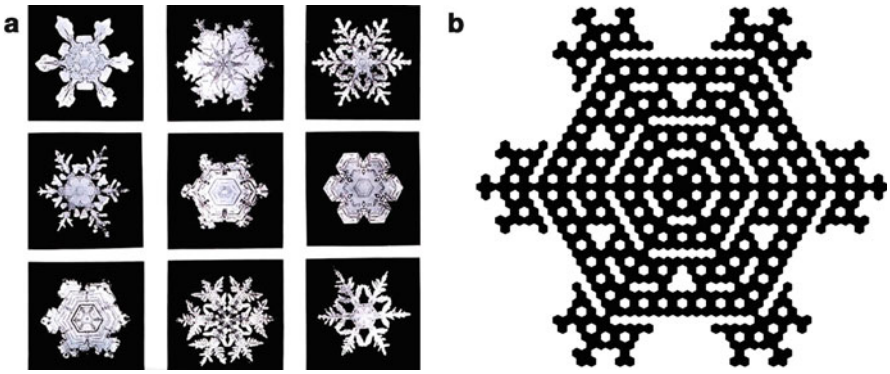
Hydrodynamic modeling: Differential equations, such as the Navier-Stokes equation, capture important macroscopic aspects of fluid dynamics; however, their implementation on a digital computer is not the equation themselves, but finite models obtained from them by truncation and roundoff. It is possible to simulate analogous macrodynamics starting directly from discrete microscopic models, cellular automata that idealize the motion and collisions of individual particles. Figure 3 shows the simulation of a flow through a circular cylinder obtained by means of a lattice-gas cellular automaton (Wolf-Gladrow 2005) (Lattice-Gas Cellular Automaton Models for Biology).

Growth processes: This is useful for the study of tumor growth dynamics in tissues. Cellular automata are used to model cell behavior and interactions at the microscopic level, in order to simulate and evaluate growth dynamics, vascularization or response to therapy. Figure 4 was originated by a model that takes into account four cell types: healthy cells, proliferating tumor cells, non-proliferating tumor cells, and necrotic tumor cells, using a simple set of rules and a set of four microscopic parameters that account for the nutritional needs of the tumor, cell-doubling time, and an imposed spherical symmetry (Kansal et al. 2000).

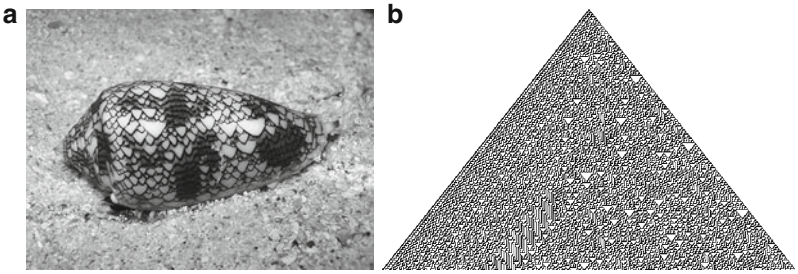


Cellular Automata, Fig. 4 A CA model of tumor growth

Pattern formation: Figures 5 and 6 show example patterns found in nature (Spatio-temporal pattern formation), together with their reproductions obtained by



Cellular Automata, Fig. 5 (**a**) Snowflakes and (**b**) a snowflake pattern reproduced by means of a two-dimensional CA on a hexagonal grid



Cellular Automata, Fig. 6 (**a**) Cone shell (conus textile) and (**b**) pattern generated by the rule 30 CA

means of cellular automata evolution. Snowflakes (Fig. 5) are obtained by running a two-dimensional CA on a hexagonal grid, while cone shell patterns (Fig. 6) are generated by a rule 30 CA.

Cross-References

- ▶ Complex System
- ▶ Complexity
- ▶ Lattice-Gas Cellular Automaton Models
- ▶ Ordinary Differential Equation (ODE)
- ▶ Parallel Computing, Data Parallelism
- ▶ Spatiotemporal Pattern Formation

References

Hedlund GA (1969) Endomorphisms and automorphisms of the shift dynamical system. Math Syst Theory 3:320–375

Kansal AR, Torquato S, Harsh GR IV, Chiocca EA, Deisboeck TS (2000) Simulated brain tumor growth dynamics using a three-dimensional cellular automaton. J Theor Biol 203:367–382

Rendell P (2002) Turing universality of the game of life. In: Adamatzky A (ed) Collision-based computing. Springer, London

Sipper M (1997) Evolution of parallel cellular machines: the cellular programming approach. Springer, Heidelberg

Toffoli M, Margolus N (1987) Cellular automata machines: a new environment for modeling. MIT Press, Boston

Von Neumann J (1966) Theory of self-reproducing automata. University of Illinois Press, Urbana

Wiener N, Rosenblueth A (1946) The mathematical formulation of the problem of conduction of impulses in a network of connected excitable elements, specifically in cardiac muscle. Arch Inst Cardiol Mex 16:205–265

Wolf-Gladrow DA (2005) Lattice-gas cellular automata and lattice boltzmann models – an introduction. Springer, Berlin

Wolfram S (2002) A new kind of science. Wolfram Media, Champaign

Zuse K (1982) The computing universe. Int J Theo Phys 21:589–600

Cellular Communication

Xiaojuan Sun
Zhou Pei-Yuan Center for Applied Mathematics, Tsinghua University of Beijing, Beijing, China

Synonyms

Cell-to-cell communication; Intercellular communication

Definition

Cell communication is an important process by which a cell sends or receives information from other cells. Individual cells, such as yeast, need to sense and respond to their environment, and communicate with other cells to have any kind of "social life." For example, when a yeast is ready to mate, it secrets a small protein called a mating factor. Neighboring yeast cells of opposite "sex" can detect this chemical mating call and respond by halting their cell cycle progress and reaching out toward the cell that emitted the signal. In a multicellular organism, cells must interpret the multitude of signals they receive from other cells to help coordinate their behaviors, such as cell fate decision in embryo development (Alberts et al. 2004).

In typical cell communications, the signaling cell produces a particular type of signal molecule that is detected by the target cell. The target cell possesses receptor proteins that recognize and respond specifically to the signal molecule. Signals can act over long or short range, depending on the way in which the signal is transported. For example, hormones produced in endocrine cell are secreted into the bloodstream and are often distributed widely throughout the body. In embryo development, signal proteins (morphogens) disperse from localized synthesis sites to form concentration gradients across the developing tissue. Neuronal signals are transmitted along axons to remote target cells. Cells with membrane-to-membrane interface can engage in contact-dependent signaling (Alberts et al. 2004).

Each cell can only respond to a limited set of signals selectively depending on the specific receptor proteins. Most extracellular signal molecules cannot pass through the plasma membrane. They bind to specific cell-surface receptor proteins to induce the intracellular signal transduction pathway. Receptor proteins act as transducers, binding to signaling molecules, and converting the signal from one physical form to another. Most cell-surface receptor proteins belong to one of three large families: ion-channel-linked receptors, G-protein-linked receptors, or enzyme-linked receptors. These families differ from each other by the nature of signals they generate when the extracellular signal molecule binds to them (Alberts et al. 2004).

References

Alberts B, Bray D, Hopkin K, Johnson AD, Lewis J, Raff M, Roberts K, Walter P (2004) Essential cell biology. Garland Science, New York

Cellular Potts Model

Anja Voß-Böhme[1], Jörn Starruß[2] and Walter de Back[2]
[1]Center for Information Services and High Performance Computing (ZIH), Technical University Dresden, Dresden, Germany
[2]Centre for High Performance Computing (ZIH), Technical University Dresden, Dresden, Germany

Synonyms

CPM; Glazier–Graner–Hogeweg model; Potts model, cellular/extended

Definition

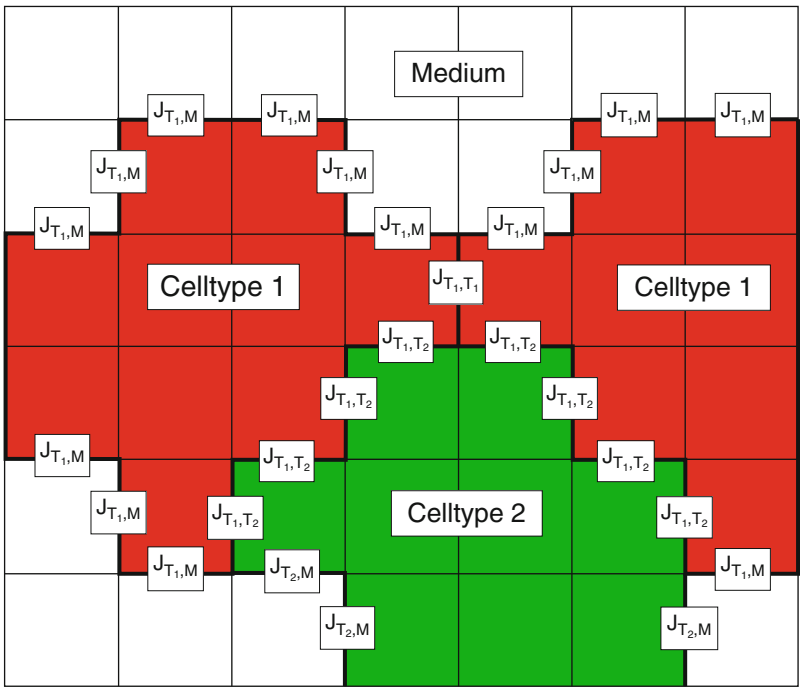
A cellular Potts model (CPM) is a spatial lattice-based formalism for the study of spatiotemporal behavior of biological cell populations. It can be used when the details of intercellular interaction are essentially determined by the shape and the size of the individual cells as well as the length of the contact area between neighboring cells.

Formally, a cellular Potts model is a time-discrete Markov chain (▶ Markov Chain). It is a lattice model where the individual cells are simply connected domains of nodes with the same cell index. A CPM evolves by updating the cells' configuration by one pixel at a time based on probabilistic rules. These dynamics are interpreted to resemble membrane fluctuations, where one cell shrinks in volume by one lattice site and a neighboring cell increases in volume by occupying this site. The transition rules follow a modified ▶ Metropolis algorithm with respect to a Hamiltonian.

Characteristics

Problem

The biological structure and function typically result from the complex interaction of a large number of



Cellular Potts Model, Fig. 1 Cell-surface interaction in the Cellular Potts Model. Three cells, each one covering several lattice sites, interact with each other at the cell surfaces. The strength J of the interaction depends on the cell types, type T1 depicted in *red*, type T2 in *green*. There are also interactions between the cells and the medium (*white*)

components. When ▶ spatiotemporal pattern formation in cellular populations or tissues is considered, one is often interested in concluding characteristics of the global, ▶ collective behavior of cell configurations from the individual properties of the cells and the details of the intercellular interaction. However, even if the basic cell properties and interactions are perfectly known, it is possible that – due to the complex structure of the system – the collective traits cannot be directly extrapolated from the individual properties. Therefore, appropriate mathematical models need to be designed and analyzed that help to accomplish this task on a theoretical basis. Cellular Potts models constitute a modeling framework that is applicable when the details of intercellular interaction are essentially determined by the shape and the size of the individual cells as well as the length of the contact area between neighboring cells.

This model class has been developed by Glazier and Graner (1993) in the context of cell sorting. The latter refers to the observed segregation of heterotypic cell aggregates into spatially confined homotypic cell clusters. The CPM was introduced to explore the tissue-scale consequences of the differential adhesion hypothesis (▶ Differential Adhesion Hypothesis) that holds that cell-type-dependent disparities in the expression of molecules that regulate intercellular adhesion are responsible for cell sorting. Since then, this formalism has been elaborated and applied to study a wide range of morphogenetic phenomena in developmental biology.

The Model

State Space

A CPM assigns a value $\eta(x)$ from a set $W = \{0, 1, ..., n\}$ to each site x of a countable set S, cp. Fig. 1. The set S resembles the discretized space and is often chosen as a two- or three-dimensional regular lattice. The set $W = \{0, 1, ..., n\}$ contains the so-called cell indices, where $n \in \mathrm{N}$ is the absolute number of cells that are considered in the model. The state of the system as a whole is described by *configurations* $\eta \in \mathrm{X} = W^S$. Given a configuration $\eta \in \mathrm{X}$, a *cell* is the set of all points in S with the same cell index, $\mathrm{cell}_w := \{x \in S : \eta(x) = w\}, w \in W \backslash \{0\}$. The value 0 is assigned to a given node, if this node is not occupied by a cell but by medium. Each cell is of a certain *cell type*, which determines the migration and interaction properties of the cell, the set of all possible cell types being denoted by Λ. Denote by $\tau : W \to \Lambda$ the map that assigns each cell its cell type. A cell with index $w \in W$ has *volume* (for the

Kronecker symbol δ it holds that $\delta(u,v)=1$ if $u=v$ and $\delta(u,v)=0$ otherwise)

$$V_w(\eta) := \sum_{x\in S} \delta(w,\eta(x)),$$

and *surface length*

$$M_w(\eta) := \frac{1}{2} \sum_{\text{interfaces } \{x,y\}} \delta(w,\eta(x)).$$

The sum in the last term is taken over all *interfaces* of a given configuration η that are all pairs of lattice neighbors which do not belong to the same cell.

Dynamics

A cellular Potts model (CPM) is a time-discrete Markov chain (▶ Markov Chain) with state space X, where the transition probabilities are specified with the help of a *Hamiltonian*. The latter is a function H: X → R which often has a special structure. Usually, it is the sum of several terms that control single aspects of the cells' interdependence structure. The standard CPM uses the following two terms. First a *surface interaction term*

$$H_s(\eta) = \sum_{\text{interfaces } \{x,y\}} J(\tau(\eta(x)),\tau(\eta(y))), \eta \in \mathrm{X}, \quad (1)$$

is specified. Here, J: $\Lambda \times \Lambda \to$ R, the matrix of so-called surface energy coefficients, is assumed to be symmetric. Second the *volume constraint*

$$H_v(\eta) = \sum_{w\in W} \lambda_{\tau(w)} (V_w(\eta) - v_{\tau(w)})^2, \eta \in \mathrm{X}. \quad (2)$$

is used. Here V_τ, the target volume, and λ_τ, the strength of the volume constraint, are cell-type-specific parameters, $\tau\in\Lambda$. Depending on the phenomenon under investigation, further summands can be included. For instance, a constraint can be put on the surface length (Ouchi et al. 2003)

$$H_m(\eta) = \sum_{w\in W} \alpha_{\tau(w)} (M_w(\eta) - m_{\tau(w)})^2, \eta \in \mathrm{X}. \quad (3)$$

Again m_τ, the target surface length, and α_τ, the strength of the surface constraint, are parameters, $\tau\in\Lambda$. Thus, the typical structure of a CPM-Hamiltonian is

$$H = H_s + H_v + H_0, \quad (4)$$

where H_s, H_v are given in (1) and (2) and H_0: X → R is a model-specific addend. See the section "Extensions and Applications" for additional examples of H_0.

Transitions from one configuration to another follow a special rule which is called *modified Metropolis algorithm* (▶ Metropolis Algorithm). First, two additional parameters are specified: a so-called temperature $T > 0$, which is a biological analogue of the energy of thermal fluctuations in statistical physics and is a measure of cell motility, and the *transition threshold* h, which accounts for energy dissipation during formation and breaking of intercellular bonds and avoids oscillatory behavior (Savill and Hogeweg 1997; Ouchi et al. 2003). Then, the following algorithm is performed (1):

1. Start with configuration η.
2. Pick a target site $x \in S$ at random with uniform distribution on S.
3. Pick a neighbor y of x at random with uniform distribution among all lattice neighbors of x.
4. Calculate the energetic difference $\Delta H_x^y := H(\eta_x^y) - H(\eta)$ of a transition $\eta \to \eta_x^y$, where $\eta_x^y(z) := \eta(y)$ if $z = x$ and $\eta_x^y(z) := \eta(z)$ otherwise.
5. Accept the transition by setting $\eta := \eta_x^y$ with probability $p(\Delta H_x^y)$, or ignore the transition with probability $1 - p(\Delta H_x^y)$, where
$$p(\Delta \mathrm{H}_x^y) = \begin{cases} 1 & \text{if } \Delta \mathrm{H}_x^y < h \\ e^{-(\Delta \mathrm{H}_x^y - h)/T} & \text{otherwise} \end{cases}$$
5. Go to 1 or end the algorithm.

Consequently, only such transitions are possible where the index of at most one lattice site is changed, resulting in a shift of the cell's center of mass. The new assignment to this lattice site is chosen from the cell indices of the neighboring lattice sites. These dynamics are interpreted to resemble membrane fluctuations, where one cell shrinks in volume by one lattice site and a neighboring cell increases in volume by occupying this site.

To complete the model, appropriate boundary conditions must be specified. If the influence of the boundary shall be neglected, periodic boundary conditions are used. This means that the space can be thought of as

being mapped onto a torus. However, fixed boundary conditions, where the interaction between cell surfaces and confining environment is explicitly modeled, can be defined as well.

Extensions and Applications

The CPM model formalism has been used for several problem-specific extensions. In general, this is done by including additional terms into the Hamiltonian (4). In some cases, these additional terms also depend on the chosen target spin, thereby changing the weights for the acceptance of a proposed transition in the modified Metropolis algorithm. The latter extensions are called *kinetic extensions*, since they directly affect the transition rates.

Cell motility emerges in the CPM implicitly from the fluctuations of the cells' center of masses. To explicitly model physical characteristics of cell motility such as cell persistence and inertia, additional terms that constrain the cell displacement per time step can be added to the difference ΔH of the standard CPM-Hamiltonian (4) that is calculated in step (3) of the modified Metropolis algorithm. Inertia, for example, has been modeled by constraining the cell velocity increment via the term

$$\Delta H_{\text{inertia}}(t) = \sum_{w \in W} \lambda_{\text{inertia}}(w) \| \overrightarrow{vel}(w,t) - \overrightarrow{vel}(w, t - \Delta t)\|^2, \quad (5)$$

where $\overrightarrow{vel}(w,t)$ denotes the instantaneous center-of-mass velocity of the cell w at time t, $\lambda_{\text{inertia}}(w)$ is a cell-specific parameter, and Δt is one or more Monte Carlo steps (Balter et al. 2007). Since the increment of the Hamiltonian depends on the proposed transition, this is a kinetic extension of the CPM.

Cell shapes arise in the CPM implicitly from satisfying the volume constraint. In the two-dimensional CPM, cells adopt approximately hexagonal shapes, producing a space tiling pattern comparable to epithelial tissues. Elongated cell shapes can be modeled by imposing a cell length constraint which renders the major axis of the ellipsoidal approximation of the cell's shape to be close to a predefined target length or ratio (Zajac et al. 2003). Rod cell shapes with particular stiffness have been modeled using a compartmentalized cell concept, where each cell consists of a row of standard CPM cells (Starruß et al. 2007).

Chemotactic response to some field $c{:}S \to [0,\infty)$ of signals can be modeled in the simplest form by an addend $H_{chemo} = \sum_{w \in W \setminus \{0\}} \lambda_{chemo}(w) \sum_{x \in cell_w} c(x)$ to the Hamiltonian, where λ_{chemo} is possibly a cell-type-specific chemotactic response parameter (Glazier et al. 2007). If $\lambda_{\text{chemo}} < 0$, the cells prefer to move up the chemotactic gradient, for $\lambda_{\text{chemo}} > 0$ they prefer to move down the gradient. There have been several more refined extensions to the CPM that model chemotaxis (Glazier et al. 2007). One example is the following kinetic extension used by Savill and Hogeweg (1997), where the positions of the target spin x and the trial spin y in a proposed transition $\eta \to \eta_x^y$ are taken into account,

$$\Delta H_{\text{chemo}} = \sum_{w \in W} \lambda_{\text{chemo}}(w)(c(y) - c(x)). \quad (6)$$

Hybrid and multiscale modeling: The CPM can be coupled to auxiliary formalisms, typically using systems of differential equations. A hybrid approach enables multiscale modeling in which molecular species are represented as continuous quantities, and cells are treated as discrete entities. For instance, CPM parameters pertaining to cellular properties can be under the control of ordinary differential equations, representing subcellular processes such as gene regulation. CPM cell behavior can also be linked to lattice-based reaction-diffusion systems representing the biochemical microenvironment through, for example, chemotaxis. A similar approach can be adopted to spatially represent the intracellular biochemistry that exerts influence on the protrusions and retractions in the CPM by kinetic modulation of transition probabilities (Marée et al. 2006).

Implementations

When applied to specific biological problems, the CPM framework is typically used with several extensions and modifications. Its analysis comprises extensive numerical simulation studies. In an effort to provide a common implementation for CPM simulations, CompuCell3D has been developed (Glazier et al. CompuCell3D). This open source software implements a large number of common CPM extensions and provides a graphical modeling interface.

Limitations and Merits

From a theoretical perspective, the CPM is poorly understood. Hence, the analysis of CPM models can effectively only be performed by numerical simulation. Important mathematical methods, such as rigorous spatiotemporal limit procedures to derive the laws that guide the behavior of certain macroscopic variables, are not yet available. Since the classical Metropolis algorithm (▶ Metropolis algorithm) is modified in the CPM, these models differ in essential aspects from classical equilibrium models. In addition, CPMs have been criticized because their calibration is often nontrivial. Cellular behaviors are specified in an indirect or phenomenological manner via the Hamiltonian and the modified Metropolis algorithm. Consequently, the relation between the parameters that control the dynamics of the CPM and the biological-physical quantities they represent often remains allusive.

Despite these limitations, the CPM formalism has found applications in many topics, mainly in the field of developmental biology. Its spatial and cell-centered nature renders it suitable for the study of phenomena where a mesoscopic description of individual cell shape and motility is important. It provides a flexible modeling framework that allows incorporation of problem-specific extensions. Moreover, coupling the CPM to auxiliary model formalisms enables the exploration of the complex interplay between several factors at different biological scales, acting at the intracellular, the intercellular, and the tissue level.

Cross-References

- ▶ Collective Behavior
- ▶ Differential Adhesion Hypothesis
- ▶ Markov Chain
- ▶ Metropolis Algorithm
- ▶ Spatiotemporal Pattern Formation

References

Balter A, Merks RMH, Poplawski NJ, Swat M, Glazier JA (2007) The Glazier-Graner-Hogeweg model: extensions, future directions, and opportunities for further study. In: Anderson ARA, Chaplain MAJ, Rejniak KA (eds) Single cell-based models in biology and medicine, Mathematics and biosciences in interaction. Birkhäuser Verlag, Basel, pp 151–167

Glazier JA et al. CompuCell3D, an open source modeling environment, www.compucell3d.org

Glazier JA, Graner F (1993) Simulation of the differential adhesion driven rearrangement of biological cells. Phys Rev E 47(3):2128–2154

Glazier JA, Balter A, Poplawski NJ (2007) Magnetization to morphogenesis: a brief history of the Glazier-Graner-Hogeweg model. In: Anderson ARA, Chaplain MAJ, Rejniak KA (eds) Single cell-based models in biology and medicine, Mathematics and biosciences in interaction. Birkhäuser Verlag, Basel, pp 79–106

Marée AFM, Jilkine A, Dawes A, Grieneisen VA, Edelstein-Keshet L (2006) Polarization and movement of keratocytes: a multiscale modelling approach. Bull Math Biol 68(5):1169–1211. doi:10.1007/s11538-006-9131-7

Ouchi NB, Glazier JA, Rieu J, Upadhyaya A, Sawada Y (2003) Improving the realism of the cellular Potts model in simulations of biological cells. Physica A 329(3–4):451–458

Savill NJ, Hogeweg P (1997) Modelling morphogenesis: from single cells to crawling slugs. J Theor Biol 184(3):229–235

Starruß J, Bley T, Sogaard-Andersen L, Deutsch A (2007) A new mechanism for collective migration in Myxococcus xanthus. J Stat Phys 128(1–2):269–286

Zajac M, Jones GL, Glazier JA (2003) Simulating convergent extension by way of anisotropic differential adhesion. J Theor Biol 222:247–259

Cellular Spaces

▶ Cellular Automata

Cellular Structures

▶ Cellular Automata

CE-MS-based Metabolomics

Tamaki Fujimori
Human Metabolome Technologies Inc., Tsuruoka, Yamagata, Japan

Synonyms

Capillary electrophoresis mass spectrometry–based metabolomics

Definition

Capillary electrophoresis (CE) mass spectrometry (MS) is an analytical technique. CE is superior in separation efficiency of ionic metabolites. MS can provide high sensitivity. The combination of CE with MS can lead to achieve the analytical platform with high separation capacity, resolution, and sensitivity. CE-MS has been often used for measurements of ionic metabolites in metabolomics. CE-MS analysis is not suited to measurements of neutral metabolites. CE-MS analysis is a complementary tool to reversed-phase LC-MS analysis which is specific to measurements of neutral metabolites.

Characteristics

Selection of Analytical Machines

Most of the metabolites in primary metabolism are ionic, hydrophilic, and polar. For example, metabolites of glycolysis, TCA cycle, amino acid metabolism, and nucleotide synthesis show these characteristics. Therefore, CE-MS analysis is suitable for measurements of metabolites in primary metabolism. Metabolomics using CE-MS analysis could lead to understandings of biochemical and biological mechanisms in primary metabolism. On the contrary, CE-MS analysis is not suited to measurements of neutral metabolites because they cannot be separated in CE. LC-MS is suitable for measurements of neutral metabolites. CE-MS analysis is a complementary tool to LC-MS analysis. Whether the usage of CE-MS or LC-MS is selected in metabolomics depends on the target metabolites.

CE

CE separates polar metabolites on the basis of their mass-to-charge ratios. The separation capacity of CE is superior to that of gas chromatography (GC) or liquid chromatography (LC). Although the sensitivity is low in CE, it can be sufficiently compensated by high sensitivity in MS. Low-molecular-weight metabolites in biological samples can be analyzed in two modes for cationic and anionic metabolites.

After sample injection, the voltage is applied to the capillary. Cationic metabolites migrate toward the cathode side and anionic metabolites migrate toward the anode side on the basis of electrochemical property. The migration speed of metabolites is determined by hydrated ionic radius and valence. Because even the metabolites with the same mass can be separated based on the nature of CE, the separation of structural isomers is possible. For example, glucose 1-phosphate, glucose 6-phosphate, and fructose 6-phosphate can be easily separated in CE. The CE analysis is also advantageous in terms of small volume requirement (a few nanoliters) for sample injection.

TOFMS

Time-of-flight mass spectrometry (TOFMS) is often used as MS in CE-MS analysis. Accelerated ions by applied voltage reach the detector under high vacuum. Ions with low m/z reach the detector earlier than those with high m/z. The value of m/z is calculated on the basis of the arrival time. TOFMS provides relatively high resolution, which enables estimation of composition formula.

Peak Annotation

Detected peak information including migration time in CE and m/z is obtained in CE-TOFMS analysis. Peaks of putative metabolites are assigned to those of known metabolites according to the peak information. Human Metabolome Technologies Inc. possesses library information about migration time and m/z of about 900 metabolites. These metabolites can be identified and quantified in reference to library information. Most amino acids, organic acids, sugar phosphates, and nucleotides are included in the library.

Metabolomics Application

CE-MS based metabolomics is applied to studies in many fields including basic sciences, medical sciences, nutrition, agriculture, and toxicology. Metabolomics could lead to elucidating of gene, enzyme, biomarker, and metabolic network. Specifically, CE-MS based metabolomics can depict the details of primary metabolism which consists of many ionic metabolites. Several applications are introduced below.

Soga et al. identified an oxidative biomarker in hepatotoxicity by CE-MS analysis. The changes in liver metabolites following acetaminophen-induced hepatotoxicity were examined. Ophthalmic acid increased in conjunction with glutathione

consumption in liver after the addition of acetaminophen. The changes were also observed in serum. Ophthalmic acid might be a new biomarker for hepatic oxidative stress.

Ishii et al. performed multi-omics (transcriptomics, proteomics, and CE-MS-based metabolomics) analysis to examine metabolic regulation of *Escherichia coli* in response to gene disruption or changes in growth rate. Metabolite levels were stable compared to gene expression and protein levels. The results showed robustness of metabolic network against perturbation such as gene disruption and changes in growth rate.

Ohashi et al. examined metabolome changes in histidine-starved *E. coli* by CE-MS-based metabolomics. The analysis quantified 198 metabolites among 375 charged and hydrophilic metabolites in primary metabolism. In *E. coli*, histidine starvation induced changes in many metabolic pathways including glycolysis, TCA cycle, amino acid metabolism, and nucleotide biosynthesis.

References

Ishii N, Nakahigashi K, Baba T, Robert M, Soga T, Kanai A, Hirasawa T, Naba M, Hirai K, Hoque A, Ho PY, Kakazu Y, Sugawara K, Igarashi S, Harada S, Masuda T, Sugiyama N, Togashi T, Hasegawa M, Takai Y, Yugi K, Arakawa K, Iwata N, Toya Y, Nakayama Y, Nishioka T, Shimizu K, Mori H, Tomita M (2007) Multiple high-throughput analyses monitor the response of *E. coli* to perturbations. Science 316(5824):593–597

Monton MR, Soga T (2007) Metabolome analysis by capillary electrophoresis-mass spectrometry. J Chromatogr A 1168(1–2):237–246

Ohashi Y, Hirayama A, Ishikawa T, Nakamura S, Shimizu K, Ueno Y, Tomita M, Soga T (2008) Depiction of metabolome changes in histidine-starved *Escherichia coli* by CE-TOFMS. Mol Biosyst 4(2):135–147

Ramautar R, Somsen GW, de Jong GJ (2009) CE-MS in metabolomics. Electrophoresis 30(1):276–291

Ramautar R, Mayboroda OA, Somsen GW, de Jong GJ (2011) CE-MS for metabolomics: developments and applications in the period 2008–2010. Electrophoresis 32(1):52–65

Soga T, Baran R, Suematsu M, Ueno Y, Ikeda S, Sakurakawa T, Kakazu Y, Ishikawa T, Robert M, Nishioka T, Tomita M (2006) Differential metabolomics reveals ophthalmic acid as an oxidative stress biomarker indicating hepatic glutathione consumption. J Biol Chem 281(24):16768–16776

Cene

Eric Werner
Department of Physiology, Anatomy and Genetics, University of Oxford, Oxford, UK
Department of Computer Science, University of Oxford, Oxford, UK
Oxford Advanced Research Foundation, Fort Myers, FL, USA

Synonyms

Cancer networks; Cenome; Developmental control networks; Developmental networks; Stem cell networks

Definition

A cene is a developmental network (▶ Developmental Control Networks) that controls the development of multicellular organisms. The nodes in the cene are cell control states. Edges in the network denote cell actions including jumps to new cell states. There are branches in the network that denote cell division where each daughter cell enters a possibly new control state. Some branches stand for stochastic changes of control states. Other branches describe cell signaling protocols (Werner 2011a, b).

Characteristics

Cenes can be linked together to form larger cenes. The global developmental control network in a genome is called the ▶ cenome. The topology of the cene determines its ideal dynamic phenotype.

Examples of cenes include ▶ stem cell networks, ▶ cancer networks and terminal, and progenitor cell ▶ developmental control networks (for details see Werner 2011a, b).

Cenes Are Executable Networks

Developmental control networks or cenes are executable networks. The cell has an interpretive executive system, the IES, that interprets and executes the directives in developmental control networks. This system co-evolved with the developmental control networks (Werner 2011a).

Sub-cenes Within Cenes

Any cene can link to another cene. This means the link has further developmental progeny generated by that sub-cene. In this way more and more complex cenes are built up.

Evolution of Cenes and Multicellular Organisms

The evolution of multicellular organisms is directly linked to the increasing complexity of their cenes. These networks are an autonomous layer on top of normal gene networks (Werner 2011a, b).

Cross-References

- ▶ Cancer Networks
- ▶ Cenome
- ▶ Developmental Control Networks
- ▶ Stem Cell Networks

References

Werner E (2011a) On programs and genomes, arXiv:1110.5265v1 [q-bio.OT]. http://arxiv.org/abs/1110.5265

Werner E (2011b) Cancer networks: a general theoretical and computational framework for understanding cancer, arXiv:1110.5865v1 [q-bio.MN]. http://arxiv.org/abs/1110.5865v1

Cenes

- ▶ Cancer Networks
- ▶ Developmental Control Networks

Cenome

Eric Werner
Department of Physiology, Anatomy and Genetics, University of Oxford, Oxford, UK
Department of Computer Science, University of Oxford, Oxford, UK
Oxford Advanced Research Foundation, Fort Myers, FL, USA

Synonyms

Cene; Developmental control networks

Definition

A cenome is the global developmental network (▶ Developmental Control Networks) that controls the development of a multicellular organism. The cenome is a type of global ▶ Cene. The nodes in the cenome are cell control states. Edges in the network denote cell actions including jumps to new cell states. There are branches in the network that denote cell division where each daughter cell enters a possibly new control state. Some branches stand for stochastic changes of control states. Other branches describe cell signaling protocols (Werner 2011a, b).

Characteristics

Cenomes consist of linked-together ▶ developmental control networks or ▶ cenes. The global developmental control network in a genome is called the Cenome. The topology of the cenome determines part of its ideal dynamic phenotype. Since most genes are shared between organisms they cannot be responsible for the unique morphologies and functional phenotypes of organisms. Cenes and the global cenome complement genetic processes with multicellular control processes.

The cenome consists of cenes (▶ Cene) that can include ▶ stem cell networks, ▶ cancer networks and terminal, and progenitor cell ▶ developmental control networks (for details see Werner 2011a, b).

Cenomes Are Executable Networks

The cenome of an organism is an executable network. Each cell in a multicellular system has an interpretive executive system, the IES, that interprets and executes the directives in the cenome. The IES co-evolved with the cenome (Werner 2011a).

Cenome and Cene Control Have a Subsumption Architecture

The cenome control network is a higher layer that subsumes lower level standard gene networks. Lower level gene networks allow the cell to react to its local environment, while higher level cenes,

which make up the cenome, guide the global development of the embryo.

Evolution of Cenes and Multicellular Organisms

Any cene can link to another cene. This means the link has further developmental progeny generated by that sub-cene. In this way, more and more complex cenes are built up within a cenome. The evolution of multicellular organisms is directly linked to the increasing complexity of their cenome (Werner 2011a, b).

Cross-References

- ▶ Cancer Networks
- ▶ Cene
- ▶ Developmental Control Networks
- ▶ Stem Cell Networks

References

Werner E (2011a) On programs and genomes, arXiv:1110.5265v1 [q-bio.OT]. http://arxiv.org/abs/1110.5265

Werner E (2011b) Cancer networks: a general theoretical and computational framework for understanding cancer, arXiv:1110.5865v1 [q-bio.MN]. http://arxiv.org/abs/1110.5865v1

Centromere

Rosella Visintin
IEO, European Institute of Oncology,
Milan, Italy

Definition

A region on the DNA, usually localized in the center of chromosomes, where sister chromatids are closely associated.

Cross-References

- ▶ Mitosis

Ceteris Paribus Laws

Max Kistler
IHPST, Université Paris 1 Panthéon-Sorbonne, Paris, France

Definition

Regularities in ▶ special sciences typically have exceptions. A law expressing a regularity that has exceptions is sometimes called a "ceteris paribus law" (or "cp-law"). A situation is an exception to a law if it does not correspond to the regularity expressed by the law, but is not taken to refute the law, because the exception can be explained to be the result of interferences that follow (other) laws. A cp-law expresses a regularity that exists in "normal" situations, or, literally, if "all else is equal," i.e., if there are no perturbations resulting in an exception. In classical genetics, Mendel's law of segregation says that, in sexually reproducing organisms, during gamete formation each member of an allelic pair separates from the other member to form the genetic constitution of an individual gamete, so that there is a 50:50 ratio of alleles in the mass of the gametes. However, some sets of gametes are exceptional with respect to this law because some organisms undergo "meiotic drive" leading to an overproduction of gametes with one allele at the expense of gametes with the other allele. No strict law is true of real systems whose evolution is subject to perturbations. One proposal for understanding ceteris paribus laws is to take them to be strictly universal generalizations that are true of *ideal* systems. However, the status of such "ideal systems" is problematic. It is not clear how laws describing systems that are not real can help explain and predict real systems. An alternative is to distinguish between *laws of nature* and laws "in situ" (Cummins 2000), which are specific for certain systems that Cartwright (1999) calls "nomological machines." *Laws of natur*e determine relations between different properties of objects and have unrestricted scope: All massive objects obey Newton's law of universal gravitation. However, no such law determines directly and by itself the evolution of any particular system: To obtain the equation of motion of a particular system, one has to build a model that

takes into account all properties of the objects composing the system and all forces due to these properties. To find the equation of motion of a charged massive particle p such as a biological molecule, one has to describe all massive and charged particles q_i in the environment of the object, then calculate the forces that the objects q_i exert on p. The motion of the molecule is determined by the sum of all these forces. Exceptions to generalizations bearing on particular systems can then be conceived as resulting from the neglect of some properties of p, or of relevant objects q_i in its environment.

References

Cartwright N (1999) The dappled world, A study of the boundaries of science. Cambridge University Press, Cambridge
Cummins R (2000) How does it work? vs. What are the laws? Two conceptions of psychological explanation. In: Keil F, Wilson R (eds) Explanation and cognition. MIT Press, Cambridge, pp 117–145

CFSE

▶ Modeling, Cell Division and Proliferation

Chain Type

▶ IMGT-ONTOLOGY, ChainType

Change of Antigenic Properties by Mutations

▶ Antigenic Drift and Shift

Change of Antigenic Properties by Rearrangement of Viral Genome Segments

▶ Antigenic Drift and Shift

Characteristic Path Length

Falk Schreiber
Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), OT Gatersleben, Stadt Seeland, Germany
Martin Luther University Halle–Wittenberg, Halle, Germany

Synonyms

Average path length

Definition

The characteristic path length $l(G)$ of a ▶ graph $G = (V, E)$ is defined as the average number of edges in the shortest paths between all vertex pairs given by

$$l(G) = \frac{1}{|V| * (|V| - 1)} \sum_{\upsilon \in V} \sum_{\upsilon' \in V \setminus \{\upsilon\}} spl(\upsilon, \upsilon'), \quad (1)$$

where $spl(\upsilon,\upsilon')$ gives the number of edges in a shortest path between vertices υ and υ'. In case of an unconnected graph, the characteristic path length is infinite, as the number of edges between two unconnected vertices is considered infinite. In this case, formula (1) is often modified to sum over just all connected vertex pairs. Alternatively, other measures such as the harmonic mean or the average inverse path length can be used.

References

Junker BH, Schreiber F (2008) Analysis of Biological Networks. Wiley Series on Bioinformatics, Computational Techniques and Engineering, Wiley, 368
Watts D, Strogatz S (1998) Collective dynamics of small-world networks. Nature 393:440–442

Checkpoints

▶ Cell Cycle Dynamics, Irreversibility

Chemical Master Equation

Hao Ge[1] and Hong Qian[2]
[1]School of Mathematical Sciences and Centre for Computational Systems Biology, Fudan University, Shanghai, China
[2]Department of Applied Mathematics, University of Washington, Seattle, WA, USA

Synonyms

Gillespie algorithm; Stochastic chemical kinetics

Definition

Chemical master equation is the stochastic counterpart of the chemical kinetic equation based on the law of mass action. It describes the kinetics of chemical reactions in a rapidly stirred tank with small volume in terms of stochastic reaction times giving rise to fluctuating copy numbers of reaction species.

Consider a system of fixed volume V at constant temperature T. Let there be well-stirred mixture of $N \geq 1$ molecular species $\{S_1, \ldots, S_N\}$ and $M \geq 1$ reactions $\{R_1, \ldots, R_M\}$. One specifies the dynamical state of this system by $X(t) = (X_1(t), \ldots, X_N(t))$, where $X_i(t)$ is the copy number of molecular species S_i in the system at time t.

One describes the time evolution of X(t) from some given initial state $X(t_0) = x_0$. Both single-molecule experimental measurements and theoretical investigations have shown that X(t) is a stochastic process because the time at which a particular reaction occurs is random.

The chemical master equation kinetics assumes that the system is well stirred such that at any moment each reaction occurs with equal probability at any position in space. Furthermore, it assumes that for each reaction R_j, there is a corresponding rate function r_j and a stoichiometry vector $v_j = (v_{j1}, \ldots, v_{jN})$, which are defined as

$$\begin{aligned} &r_j(x)dt = \text{the probability, given } X(t) = x, \\ &\text{that one reaction } R_j \text{ will occur some where} \\ &\text{in the next infinitestimal time interval } [t, t+dt), \\ &(j = 1, \cdots, M). \end{aligned} \tag{1}$$

and

$$\begin{aligned} &v_{ji} = \text{the change in the number of } S_i \text{ molecules} \\ &\text{caused by one } R_j \text{ reaction}, (j = 1, \cdots, M; i = 1, \cdots, N). \end{aligned} \tag{2}$$

The stoichiometry vector $\{v_{ji}\}$ can be obtained from the difference between the numbers of a molecular species that are consumed and produced in the reaction.

Exact description of the rate function associates with a reaction that can be either from phenomenological models for stochastic chemical kinetics or from more fundamental molecular physics based on the concept of elementary reactions. In general, the function r_j has the mathematical form:

$$r_j(x) = k_j h_j(x). \tag{3}$$

Here k_j is the specific probability rate constant for reaction R_j, which is defined such that $k_j dt$ is the probability that a randomly chosen combination of the reactants of R_j will react accordingly in the next infinitesimal time interval dt. k_j is intimately related to the rate constants in the traditional law of mass action kinetics.

The function $h_j(x)$ in Eq. 3 measures the number of distinct combinations of R_j reactant molecules available in the state x. It is a combinatorial factor that can be easily obtained from the reaction R_j, i.e., $\mathbf{h_j(x)} = \prod_{k=1}^{N} \frac{x_k!}{m_{jk}!(x_k - m_{jk})!}$. In the transitional mass-action kinetics, this term is related to the product of the concentrations of all reactants.

In general, for a chemical reaction

$$R_j : m_{j1}S_1 + \cdots + m_{jN}S_N \rightarrow n_{j1}S_1 + \cdots + n_{jN}S_N, \tag{4}$$

one has

$$\begin{aligned} v_{ji} &= n_{ji} - m_{ji}, \\ r_j(x) &= k_j \prod_{k=1}^{N} \frac{x_k!}{m_{jk}!(x_k - m_{jk})!} \end{aligned} \tag{5}$$

If for any k and j, have $x_k >> m_{jk}$, then approximately

$$r_j(x) = k_j \prod_{k=1}^{N} \frac{x_k{}^{m_{jk}}}{m_{jk}!} \tag{6}$$

Then

$$\frac{r_j(x)}{V} = \left(\frac{k_j V^{n-1}}{\prod_{k=1}^{N} m_{jk}!} \right) \prod_{k=1}^{N} \left(\frac{x_k}{V} \right)^{m_{jk}}, \; n = \sum_{k=1}^{N} m_{jk}. \tag{7}$$

is precisely the reaction flux per unit volume in the mass-action kinetics; the (x_k/V) is the concentration of species k, and the term in the parenthesis, $k_j' = \frac{k_j V^{n-1}}{\prod_{k=1}^{N} m_{jk}!}$, is the reaction rate constant in the traditional chemical kinetics. Here, k_j is regarded as "stochastic reaction constant," while k_j' is the corresponding reaction constant for Law of Mass Action.

The reaction rate constant is deduced only from experiments, or calculated based on Kramers-Marcus theory. It is usually a function of temperature but is independent of the volume of a reaction system. However, the stochastic reaction constant k_j depends on the system volume and the temperature.

From the rate function given from above, the state vector X(t) is a Markov jump process on the nonnegative N-dimensional integer lattice.

Any Markov process can be described by two completely different, complementary mathematical methods. One follows its stochastic trajectories and one consider its probability distribution function changing with time. The Gillespie algorithm gives the former, and the chemical master equation is for the latter.

In the chemical master equation perspective, one no longer asks what are the copy number (or concentration) of species i at time t, but rather what is the probability of the system having x_i copies of species X_i at time t.

We focus on the probability

$$P(x,t) = \Pr\{X(t) = x\}, \tag{8}$$

and now derive its evolutionary equations, i.e., the chemical master equation.

We take a time increment dt and consider the variation between the probability of X(t) = x and of X(t + dt) = x. This variation is

$$\begin{aligned} P(x,t+dt) - P(x,t) = {} & \text{Increasing of the probability in dt} \\ & - \text{Decreasing of the probability in dt} \end{aligned} \tag{9}$$

We take dt so small such that the probability of having two or more reactions in dt is negligible compared to the probability for only one reaction. Then increasing of the probability in dt occurs when a system with state X(t) = x − v_j reacts according to R_j in (t, t + dt), the probability of which is $r_j(x - v_j)dt$. Thus,

$$\text{Increasing of the probability in dt} = \sum_{j=1}^{M} r_j(x - v_j) P(x - v_j, t)\, dt. \tag{10}$$

Similarly, when a system with state X(t) = x reacts according to any reaction channel R_j in (t, t + dt), the probability P(x, t) will decrease. Thus,

$$\text{Decreasing of the probability in dt} = \sum_{j=1}^{M} r_j(x) P(x,t)\, dt. \tag{11}$$

Substituting Eqs. 10 and 11 into Eq. 9, we obtain

$$\begin{aligned} P(x,t+dt) - P(x,t) = {} & \sum_{j=1}^{M} r_j(x - v_j) P(x - v_j, t)\, dt \\ & - \sum_{j=1}^{M} r_j(x) P(x,t)\, dt, \end{aligned} \tag{12}$$

which yields, with the limit dt → 0, the chemical master equation:

$$\begin{aligned} \frac{\partial}{\partial t} P(x,t) = {} & \sum_{j=1}^{M} r_j(x - v_j) P(x - v_j, t)\, dt \\ & - \sum_{j=1}^{M} r_j(x) P(x,t)\, dt. \end{aligned} \tag{13}$$

Characteristics

Consistent with Law of Mass Action

The relationship between deterministic law of mass action and stochastic chemical master equation for chemical reactions was established by T. Kurtz in a general theory in which a stochastic Markov chain model for general chemical reaction is studied alongside its deterministic counterpart. Recall that both Gillespie algorithm and the chemical master equation are two different descriptions of the same stochastic, jump process. Let $X^V(t)$ denote this stochastic process, where V is the volume of the system, and the initial condition (in the thermodynamic limit) is

$$\lim_{V\to\infty} \frac{X^V(0)}{V} = x_0 \tag{14}$$

Then, the solution to the initial value problem of the corresponding chemical kinetics based on the law of mass action model is denoted by $c(t,x_0)$. Kurtz has shown that the relationship between these two solutions is

$$\lim_{V\to\infty} \Pr\left\{ \sup_{s\le T} \left| \frac{X^V(t)}{V} - c(t,x_0) \right| > \varepsilon \right\} = 0, \tag{15}$$

for every T and $\varepsilon > 0$,

It is important to note the mathematical subtlety of $s \le T$. The above result is only valid for finite time interval, even though it can be as long as one likes: The ε is a function of T; greater the T, larger the ε.

Compare the solution to the chemical master equation with initial distribution concentrated at x_0, $P(x,t|x_0)$ and the deterministic kinetics $c(t,x_0)$; if the latter approaches to a steady-state value then the former cumulates probability at the steady state. In general, there is an agreement between the steady states of the deterministic kinetics and the peaks of the stationary distribution of the chemical master equation.

However, in the limit of infinitely large volume, the situation is quite different: The peaks of the stationary distribution of chemical master equation may not be consistent with the stable fixed points of the mass-action kinetics. This is exactly due to the $s \le T$ in Eq. 15. This can be explained by a theory of multiple timescales.

Numerical Simulation: Gillespie Algorithm

We now introduce stochastic simulation methods that generate the random trajectories of X(t). Once we have enough sample trajectories of a stochastic process, we will be able to calculate the probability distribution function P(x, t) and all other statistical behaviors including the mean trajectory, variances, and correlations.

The stochastic simulation algorithm (SSA) (also known as the Gillespie algorithm) is based on the following next-reaction probability distribution: assume the system is in state x at time t, and let

$$\begin{aligned} &p(\tau, j|x,t) = \text{probability that, given } X(t) = x, \\ &\text{the next reaction will occur in the} \\ &\text{infinitesimal time interval } [t+\tau, t+\tau+d\tau), \\ &\text{and will be the } R_j \text{ reaction.} \end{aligned} \tag{16}$$

In probability theory, there is an elementary theorem, which states that

$$p(\tau, j|x,t) = r_j(x)\exp(-r_0(x)\tau),\ 0 \le \tau < \infty, j = 1,\ldots,M \tag{17}$$

where

$$r_0(x) = \sum_{k=1}^{M} r_k(x) \tag{18}$$

The key step of this simulation method is to generate the pair of numbers (τ, j) in accordance with the probability Eq. 18: First generate two random numbers a_1 and a_2 from the uniform distribution in the unit interval, and then take

$$\tau = \frac{-\ln(a_1)}{r_0(x)},$$

$$j = \text{the smallest integer satisfying } \sum_{i=1}^{j} r_i(x) > a_2 r_0(x) \tag{19}$$

Below are main steps in the stochastic simulation algorithm.

1. Initialization: Let the initial state as $X = X_0$, and $t = t_0$.
2. Simulation: Generate a pair of random numbers (τ, j) according to the probability density function (17).
3. Update: Increase the time by τ, and replace the molecule numbers by $X + v_j$.
4. Iterate: Go back to Step 2 unless the end of the simulation procedure.

The exact stochastic simulation algorithm consists with the chemical master equation and gives an exact sample trajectory of the real system.

Cross-References

- ▶ Fokker–Planck Equation
- ▶ Law of Mass Action
- ▶ Markov Chain
- ▶ Stochastic Differential Equation

References

Deard DA, Qian H (2008) Chemical biophysics: quantitative analysis of cellular systems, Cambridge texts in biomedical engineering. Cambridge University Press, Cambridge (Chap 11)

Ge H, Qian H (2009) Thermodynamic limit of a nonequilibrium steady-state: Maxwell-type construction for a bistable biochemical system. Phys Rev Lett 103:148103

Gillespie DT (1977) Exact stochastic simulation of coupled chemical reactions. J Phys Chem 81:2340–2361

Kurtz TG (1972) The relationship between stochastic and deterministic models for chemical reactions. J Chem Phys 57:2976–2978

Lei JZ (2010) Stochastic modeling in systems biology. J Adv Math Appl (to appear)

Liang J, Qian H (2010) Computational cellular dynamics based on the chemical master equation: a challenge for understanding complexity. J Comput Sci Technol (Rev) 25:154–168

McQuarrie DA (1967) Stochastic approach to chemical kinetics. J Appl Probab 4:413–478

Qian H, Bishop LM (2010) The chemical master equation approach to nonequilibrium steady-state of open biochemical systems: linear single-molecule enzyme kinetics and nonlinear biochemical reaction networks. Int J Mol Sci 11(9):3472–3500

Vellela M, Qian H (2007) Quasistationary analysis of a stochastic chemical reaction: Keizer's paradox. Bull Math Biol 69:1727–1746

Vellela M, Qian H (2009) Stochastic dynamics and nonequilibrium thermodynamics of a bistable chemical system: the Schlögl model revisited. J R Soc Interface 6:925–940

C

ChIP

- ▶ Chromatin Immunoprecipitation

ChIP-on-Chip Assay

Nobuo Shimamoto
Faculty of Life Sciences, Kyoto Sangyo University, Kyoto, Japan

Definition

ChIP-on-CHIP Assay is a method to determine the location of the genomic DNA to which a DNA-binding protein binds. The protein is cross-linked to DNA, usually in vivo, and the DNA is fragmented. The DNA fragments cross-linked with the protein are then immunoprecipitated, and hybridized to DNA chip or directly sequenced to determine the location.

Cross-References

- ▶ Transcription in Bacteria

Chi-Squared Test

Larissa Stanberry
Bioinformatics and High-throughput Analysis Laboratory, Seattle Children's Research Institute, Seattle, WA, USA

Synonyms

Pearson's chi-squared test

Definition

The chi-squared test of goodness of fit evaluates whether the observed frequencies differ from theoretical values.

Characteristics

The chi-squared test of independence is applied to count data that could be represented by the $I \times J$ contingency table. For example, counting sightings of three different bird species in four different locations would result in a 3×4 table, where each cell frequency n_{ij} gives the number of sightings of the i-th species in the j-th location. One might be interested to evaluate whether the abundance of species differ between the locations. The test statistic is given by:

$$X^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{n_{ij} - \mu_{ij}}{\mu_{ij}},$$

where μ_{ij} are the expected frequencies under the null hypothesis. The expected frequencies are typically unknown, but could be estimated by $\hat{\mu}_{ij} = \sum_i n_{ij} \sum_j n_{ij}/n$, where n is the total number of counts. Substituting $\hat{\mu}_{ij}$ in place of μ_{ij}, we obtain that, asymptotically, the resulting test statistic X^2 follows a chi-squared distribution with $(I-1)(J-1)$ degrees of freedom (Agresti 2002).

Conclusions

The chi-square test provides evidence of association between the variables; however, it does not indicate how the variables relate to each other. To truly understand the nature of the association, the chi-squared test alone is not sufficient and should be followed by more detailed analysis, i.e., residual analysis, decomposition, odds ratios, etc. The chi-squared test relies on asymptotic distribution and is applicable only when the cell frequencies are sufficiently large. For small sample sizes, the Fisher's Exact Test can be used. Furthermore, the chi-squared test is invariant with respect to reodering of rows and columns. Hence, if one of the variables is ordinal, the appropriate statistics that respect the ordinality should be chosen.

References

Agresti A (2002) Categorical data analysis. Wiley, Hoboken

Chondrocytes

Steven D. Rhodes
School of Medicine, Indiana University, Indianapolis, IN, USA

Definition

Chondrocytes are a special type of connective tissue cells that form cartilage. They are derived from mesenchymal stem cells of the bone marrow cavity and reside in joints where cartilage is required to cushion bone-on-bone friction. Chondrocyte differentiation can be induced in vitro from mesenchymal stem cell cultures in the presence of ascorbic acid plus cytokines such as transforming growth factor beta 1 (TGF-β1) or TGF-β3. Chondrocytes can be defined phenotypically by the expression of Cbfa1/Runx2, Type-II collagen, Type-IX collagen, and Aggrecan. Histologically, mature chondrocytes stain positively with toluidine blue, signifying an abundance of proteoglycans within the extracellular matrix.

Cross-References

► Single Cell Assay, Mesenchymal Stem Cells

Chorioallantoic Membrane (CAM) Assay

Marsha A. Moses
Department of Surgery/Harvard Medical School, Vascular Biology Program/Children's Hospital Boston, Boston, MA, USA

Definition

The chorioallantoic membrane (CAM) assay is one of the earliest in vivo angiogenesis assays to be developed. The CAM of fertilized chicken eggs is

challenged with either potential stimulators or inhibitors of neovascularization by placement of polymers or other delivery systems impregnated with the test substances at a site on the CAM immediately above the subectodermal plexus (Ausprunk et al. 1975). Zones of vessel clearance indicate anti-angiogenic activity, whereas increased vessel development in a spokewheel configuration indicates angiogenic stimulation (Fernandez et al. 2003; Moses et al. 1990).

Cross-References

► Neovascularization

References

Ausprunk DH, Knighton DR, Folkman J (1975) Vascularization of normal and neoplastic tissues grafted to the chick chorioallantois. Role of host and preexisting graft blood vessels. Am J Pathol 79:597–618

Fernandez CA, Butterfield C, Jackson G, Moses MA (2003) Structural and functional uncoupling of the enzymatic and angiogenic inhibitory activities of tissue inhibitor of metalloproteinase-2 (TIMP-2). J Biol Chem 278:40989–40995

Moses MA, Sudhalter J, Langer R (1990) Identification of an inhibitor of neovascularization from cartilage. Science 248:1408–1410

Choristoma

Barbara J. Davis
Section of Pathology, Tufts Cummings School of Veterinary Medicine Biomedical Sciences, North Grafton, MA, USA

Definition

A heterotopic rest of cells in normally organized tissue in an abnormal site – common to find pancreatic tissue in the liver.

Cross-References

► Cancer Pathology

Chromatin

Vani Brahmachari and Shruti Jain
Dr. B. R. Ambedkar Center for Biomedical Research, University of Delhi, Delhi, India

Synonyms

Compact DNA, Nucleosomes

Definition

The genome which is a linear DNA molecule is very long relative to the size of the nucleus in the cells. For example, it is estimated that all the 23 pairs of chromosomes present in the human cells if completely stretched and lined up end to end, it will extend over 2 m, which is of many orders more than the size of the nucleus in the cell. However, the DNA is wrapped around a ball of proteins called the histones and further folded to be accommodated in the nucleus. This compact DNA-protein complex is called chromatin. The DNA-protein complex forms repetitive units called ► nucleosomes. The compaction state of chromatin is correlated to the transcriptional activity of DNA. The chromatin provides the platform for interaction with various regulatory factors that control transcription.

Cross-References

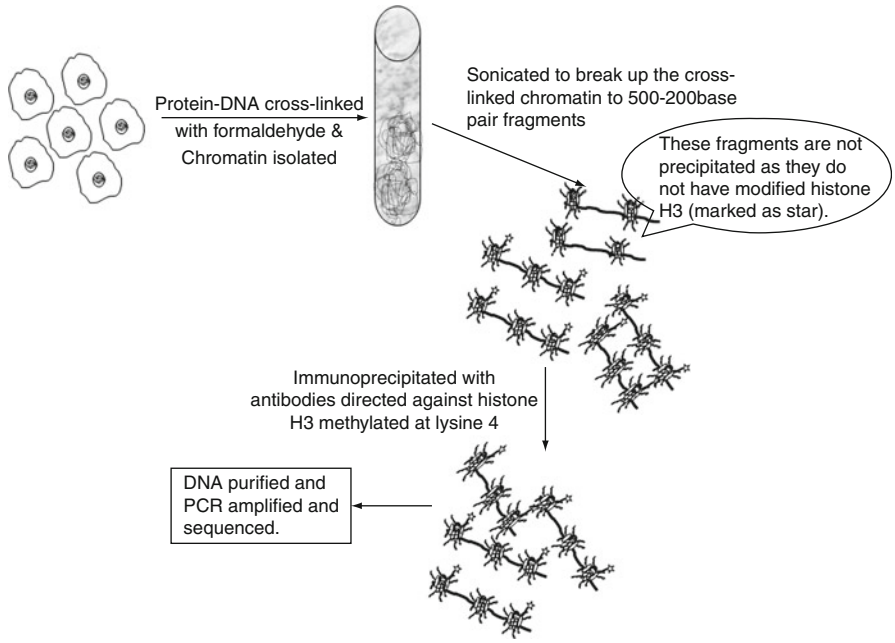
► Epigenetics
► Nucleosomes

Chromatin Immunoprecipitation

Vani Brahmachari and Shruti Jain
Dr. B. R. Ambedkar Center for Biomedical Research, University of Delhi, Delhi, India

Synonyms

ChIP



Chromatin Immunoprecipitation, Fig. 1 Outline of chromatin immunoprecipitation (ChIP). Immunoprecipitation with anti-H3K4me is shown as an example. The modification on histone is marked as a star

Definition

Chromatin immunoprecipitation (ChIP) is a type of immunoprecipitation technique used to analyze the protein-protein and DNA-protein interactions (Collas 2010). It aims to determine whether specific proteins, like transcription factors are associated with specific genomic regions like promoter elements or regulatory elements. It can also be used to determine the specific histone (► epigenetics) modifications associated at the target sites in the genome, indicating gene expression profiles.

The steps involved are as follows (Fig. 1):

(a) The cells are taken and the protein-DNA complexes (chromatin) are crosslinked by formaldehyde treatment.
(b) The DNA-protein complexes are then sheared and DNA fragments associated with the protein(s) of interest are selectively immunoprecipitated using antibody against the specific protein.
(c) The associated DNA fragments are purified and sequences are identified by amplification or DNA sequencing reactions.

Cross-References

► Epigenetics
► Epigenetics, Drug Discovery

References

Collas P (2010) The current state of chromatin immunoprecipitation. Mol Biotechnol 45:87–100

Chromatin Proteins

► Histones

Chromodomain Helicase DNA Binding (CHD)

Tetsuro Kokubo
Department of Supramolecular Biology, Graduate School of Nanobioscience, Yokohama City University, Yokohama, Kanagawa, Japan

Definition

In metazoans such as fly and human, the CHD class of ATP-dependent nucleosome-remodeling factors

includes many members, e.g., CHD1 and Mi-2/NuRD. In yeast, by contrast, the family has only a single member, CHD1 (Clapier and Cairns 2009) (summarized in Table 1).

CHD appears to function redundantly and/or cooperatively with ISWI in remodeling nucleosomes. For instance, CHD and ISWI are both required in yeast for nucleosome eviction from the *PHO5* promoter (Ehrensberger and Kornberg 2011) and for appropriate spacing of nucleosome arrays in the coding region of *ADH2* (Xella et al. 2006). CHD is involved in multiple steps of mRNA biogenesis, including transcriptional initiation, elongation, termination, and splicing.

Mi-2/NuRD is a multi-protein complex that contains HDACs and methyl CpG-binding proteins (see the definition "▶ mammalian HDAC"). This complex plays a crucial role in transcriptional repression rather than in activation.

Chromodomain Helicase DNA Binding (CHD), Table 1 Remodeler composition and orthologous subunits

		Organisms						
Family and composition		Yeast			Fly		Human	
SWI/SNF	Complex	SWI/SNF	RSC		BAP	PBAP	BAF	PBAF
	ATPase	Swi2/Snf2	Sth I		BRM/Brahma		hBRM or BRGI	BRGI
	Noncatalytic homologous subunits	Swi1/Adr6			OSA/ eyelid		BAF250/ hOSAI	
						Polybromo BAP170		BAF 180 BAF 200
		Swi3	Rsc8/Swh3		MOR/BAP155		BAF155, BAF170	
		Swp73	Rsc6		BAP60		BAF60a or b or c	
		Snf5	Sfh1		SNRI/BAP45		hSNF5/BAF47/INII	
					BAPI11/dalao		BAF57	
		Arp7, Arp9			BAP55 or BAP47		BAF53a or b	
					Actin		β-actin	
	Unique	[a]	[b]					
ISWI	Complex	ISW1a	ISW1b	ISW2	NURF	CHRAC ACF	NURF	CHRAC ACF
	ATPase	Isw1		Isw2	ISWI		SNF2L	SNF2H[c]
	Noncatalytic homologous subunits			Itc1	NURF301	ACF1	BPTF	hACFI/ WCRF180
						CHRAC14		hCHRAC17
						CHRAC16		hCHRAC15
					NURF55/ p55		RbAp46 or 48	
	Unique	Ioc3	Ioc2, Ioc4		NURF38			
CHD	Complex	CHD1			CHD1	Mi-2/NuRD	CHD1	NuRD
	ATPase	Chd1			dCHD1	dMi-2	CHD1	Mi-2α/CHD3, Mi-2β/CHD4
	Noncatalytic homologous subunits					dMBD2/3		MBD3
						dMTA		MTA1,2,3
						dRPD3		HDAC1,2
						p55		RbAp46 or 48
						p66/68		p66α,β
	Unique							DOC-17

(*continued*)

Chromodomain Helicase DNA Binding (CHD), Table 1 (continued)

		Organisms						
Family and composition		Yeast			Fly		Human	
INO80	Complex	INO80	SWR1	Pho-dINO80	Tip60	INO80	SRCAP	TRRAP/Tip60
	ATPase	Ino80	Swr1	dIno80	Domino	hIno80	SRCAP	p400
	Noncatalytic homologous subunits	Rvb1,2		Reptin, Pontin		RUVBL1,2/Tip49a,b		
		Arp5,8	Arp6	dArp5,8 dActin1	BAP55 Actin87E	BAF53a		
		Arp4, Actin1				Arp5,8	Arp6	Actin
		Taf14	Yaf9		dGAS41		GAS41	
		Ies2,6				hIes2,6		
			Swc4/ Eaf2		dDMAP1		DMAPI	
			Sw2/ Vps72		dYL-1		YL-1	
			Bdf1		dBrd8			Brd8/TRC/ p120
			H2AZ, H2B		H2Av,H2B		H2AZ, H2B	
			Swc6/ Vps71				ZnF-HITI	
					dTra1			TRRAP
					dTip60			Tip60
					dMRG15			MRG15 MRGX
					dEaf6			FLJI1730
					dMRGBP			MRGBP
					E(Pc)			EPCI, EPC-like
					dING3			ING3
	Unique	Ies1,les3-5, Nhp10	Swc3,5,7	Pho		[d]		

[a]Swp82, Taf14, Snf6, SnfI 1
[b]Rsc1 or Rsc2, Rsc3-5, 7, 9, 10, 30, Htl1, Ldb7, Rtt102
[c]In addition, SNF2H associates respectively with Tip5, RSF1, and WSTF to form **NoRC**, **RSF**, and **WICH** remodelers
[d]Amida, NFRKB, MCRS1, UCH37, FLJ90652, FLJ20309

Cross-References

► Mechanisms of Transcriptional Activation and Repression

References

Clapier CR, Cairns BR (2009) The biology of chromatin remodeling complexes. Annu Rev Biochem 78:273–304

Ehrensberger AH, Kornberg RD (2011) Isolation of an activator-dependent, promoter-specific chromatin remodeling factor. Proc Natl Acad Sci USA 108(25):10115–10120

Xella B, Goding C, Agricola E, Di Mauro E, Caserta M (2006) The ISWI and CHD1 chromatin remodelling activities influence ADH2 expression and chromatin organization. Mol Microbiol 59(5):1531–1541

CIA

► CIA/Asf1

CIA/Asf1

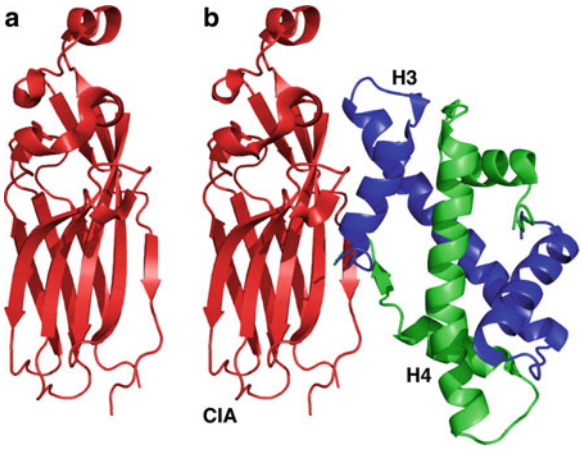
Toshiya Senda[1] and Naruhiko Adachi[2]
[1]Biomedicinal Information Research Centre (BIRC), National Institute of Advanced Industrial Science and Technology (AIST), Tokyo, Japan
[2]Structure-guided Drug Development Project, JBIC Research Institute, Japan Biological Informatics Consortium, Tokyo, Japan

Synonyms

Asf1; CIA

Definition

CIA (CCG1 interacting factor A)/Asf1 (anti-silencing function 1) is the most conserved histone chaperone in eukaryotes and plays important roles in various nuclear events such as transcription, replication, and DNA repair (Eitoku et al. 2008). The gene of this factor was first identified in genetic screening as a factor with an anti-silencing function. However, at that time, the biochemical activity of Asf1 was unknown. After a few years, two groups have independently rediscovered this factor as one relevant to nucleosome formation with histone H3–H4-binding activity. One group identified the CIA/Asf1–histone-H3–H4 complex as an activation factor of CAF-1, which is involved in the nucleosome formation in DNA replication. The other group identified CIA/Asf1 as a histone chaperone that interacts with the CCG1 (cell cycle gene 1) subunit of the general transcription factor TFIID. Further analysis revealed two subtypes of CIA/Asf1, Asf1a and Asf1b (CIA-I and CIA-II, respectively). The CIA/Asf1 molecule is composed of two parts, the N-terminal structured region with 155 amino acid residues and a C-terminal acidic tail. The amino acid sequence of the N-terminal structured region is highly conserved among all eukaryotes; amino acid sequences of human and yeast CIA/Asf1 show approximately 60% sequence identity. Genetic, biological, and biochemical studies have revealed that CIA/Asf1 is involved in transcription, replication, and DNA repair. In these nuclear processes, CIA/Asf1 seems to play a role(s) in histone transfer and nucleosome structural change. Furthermore, CIA/Asf1, as described above, forms a complex with CAF-1 and HIRA, which have been considered to be involved in nucleosome formation in replication-dependent and -independent manners, respectively. The tertiary structure of CIA/Asf1 revealed that it has a β-sandwich structure like immunoglobulin (Fig. 1a). CIA/Asf1 forms a complex with the histone H3–H4 dimer as elucidated by crystal structure analyses (Fig. 1b). Biochemical analysis showed that CIA/Asf1 has an activity that disrupts the histone $(H3–H4)_2$ tetramer into two dimers through the formation of a CIA/Asf1–histone-H3–H4 complexes.



CIA/Asf1, Fig. 1 Crystal structures of CIA/Asf1 and its histone complex. (**a**) Crystal structure of CIA (PDB code: 1ROC). (**b**) Crystal structure of the CIA–histone-H3–H4 complex (PDB code: 2IO5). CIA, H3, and H4 are shown in *red*, *blue*, and *green*, respectively

Cross-References

▶ Histone Post-translational Modification to Nucleosome Structural Change

References

Eitoku M, Sato L, Senda T, Horikoshi M (2008) Histone chaperones: 30 years from isolation to elucidation of the mechanism of nucleosome assembly and disassembly. Cell Mol Life Sci 65:414–444

Circadian

▶ Circadian Rhythm

Circadian Rhythm

Jinzhi Lei
Zhou Pei-Yuan Center for Applied Mathematics, Tsinghua University of Beijing, Beijing, China

Synonyms

Biological clock; Circadian; Human clock

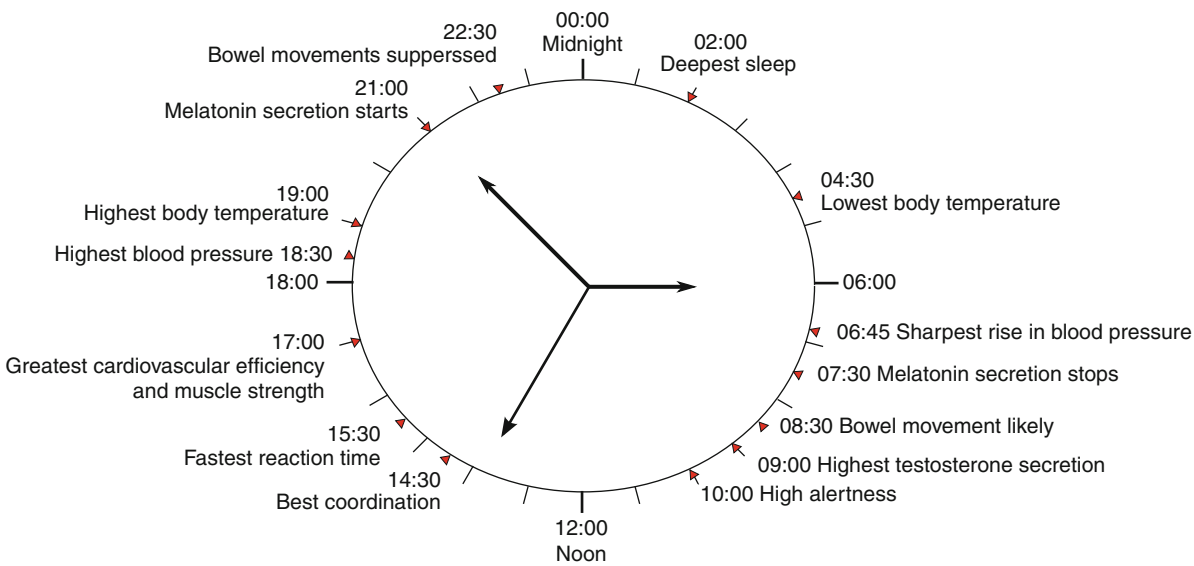
Definition

A circadian rhythm is the daily cycle of biological activity based on a 24-h period and influenced by regular variations in the environment, such as the alternation of night and day. Circadian rhythms include sleeping and waking in animals, flower closing and opening in angiosperms, and tissue growth and differentiation in fungi.

Figure 1 shows the circadian patterns of typical human who rises early in the morning, eats lunch around noon, and sleeps at night.

In mammals, the focus point of circadian rhythms is a master clock, located in the suprachiasmatic nuclei (SCN) of the anterior hypothalamus, which orchestrates the circadian program. Circadian timing in mammals is organized in a hierarchy of multiple circadian oscillators. The oscillatory machinery of the master clock is contained within single neurons ("clock cells"), and the SCN is composed of numerous clock cells. The SCN receives light information by a direct retinohypothalamic tract (RHT) to entrain the clock to the 24-h day, and in turn coordinates the timing of slave oscillators in other brain areas and in peripheral organs (Reppert and Weaver 2002).

Differences in clock protein levels and/or kinetics are considered as the main molecular basis of circadian timing in slave oscillators. The intracellular clock mechanism in mammals involves interacting positive and negative transcriptional feedback loops that drive recurrent rhythms in the RNA and protein levels of key clock components (Reppert and Weaver 2002). A detailed predictive model of the mammalian circadian clock was developed in Forger and Peskin (2003).



Circadian Rhythm, Fig. 1 Diagram of the circadian patterns of a typical human (Smolensky and Lamberg 2000)

References

Forger DB, Peskin CS (2003) A detailed predictive model of the mammalian circadian clock. Proc Natl Acad Sci USA 100:14806–14811

Reppert SM, Weaver DR (2002) Coordination of circadian timing in mammals. Nature 418:935–941

Smolensky M, Lamberg L (2000) The body clock guide to better health: how to use your body's natural clock to fight illness and achieve maximum health. Henry Holt and Company, New York

Cis-Acting Sequences in the Core Promoter

▶ Core Promoter Elements

Cis-Elements

▶ Transcription Factors and DNA Elements in Eukaryote

Classification

Eyke Hüllermeier, Thomas Fober and Marco Mernberger
Philipps-Universität Marburg, Marburg, Germany

Definition

In statistics and machine learning, classification refers to the problem of learning a classifier in the form of a mapping C from an instance space $\mathcal{X}$ to a finite set of class labels $\mathcal{Y} = \{y_1, y_2, \ldots, y_k\}$. Thus, classification is closely related to regression (▶ Regression Analysis), where the output space is numerical instead of categorical. In order to induce a classifier, the learning algorithm has access to training data, which typically consists of a set of examples $\{(x_1, y_1), \ldots, (x_n, y_n)\} \subset \mathcal{X} \times \mathcal{Y}$, that is, instances together with a corresponding class label. Classification is a specific type of supervised learning (▶ Learning, Supervised), and the main goal is to induce a classifier that generalizes well beyond the training data. Roughly speaking, this means that C should be able to classify so far unseen instances $x \in \mathcal{X}$ correctly. Apart from predictive accuracy, other criteria may of course play a role. For example, interpretable classifiers such as rule-based models (▶ Rule-based Methods) are often preferred to "black box" models delivering predictions which are not comprehensible by a human user. Besides, the efficiency and scalability of the underlying learning algorithms is of major importance.

Characteristics

The predictive accuracy of a classifier C is typically measured in terms of its *expected loss* (or risk) $R(C)$, that is, the expected value of the loss $L(C(X),Y)$, where (X,Y) is a sample taken at random from $\mathcal{X} \times \mathcal{Y}$ according to a probability distribution $\mathbf{P}$. Moreover, $L : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$ is a (real-valued) loss function that penalizes class predictions $\hat{y} = C(x)$ deviating from the true class y of an instance x:

$$R(C) = \mathbf{E}(L(C(X), Y)) = \int_{\mathcal{X} \times \mathcal{Y}} L(C(x), y) d\mathbf{P}(x, y)$$

The simplest loss function is the 0/1 loss, defined as $L(\hat{y}, y) = 0$ if $\hat{y} = y$ and $L(\hat{y}, y) = 1$ otherwise, though other losses are used as well. In *cost-sensitive classification*, for example, a different loss value $c_{ij} = L(y_i, y_j)$ can be specified for each pair of classes $(y_i, y_j) \in \mathcal{Y}^2$.

Since the probability distribution $\mathbf{P}$ is in general not known, the expected loss of a candidate model can only be estimated. An obvious estimate is the so-called empirical risk of a classifier C, namely, the average loss on the training data:

$$R_{emp}(C) = \frac{1}{n} \sum_{i=1}^{n} L(C(x_i), y_i)$$

However, for a classifier C learned on the training data, the empirical risk will normally underestimate the true risk. Roughly speaking, this is because C has been optimally adapted to the training data, and thus may fit this data even better than the true risk minimizer $C^* = \arg\min_{C' \in \mathcal{H}} R(C')$. In other words, minimizing the empirical risk on the training data comes with the danger of "▶ overfitting" the data, which

means that the true risk of a classifier C is much higher than its empirical risk.

The danger of "over-fitting" typically increases with the flexibility (capacity) of the underlying ▶ hypothesis space $\mathcal{H}$, that is, the set of candidate models from which a classifier C is chosen. This flexibility depends on the classification method or, more precisely, on the underlying model representation. ▶ Linear models, for example, separate classes by fitting a hyperplane in the instance space. They are less flexible than artificial neural networks, ▶ decision trees, or nearest neighbor methods, which can represent highly nonlinear decision boundaries. The restriction to certain types of decision boundaries is called the *representation bias* of a classification method. It largely determines the *inductive bias* of the method, which, informally speaking, produces a preference for specific types of models. A bias of that kind is needed to enable a choice between a potentially infinite number of mappings $\mathcal{X} \rightarrow \mathcal{Y}$ which explain the training data equally well. Implementing a suitable inductive bias and finding a proper trade-off between the complexity of a model and its performance on the training data are key prerequisites for successful learning (Bishop 2006).

Binary Versus Multi-class Classification

The simplest type of classification problem is binary (dichotomous) classification, in which $\mathcal{Y}$ consists of only two classes, typically referred to as the positive and negative class, respectively. For such problems, a multitude of efficient and theoretically well-founded classification methods exists. In fact, the representation of models is often geared toward the binary case, and sometimes even restricted to this problem class. For example, methods such as ▶ logistic regression and support vector machines produce decision boundaries in the form of a separating hyperplane that can only divide the instance space into two parts.

Other methods, such as decision trees and the naïve Bayes classifier, are not restricted in this way. Instead, they are directly applicable to the case of multi-class (polytomous) classification, that is, problems involving more than two classes. Algorithms that are inherently binary can still be used in a multi-class setting, namely, through *class binarization* techniques. A popular example is the one-against-rest binarization, where one takes each class in turn and learns a binary classifier that discriminates this class (considered as the positive class) from all other classes (considered as negative). At prediction time, each binary classifier predicts whether the query input is positive or not. Tie breaking techniques (typically based on confidence estimates for the individual predictions) are used in the case of a conflict, which may arise when no class or more than one class is predicted.

Ordinal and Hierarchical Classification

In standard classification, the set of class labels $\mathcal{Y}$ does not have any internal structure. In many applications, however, the classes are not completely unordered. Two important special cases are a total order $(y_1 \prec y_2 \prec \ldots \prec y_k)$ and a hierarchy, giving rise, respectively, to *ordinal classification* (also called ordinal regression in statistics) and *hierarchical classification*. As examples consider, respectively, learning to predict the expression of a gene on the scale $\mathcal{Y} = \{\text{underexpressed, normal, overexpressed}\}$ and the functional class of a protein according to the hierarchical EC nomenclature for enzymes.

From a learning point of view, the structure of $\mathcal{Y}$ is additional information that a learner should try to exploit, and this is what existing methods for ordinal and hierarchical classification essentially seek to do. The basic assumption in this regard is that the structure of $\mathcal{Y}$ is reflected in the topology of the class distribution in the instance space $\mathcal{X}$. Moreover, corresponding algorithms normally seek to minimize loss functions other than the simple 0/1 loss, which is arguably inappropriate for a structured output space $\mathcal{Y}$. For example, despite being wrong, predicting the functional class α-amylase (EC 3.2.1.1) for a protein that actually belongs to the class of β-amylases (EC 3.2.1.2) is still better than predicting a phosphorylase (EC 2.4.1.1) – the former two classes are both glycosidases, and hence functionally related.

Feature Representation

Instances $x \in \mathcal{X}$ are normally described in terms of an attribute-value representation or "feature vector", which means that x is a vector $(x_1, \ldots, x_m)$ and $\mathcal{X}$ the Cartesian product $\mathcal{X}_1 \times \ldots \times \mathcal{X}_m$, with $\mathcal{X}_i$ the domain of the i-th attribute $(i = 1, \ldots, m)$. In fact, many methods for classification as well as widely used software systems, such as the WEKA machine learning toolbox (Hall et al. 2009), expect exactly this type of representation.

Prior to actually training a classifier, this representation is normally subjected to a number of preprocessing steps, including normalization, dimensionality reduction (like ▶ Principal Component Analysis (PCA)), and ▶ feature selection or weighing. The selection of a small to moderate number of attributes (features) is important for many learning algorithms, the performance of which may deteriorate in the presence of irrelevant, noisy, or simply too many features. This aspect is especially critical in applications like gene expression analysis, where the number of features (the genes) may largely exceed the number of instances (e.g., different types of tissue).

Classification of Structured Data

Especially in the life sciences, however, a feature representation of the above kind is often not natural, and representing an object in terms of a fixed number of predefined features may come along with a considerable loss of information. Instead, other representations, such as sequences and graphs, appear to be more appealing. A small molecule, for example, is naturally represented in terms of a graph, with each atom corresponding to a node, and a bond between a pair of atoms modeled as an edge; likewise, biological networks are commonly represented as graphs or hypergraphs. Another example is modeling molecular structures on the sequence level, which comes down to representing them in terms of a sequence over a finite number of symbols, where different sequences can differ in length.

Dedicated classification algorithms for handling this type of data have been developed in the recent years. Especially interesting in this regard is kernel-based machine learning (▶ Learning, Kernel-based), including support vector machines, which allows for generalizing ("kernelizing") conventional methods by means of a mathematical construct called *kernel function* (Shawe-Taylor and Christianini 2004). A kernel function K is a mapping $\mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ satisfying special properties, and $K(x, x')$ can often be interpreted as a degree of similarity between the instances x and x'. A kernel of that kind can also be defined for structured data objects like graphs and sequences, thereby making kernel-based algorithms amenable to this type of data.

Ensemble Learning

In recent years, several methods for improving predictive performance have been developed that go beyond the learning of a single classifier. Instead, such methods can be used on a "meta level," more or less independently of the underlying classifier.

An important example of this type of approach is ▶ *ensemble learning*, where a potentially large number of classifiers is trained instead of a single one (Rokach 2010). At prediction time, each member of the ensemble is queried, and the predictions thus obtained are combined in one way or the other, for example, by means of a simple voting scheme or so-called stacking, where the optimal combination itself is formalized as a classification problem. In order to generate a (diverse) ensemble of classifiers, different methods can be used, including resampling methods such as ▶ bagging and boosting.

Classification in Systems Biology

Systems biology offers a multitude of applications for classification methods. For example, classification has already been used for the construction and analysis of metabolic networks, signaling and protein interaction networks, gene regulatory networks, the analysis of microarray data, and the prediction of the influence of toxins on biochemical pathways (Muggleton 2005; Larranaga et al. 2006; Fogel et al. 2007). The use of powerful predictors for protein–protein or protein–ligand interactions can help to reduce and organize the "wet" laboratory work necessary to verify such interactions in order to construct interaction networks. Moreover, classification tasks arise when extracting the information contained in huge biological networks (Rapaport et al. 2007).

Cross-References

- ▶ Bagging
- ▶ Data Mining–based Transcriptional Regulatory Network Construction
- ▶ Decision Tree
- ▶ Ensemble
- ▶ Feature Selection
- ▶ Learning, Kernel-Based
- ▶ Learning, Supervised
- ▶ Linear Model
- ▶ Logistic Regression
- ▶ Overfitting
- ▶ Principal Component Analysis (PCA)
- ▶ Regression Analysis
- ▶ Rule-based Methods

References

Bishop CM (2006) Pattern recognition and machine learning. Springer, Berlin
Fogel G, Fogel GB, Corne DW, Pan Y (2007) Computational intelligence in bioinformatics. Wiley-IEEE Press, New York
Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH (2009) The WEKA data mining software: an update. SIGKDD Explorations 11(1):10–18
Larranaga P, Calvo B, Santana R, Bielza C, Galdiano J, Inza I, Lozano JA, Armananzas R, Santafé G, Perez A, Robles V (2006) Machine learning in bioinformatics. Brief Bioinform 7(1):86–112
Muggleton SH (2005) Machine learning for systems biology. In: Kramer S, Pfahringer B (eds) Inductive logic programming. Springer, Berlin, pp 416–423
Rapaport F, Zinovyev A, Dutreix M, Barillot E, Vert JP (2007) Classification of microarray data using gene networks. BMC Bioinformatics 8(1):35–50
Rokach L (2010) Ensemble-based classifiers. Artif Intell Rev 33:1–39
Shawe-Taylor J, Christianini N (2004) Kernel methods for pattern anaylsis. Cambridge University Press, Cambridge

Classification Assessment and Optimization

▶ Receiver Operating Characteristic (ROC) Curve

Classification of Cancer Genesis

Carlos Sonnenschein and Ana M. Soto
Department of Anatomy and Cellular Biology, Tufts University School of Medicine, Boston, MA, USA

Definition

Cancers can be grouped into two main types regarding their genesis in the lifetime of humans. They are:
(a) Inborn errors of development
(b) *Sporadic* cancers

The *sporadic* cancers have been estimated to represent about 95% of the clinical cancers. As of recently, the *inborn errors of development* were considered as representing less than 5% of all cancers. These percentages are subject to reinterpretation as new evidence accumulates (Table 1).

Classification of Cancer Genesis, Table 1 Types of cancers

(A) Inborn errors of development
(A1) Inherited error of development ($\leq$5% of all clinical cancers)
(A2) Induced errors of development (?% of all clinical cancers)
(B) Sporadic cancers
~90% (?) of all clinical cancers

Characteristics

Inborn Errors of Development

Among these, two varieties have been distinguished, namely, (A1) *inherited* inborn errors of development and (A2) *induced* inborn errors of development (Table 1). The *inherited* inborn errors of development are the result of a process initiated by germ line mutation(s) in the genome of one or both gametes (sperm and/or ovum). If a mutated conceptus develops from this mutated egg, necessarily all of its cells will carry such genomic mutation(s); however, tumors usually appear in a single organ. This type of tumor (or malformation) represents the outcome of abnormal organogenesis in one or more tissue(s) in said offspring. Examples of this variety of inborn errors of development include retinoblastoma, Gorlin syndrome, xeroderma pigmentosa, and BRCA-1 breast neoplasias. For a comprehensive listing and description of these tumors, see Pizzo and Poplack (2011).

Tumors belonging to the *(A2) induced* inborn error of development variety appear either perinatally, during the first two decades after birth, or even later. Recent experimental evidence collected in laboratories worldwide using rodents verified that when embryos, fetuses, and/or neonates become exposed to chemical (even minute concentrations of environmental endocrine disruptors) and/or physical (radiation) carcinogens, a number of premalignant and malignant neoplastic lesions appeared in young and adult populations. There is evidence of a similar process in humans. For instance, in utero exposure of human fetuses to diethylstilbestrol (DES) during pregnancy was responsible for the appearance of rare vaginal adenocarcinomas around puberty and for an increased incidence of breast tumors in those women at the age of prevalence of this cancer (Hatch et al. 1998; Palmer et al. 2006). Given the greater incidence of breast cancers among increasingly younger women and the experimental data already accumulated,

it has been inferred that a good number of what so far have been considered *sporadic* cancers might, instead, belong to this *induced* inborn error of development group (Soto and Sonnenschein 2010). However, data have yet to be collected before accurate estimates of the percentage of what have been once considered to be *sporadic* cancers become, in fact, examples of *induced* inborn errors of development.

Sporadic Tumors

These represent the bulk of tumors seen by clinicians and diagnosed by pathologists that appear in adults which have no obvious link to germ line mutations. Their exact pathogenesis is currently the subject of intense controversy between the somatic mutation theory (SMT) (Weinberg 2006) and the tissue organization field theory of carcinogenesis (TOFT) (Soto and Sonnenschein 2011) (▶ Cancer Theories).

The SMT states that tumor formation is due to the effect of one or the accumulation of more mutations over a lifetime in a single founder cell in a host who was exposed to one or many carcinogens. Alternatively, the TOFT posits that carcinogens target *tissues* as units in the process of carcinogenesis. According to the TOFT, the disrupted interaction between tissues and among cells would fully explain the formation of a tumor. While the SMT explicitly proposes that all tumors are originally monoclonal, the TOFT postulates that the tumor mass is made up of multiple heterogeneous populations of cells that, from the beginning of the carcinogenic process, have the capacity to either evolve into a full-blown tumor or to regress and become normalized depending of the degree of tissue disruption the target tissue undergoes (Soto and Sonnenschein 2011).

For the purpose of classifying tumors as either *induced* inborn errors of development or of the *sporadic* type, a bone of contention has been whether carcinogens have to be mutagenic. According to the SMT, carcinogens are mutagens. The TOFT acknowledges, instead, that disruptors of normal tissue-tissue or cell-cell interactions can be carcinogens regardless of whether they are of a physical or chemical nature. Increasing amounts of experimental data suggest that carcinogens need not be mutagenic. This is exemplified by environmental endocrine disruptors that are not known to be mutagenic while retaining effective carcinogenic properties (Sonnenschein et al. 2011).

Regarding ▶ metastases, the TOFT proposes that the microenvironment in which the migrating cancer cells (epithelial plus stromal emboli) land plays either a "normalizing" (inhibiting) or a facilitating effect on the proliferation of those cells. In this context, the inhibiting or the facilitating effect of the microenvironments is susceptible to change as the host ages or if the host is exposed to additional carcinogens later in life. Additionally, cancer treatments (chemotherapy, radiation) may affect the microenvironment in which the cancer cells (epithelial plus stromal) that became detached from the primary tumor land. These contingencies will decide whether or not and when those "seeds" will become metastases. In contrast, the SMT claims that metastases will develop as a result of the acquisition of additional mutations or their activation in the cells of the primary tumor (which would induce them to migrate) and/or in those cells that already migrated from the primary and landed at a distance of it.

Cross-References

▶ Cancer Theories
▶ Metastasis

References

Pizzo PA, Poplack DG (2011) Principles and practice of pediatric oncology, 6th edn. Lippincott Williams & Wilkins, Philadelphia

Hatch EE, Palmer JR, Titus-Ernstoff L, Noller KL, Kaufman RH, Mittendorf R, Robboy SJ, Hyer M, Cowan CM, Ervin A et al (1998) J Am Med Assoc 280:630–634

Palmer JR, Wise LA, Hatch EE, Troisi R, Titus-Ernstoff L, Strohsnitter W, Kaufman R, Herbst AL, Noller KL, Hyer M et al (2006) Cancer Epidem Biomar 15:1509–1514

Sonnenschein C, Wadia PR, Rubin BS, Soto AM (2011) Journal of Developmental Origins of Health and Disease 2:9–16

Soto AM, Sonnenschein C (2010) Nat Rev Endocrinol 6:363–370

Soto AM, Sonnenschein C (2011) Bioessays 33:332–340

Weinberg RA (2006) The biology of cancer. Taylor & Francis, New York

Classification Performance Evaluation

▶ Receiver Operating Characteristic (ROC) Curve

Classification Tree

▶ Decision Tree

Clinical Aspects of the Toponome Imaging System (TIS)

Michael Khan[1] and Christine Waddington[2]
[1]Department of Biological Science, Biomedical Research Institute, University of Warwick, Coventry, UK
[2]MOAC/University of Warwick, Coventry, UK

Definition

The Toponome Imaging System (TIS) (▶ TIS Robot) has several key clinical applications including disease diagnosis and drug design. TIS can visualize the co-location of many proteins present on the surface and inside cells. Location patterns of the resulting protein clusters can be recognized that are disease or drug reaction specific. Additionally, by using known protein cluster patterns linked to a cell's phenotype (combinatorial molecular phenotype, CMP), and in combination with other systems biology tools, conclusions can be made about cell protein networks as a cell enters the diseased state. Drug therapies can be targeted so that they are disease specific, and the opportunities for early disease diagnosis are enhanced.

Characteristics

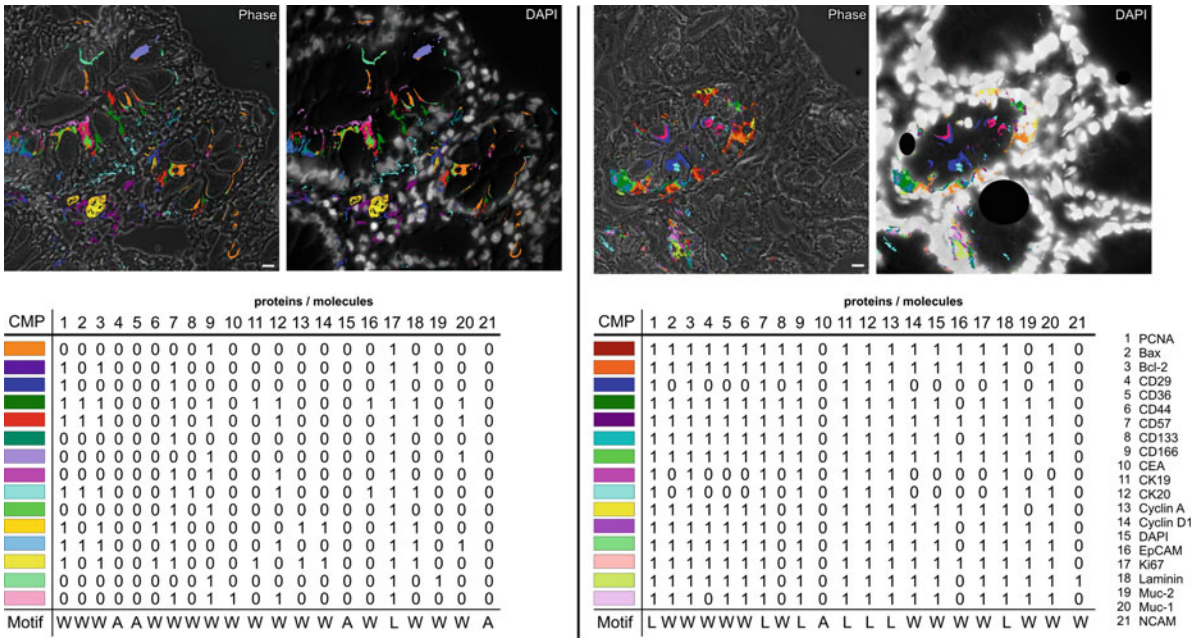
The Toponome Imaging System (TIS) (▶ TIS Robot) is an automated fluorescence technique with the ability to co-map at least 100 different proteins, or other TAG-recognizable biomolecules, on a single tissue section. To enable co-mapping, the system will generate a location image for every protein, or biomolecule, being studied.

Proteins are known to interact with each other in a cell, and the challenge of understanding how this cellular interaction occurs is a major stated goal in both medical research and systems biology. Most systems biology techniques, such as gene arrays or mass spectrometry, rely on tissue disruption before molecular phenotyping is undertaken. Clearly, as spatial and anatomical information is not preserved, the resultant lists of genes, transcripts, or proteins can only be partially translated into putative functional molecular networks by the use of informatics and this will be "error prone." There is certainly no possibility to compare protein anatomical location between chosen cells, or to directly link molecular phenotype with visible biology.

The current goal now is to microscopically examine protein expression in tissue sections or cell cultures. TIS is a new technique that combines traditional fluorescence microscopy with the ability to visualize expression of several hundred proteins simultaneously in the same tissue section or cell culture preparation. In disease diagnosis and subclassification, it is often critical to be able to directly relate molecular phenotype to anatomy. Importantly, to define a particular cell lineage or stem cell identity, observation of several proteins and other molecules may be needed and with TIS this can be done in situ, allowing study of such cells in their relevant "niche."

Finally, the ability to visualize protein co-localization on a large scale is potentially a very powerful surrogate for protein interaction. TIS looks directly at protein co-locations on the cell surface and inside the cell, and so can infer possible proteins working together. Although one cannot be certain that proteins co-localizing are physically interacting, proximity can be employed to greatly narrow down options defining functional protein networks.

To discover where proteins are co-locating in clusters, comparison is made between the individual protein fluorescent images. Patterns of clustering between specimens of similar types can then be determined, and visualized, as a mosaic of protein clusters directly linking to the cell's phenotype, these analyzed results being known as ▶ combinatorial molecular phenotypes (CMPs). Combining this with knowledge of associated protein networks, active proteins can be identified, and predictions made of the functional importance of these proteins for the different cell phenotypes. Samples from different subjects will have a natural CMP variation; however, comparable features between similar cell types are still observed due to the similarity of the underlying spatial arrangement of the protein networks.



Clinical Aspects of the Toponome Imaging System (TIS), Fig. 1 A TIS comparison of CMPs between cancerous colon cells (*on the left*) and normal colon cells in the same patient (*on the right*) (Bhattacharya et al. 2010)

Diseased and nondiseased tissues will show unique and different CMP patterns, and thus the patterns can act like "fingerprints" in further comparative analysis of other tissue. An example of these different CMP patterns can be seen in Fig. 1. The pictures show the CMP clusters displayed in color over a visual image (Phase) and over a fluorescent image (DAPI). The left-hand side shows colon cancer tissue, and the right-hand side normal colon tissue (both tissues are from the same patient). In this example, 21 proteins were used (given acronyms such as PCNA, Bax, etc.) and it can be seen for each color of protein cluster which proteins contribute (shown as a 1 in the table). Lead proteins (L) are present in all the main CMPs, while other proteins may be absent in the clusters (A). Wild card proteins (W) are present in some but not all protein clusters and may have key significance in the cell networks.

Using TIS analysis is a clinically powerful method because, as well as revealing which proteins are present and where, it indicates how they are co-located, giving further information about particular cells and subcellular regions. Importantly, being able to use a large number of biomarkers allows subtle differences between similar samples to be determined. For example, compared to normal skin, psoriasis and atopic dermatitis exhibit distinct protein clusters: Some of these clusters are specific to either psoriasis or atopic dermatitis, whereas others are common to both conditions (Schubert et al. 2006). If the clinician knows the unique "fingerprint" of each disease, only a small skin sample is needed to give a clear diagnosis. Disease fingerprinting will also aid the clinician in situations of uncertain or limited clinical symptoms, helping to prevent situations of misdiagnosis.

Another example of the clinical possibilities for using TIS is in cancer studies. In cancer, the ability to look at the tissue for the visible hallmarks of malignancy alongside molecular phenotype is critical, in particular, as it enables the examination of single cells, which may be cancer cells that are invading, dividing, or could be immune or stromal cells. Because TIS can detect individual cell protein patterns, a very small number of affected cells can give a positive cancer result; this enhances the chances of early detection, especially because cancer cells show very distinctive differences in their topology in comparison to normal, noncancerous cells. While much is known about the biochemical pathways in cancers, an understanding is needed in cell organization of functions, and this can be achieved using ▶ toponome analysis to create insight into the key protein networks driving the

disease. Schubert et al. (2009) analyzed prostate cancer tissue, determining a mosaic of distinct protein clusters that were specific to the observed cancerous cells. Using disease-specific CMP knowledge, two routes can now be taken: The first to target lead proteins in the development of the cancer with specifically designed drug therapies, the second to aid the early detection of cancer development. In instances where two diseases can have shared but distinct disease pathways, the subtle differences in lead proteins can aid understanding of the triggers for pathogenic events that result in cancer. In addition, with the CMP clusters not likely to occur by chance, a correct diagnosis can still be made even if the cells are at the early stages of cancer formation. As technological advances are made, these protein clusters can be looked for in other body samples, such as from the blood, enabling a less invasive method of early cancer detection.

Sometimes the knowledge of genomic loci involved in the pathogenesis of a disease is limited. Because the TIS technique can furnish a toponomic picture of what is occurring in the cell, proteins can be detected assembling on the cell surface. This surface assembly is part of a complex communication network involved with many different cell functionalities and, by monitoring these protein networks, key proteins can be determined and disease-altered network dynamics detected. This is valuable in situations where diseases are rare, or there is limited knowledge of cell protein pathways in a disease. Given sufficient typical protein clusters, by using comparisons with protein clusters from known networks, it is possible to predict disease behavior and effective therapies.

Using a systems biology approach, by combining the TIS results with proteome analysis and transcriptome analysis, an interpretation of multiple levels of gene expression can take place leading to further analysis, for example, comparing the effect of drug interactions on different patients. Drugs are generally designed to influence protein interactions and so the effects of the drug can be easily detected using TIS: Changes in the protein networks of synapses due to the analgesic drug dipyrone have already been observed (Linke et al. 2009).

While TIS is a powerful and widely applicable stand-alone tool, it also works well in complement with other tools as part of a systems biology pipeline. The experimenter specifies which proteins or other molecules to study using TIS, and new target proteins can be found, for example, by also using mass spectrometry or gene analysis. Despite having pre-decided on the proteins to investigate, TIS remains a tool for discovery science because previously unsuspected protein interactions may still be identified. Combining TIS with new knowledge from genomics, transcriptomics, and proteomics allows various differentially expressed proteins to be studied in the context of intact anatomy. The promise that TIS can detect clear differences between normal and unhealthy tissue, even if there are few cells and the differences are subtle, opens up the prospect of easy, early-stage cancer detection using blood or, for colon cancer, stool samples. There is further potential that cancer stem cells will become detectable, and drug therapies will be developed to specifically target stem cells restricting cancer growth. With greater understanding of the cell's network pathways, drug design can be made more disease specific and enable more precise targeting, whether this is in the form of an analgesic or perhaps a unique drug for a rare disease.

Cross-References

- ▶ TIS Robot
- ▶ Toponome Analysis

References

Bhattacharya S, Mathew G, Ruban E, Epstein D, Krusche A, Hillert R, Schubert W, Khan M (2010) Toponome Imaging System: In situ protein network mapping in normal and cancerous colon from the same patient reveals more than five-thousand cancer specific protein clusters and their subcellular annotation by using a three symbol code. J Proteome Res 9:6112–6125

Linke B, Pierre S, Coste O, Angioni C, Becker W, Maier T, Steinhilber D, Wittpoth C, Geisslinger G, Scholich K (2009) Toponomics analysis of drug-induced changes in arachidonic acid-dependent signalling pathways during spinal nociceptive processing. J Proteome Res 8:4851–4859

Schubert W, Bonnekoh B, Pommer A, Philipsen L, Böckelmann R, Malykh Y, Gollnick H, Friedenberger M, Bode M, Dress A (2006) Analyzing proteome topology and function by automated multidimensional fluorescence microscopy. Nat Biotechnol 24:1270–1278

Schubert W, Gieseler A, Krusche A, Hillert R (2009) Toponome mapping in prostate cancer: Detection of 2000 cell surface protein clusters in a single tissue section and cell type specific annotation by using a three symbol code. J Proteome Res 8:2696–2707

Clinical Decision Support Systems

▶ Biomedical Decision Support Systems

Clinical Systems Pathology

▶ Systems Pathology

Cliquishness

▶ Clustering Coefficient

Clonal Division

▶ Modeling, Cell Division and Proliferation

Closed Chromatin

▶ Heterochromatin

Closed World Assumption

C. Maria Keet
KRDB Research Centre, Free University of Bozen-Bolzano, Bolzano, Italy

Definition

The Closed World Assumption (CWA) is the assumption that that what is not known to be true, is false, so that absence of information is interpreted as *negative* information. It assumes that *complete* information about a given state of affairs is provided, which is useful for constraining information and validating data in an application such as a relational database. This is contrasted with the Open World Assumption.

Closed World Assumption, Table 1 Sample instances about alumni and their degrees obtained

Alumnus	Degree obtained
Delani	PhD in Molecular Biology
Anna	PhD in Ecology
Peter	MSc in Informatics
Dalila	PhD in Genetics

Example. Take the sample data in Table 1 and a query "Which alumni do not have a PhD?". Then under the CWA, it answers with "Peter" because under the CWA the system assumes this information is all there is in the world.

Closure, Causal

Matteo Mossio
IAS-Research Philosophy of Biology Group, Department of Logic and Philosophy of Science, University of the Basque Country (UPV/EHU), Donostia – San Sebastian, Spain

Definition

In biological systems, closure refers to a holistic feature such that their constitutive processes, operations, and transformations (1) depend on each other for their production and maintenance and (2) collectively contribute to determine the conditions at which the whole ▶ organization can exist.

According to several theoretical biologists, the concept of closure captures one of the central features of biological organization since it constitutes, as well as evolution by natural selection, an emergent and distinctively biological causal regime. In spite of an increasing agreement on its relevance to understand biological systems, no agreement on a unique definition has been reached so far.

Characteristics

The concept of closure plays a relevant role in biological explanation since it is taken as a naturalized grounding for many distinctive biological dimensions,

as purposefulness, normativity, and functionality (Chandler and Van De Vijver 2000).

The contemporary application of closure to the biological domain comes from a philosophical and theoretical tradition tracing back at least to Kant who claimed, in the *Critique of Judgment*, that biological systems should be understood as natural purposes (*Naturzwecke*), i.e., systems in which the parts are reciprocally causes and effects of the others, such that the whole can be conceived as organized by itself, self-organized. The essence of living system is a form of internal and circular causality between the whole and the parts, distinct from both efficient causality of the physical world and the final causality of artifacts (Kant 1985).

One of the most influential contemporary characterizations of closure in the biological domain has been provided by Francisco Varela (1979). In his account, he builds on an algebraic notion, according to which "a domain K has closure if all operations defined in it remain within the same domain. The operation of a system has therefore closure, if the results of its action remain within the system (Bourgine and Varela 1992, p. xii)."

Applied to biological systems, closure is realized as what Varela labels *operational* (or *organizational*) *closure*, which designates an organization of processes such that "(1) the processes are related as a network, so that they recursively depend on each other in the generation and realization of the processes themselves, and (2) they constitute the system as a unity recognizable in the space (domain) in which the processes exist" (Varela 1979, p. 55).

It should be noted that Varela himself has proposed, over time, slightly different definitions of operational closure. In addition, more recent contributions have introduced a theoretical distinction between organizational and operational closure. Whereas "organizational" closure indicates the abstract network of relations that define the system as a unity, "operational" closure refers to the recurrent dynamics and processes of such a system (Thompson 2007).

In Varela's view, operational closure is closely related to ▶ autonomy, the central feature of living organization. More precisely, he enunciates the "Closure Thesis," according to which "every autonomous system is operationally closed" (Varela 1979, p. 58). In principle, the class of autonomous systems realizing operational closure is larger than the class of biological systems. As a consequence, operational closure is taken as a necessary but not sufficient condition to define biological organization. Biological systems, in fact, constitute a subclass of autonomous systems, which realize a specific form of operational closure, which Varela labels, with Humberto Maturana, autopoiesis (▶ Systems, Autopoietic) (Varela 1979). The specificity of operational closure as autopoiesis is that, unlike other possible forms, it describes the system at the chemical and molecular level, and supposes relations of material production among its constituents.

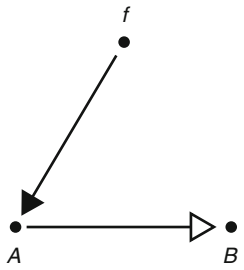
A crucial distinction is usually made between organizational/operational and *material* closure, where the latter indicates the absence or incapacity to interact. While being organizationally closed, biological systems are structurally coupled with the environment, with which they exchange matter, energy, and information. The concept of biological closure implies then a distinction between two causal levels, an open and a closed one – an issue which have been more explicitly addressed by the account proposed by Robert Rosen (Rosen 1991).

Rosen's account is based on a rehabilitation and reinterpretation of the Aristotelian categories of causality and, in particular, on the distinction between efficient and material cause. Let us consider an abstract mapping f between the sets A and B, such that $f: A => > B$. Represented in a relational diagram, we have (Fig. 1):

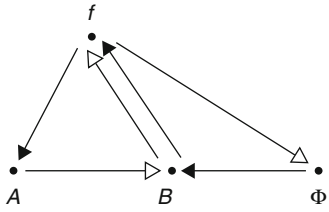
When applied to model natural systems, Rosen claims that the hollow-headed arrow represents material causation, a flow from A to B, whereas the solid-headed arrow represents efficient causation, a ▶ constraint exerted by f on this flow.

Rosen's central thesis is that "a material system is an organism [a living system] if, and only if, it is closed to efficient causation" (Rosen 1991, p. 244), whereas a natural system is closed to efficient causation if and only if its relational diagram has a closed path that contains all the solid-headed arrows. It is worth noting that, unlike the varelian tradition, Rosen takes closure as the *definition* of biological organization.

According to Rosen, the central feature of a biological system consists in the fact that all components having the status of efficient causes are materially produced by and within the system itself. At the most general level, closure is realized in biological systems among three classes of efficient causes corresponding to three broad classes of biological functions (▶ Function, Distributed) that Rosen denotes as *metabolism*



Closure, Causal, Fig. 1 Relational diagram, descriprion see text



Closure, Causal, Fig. 2 Relational diagram, descriprion see text

(f: A=>> B), *repair* (Φ: B=>> f), and *replication* (B: f=>>Φ) (Fig. 2).

By providing a clear-cut theoretical and formal distinction between material and efficient causation, Rosen's characterization explicitly spells out that biological organization consists of two coexisting causal regimes: closure to efficient causation, which grounds its unity and distinctiveness, and openness to material causation, which allows material, energetic, and informational interactions with the environment.

More recently, the scientific work on biological closure has been developed in various directions (Chandler and Van De Vijver 2000). In particular, a thriving research line has specifically focused on the critical nature of systems realizing closure, which must maintain a continuous flow of energy and matter with the environment in conditions far from thermodynamic equilibrium. To capture this dimension of closure, Stuart Kauffman has proposed the notion of Work-Constraint cycle (Kauffman 2000).

The Work-Constraint cycle represents an interpretation of organizational closure that links the idea of "work" to that of "► constraint," the former being defined, as "constrained release of energy into relatively few degrees of freedom." A system realizes a Work-Constraint cycle if it is able to use its work to regenerate at least some of the constraints that make work possible. The cycle is a thermodynamic irreversible process, which dissipates energy and requires a coupling between exergonic (spontaneous, which release energy) and endergonic (non spontaneous, which require energy) reactions, such that exergonic processes are constrained in a specific way to produce a work, which can be used to generate endergonic processes, which in turn generate those constraints canalizing exergonic processes. In Kauffman's terms: "Work begets constraints beget work" (Kauffman 2000).

A complementary account of closure has been proposed by Howard Pattee, who focused on its informational dimension (Pattee 1982). In his view, biological organization consists of the integration of two intertwined dimensions, which cannot be understood separately. On the one side, the organization realizes a dynamic and autopoietic network of mechanisms and processes, which defines itself as a topological unit, structurally coupled with the environment. On the other side, it is shaped by the material unfolding of a set of symbolic instructions, stored and transmitted as genetic ► information.

According to Pattee, the dynamic/mechanistic and informational dimensions realize a distinct form of closure between them, which he labels *semantic closure*. By this notion, he refers to the fact that while symbolic information, to be such, must be interpreted by the dynamics and mechanisms that it constrains, the mechanisms in charge of the interpretation and the "material translation" require that very information for their own production. Semantic closure, as an interweaving between dynamics and information, constitutes then an additional dimension of organizational closure of biological systems, complementary to the operational/efficient one.

Cross-References

- ► Autonomy
- ► Constraint
- ► Emergence
- ► Explanation in Biology
- ► Explanation, Functional
- ► Function, Distributed
- ► Holism
- ► Information, Biological
- ► Organization

References

Bourgine P, Varela FJ (1992) Toward a practice of autonomous systems. MIT Press/Bradford Books, Cambridge, MA

Chandler JLR, Van de Vijver G (eds) (2000) Closure: emergent organizations and their dynamics. Ann NY Acad Sci 901

Kant I (1987/1790) Critique of judgment. Hackett, Indianapolis

Kauffman S (2000) Investigations. Oxford University Press, Oxford

Pattee HH (1982) Cell psychology: an evolutionary approach to the symbol–matter problem. Cogn Brain Theory 5(4):325–341

Rosen R (1991) Life itself: a comprehensive inquiry into the nature, origin, and fabrication of life. Columbia University Press, New York

Thompson E (2007) Mind in life: biology, phenomenology, and the sciences of mind. Harvard University Press, Cambridge, MA

Varela FJ (1979) Principles of biological autonomy. North Holland, New York

Cluster Analysis

Tim W. Nattkemper
Biodata Mining Group, Faculty of Technology, Bielefeld University, Bielefeld, Germany

Synonyms

Vector quantization

Definition

A large and high-dimensional TIS (TIS robot) (► Clinical Aspects of the Toponome Imaging System (TIS); ► TIS Robot) data set is reduced to a small set of protein co-location prototypes without the application of thresholds. Different indices are proposed to evaluate the ► clustering result or to estimate an appropriate number of clusters in a data set. To allow visual diagnostics and to apply biological expert knowledge, a special visualization icon is proposed. This way, important molecular co-location patterns can be identified and localized in the image, which helps in different fields of systems biology (► Systems Biology Pathway Exchange (SBPAX)) like ► pathway analysis or protein network analysis or other ► toponomics applications.

Characteristics

The main idea behind TIS (TIS robot) (► Clinical Aspects of the Toponome Imaging System (TIS); ► TIS Robot) data clustering is to find a way to interpret ► toponome images without applying thresholds. Thus, a toponome image recorded with an antibody library of N tags is considered a N-dimensional data of fluorescence grey value patterns $\mathbf{g}^{(x,y)} = (g_1, \ldots, g_N)^{(x,y)}$ with (x, y) as pixel coordinates and $g_j^{(x,y)}$ denoting the grey value for the j-th tag at pixel (x, y). ► Clustering provides a valuable method for partitioning this data into groups of similar grey value patterns. This partitioning has two aims: First, clustering achieves a data reduction, so that data sets can be analyzed and compared on a meta-level. Second, one wants to address the fact that similar co-location patterns may also belong to the same functional group or to the same hierarchically organized network. If cluster analysis is to be applied to TIS data, the following aspects have to be considered.

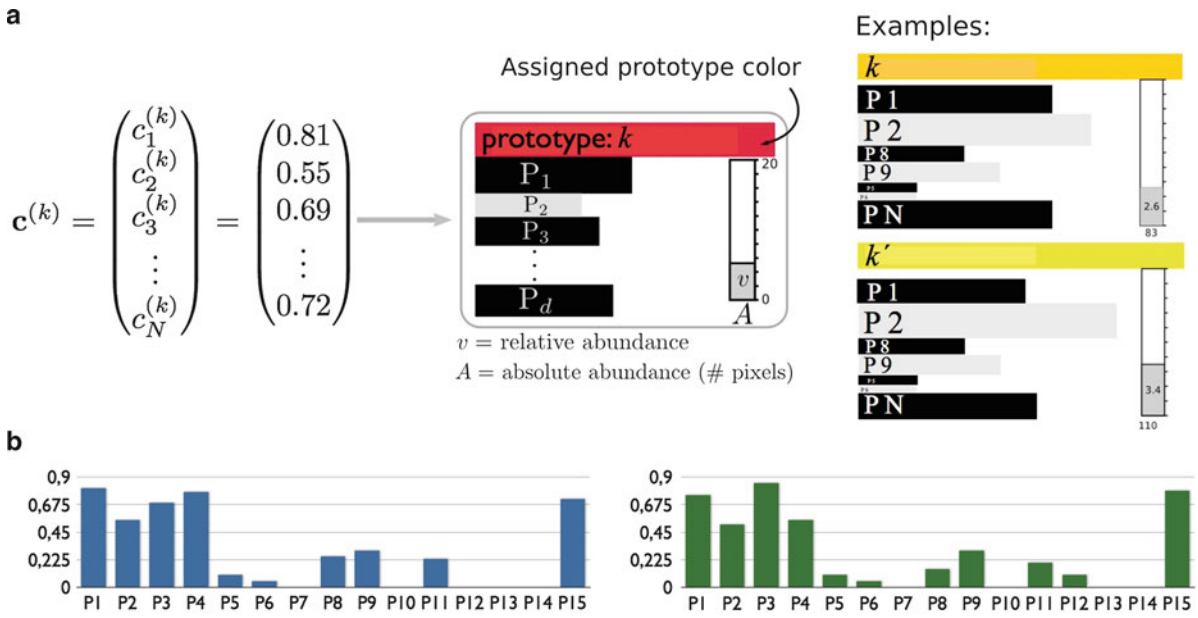
Like in any other clustering application, it is not possible to define a criterion for an optimum cluster result, since clustering tries to achieve two opposing goals: connectedness and compactness (Handl et al. 2005). Transferred to the context of toponomics data ► clustering this can be seen as the conflict between the two ideas:(1) to generate large clusters that represent one frequent co-fluorescence pattern with some variation, and (2) to form small compact clusters that contain only those proteins which are utmost similar. Cluster analysis algorithms can be divided into the following two groups.

Hierarchical cluster analysis methods organize the input data (i.e., the fluorescence grey value patterns of one or two TIS images) into a tree structure exposing the relationships from the most similar to the most different fluorescence grey value patterns. In contrast, partitioning or vector quantization cluster algorithms successively assign each pattern $\mathbf{g}^{(x,y)}$ to one distinct group, i.e., cluster. In the clustering result, each of the clusters is characterized by a typical representative or the cluster center, which is usually referred to as the prototype or the codebook vector. The most popular partitioning cluster method is the ► K-means approach (Lloyd 1982). The objective is to find a set of K prototypes so that the distances between each

data item and its closest prototype are minimized. Although the obtained clustering result may represent only a locally optimal solution rather than the global optimum, the strategy is often used due to its algorithmic simplicity and efficiency, and has been recently voted in the list of top ten algorithms in data mining (Wu et al. 2008). For large data sets such as toponome data, the consideration of each data item for prototype adaptation can be computationally expensive, which motivates an online K-means version applying the *winner-takes-all* (*WTA*) rule. It is common to repeatedly run the algorithm with different initializations (Hastie et al. 2001) and to take that solution as the clustering result, which shows the smallest intra-class variance. Neural Gas clustering claims to be an enhancement as it takes into account a "neighborhood ranking" of all proteins that are assigned to a cluster – an advantage bought by an increase in computational running time. In TIS cluster analysis, one should usually favor partitioning cluster algorithms, since the tree plot for an entire image should be too large for a visual analysis. Thus, the following part of this entry focuses on partitioning cluster algorithms.

A number of quality measures have been proposed to evaluate the outcomes of cluster algorithms. Because of its opposing goals, no definite criterion can be formulated that describes an optimal clustering of a dataset. This pertains not only to the applied cluster algorithm but also to the "true" number of clusters of a dataset. Proposed measures that base solely on the clustering itself and the underlying dataset range from early approaches (Calinski and Harabasz 1974; Davies and Bouldin 1979) up to novel instruments (Yeung et al. 2001; Maulik and Bandyopadhyay 2002). Another kind of assistance in choosing a cluster algorithm was recently proposed in Yeung et al. (2001). They delineated an instrument called Figure of Merit (FOM) to evaluate cluster solutions. The idea of their method is to integrate a kind of bootstrapping approach and thereby estimate the predictive power of a cluster algorithm.

One important aspect in TIS cluster analysis is the evaluation of the outcome, i.e., the cluster result, which is, in case of partitioning methods, represented by the set of K prototypes $\{\mathbf{c}^{(k)}\}k = 1,\ldots, K$. Usually, the grey values of the images are transformed to a certain range (e.g., [0; 1]) using a linear scaling or some nonlinear scaling function such as tan h ($g_j^{(x,y)}$), so for all prototype components it is $c_j^{(k)} \in [0; 1]$. An essential first step in the evaluation is the visual inspection of the prototypes, so a graphical display is needed to support visual ▶ data mining. A visual inspection of the prototypes allows the identification of interesting protein co-location patterns in the data and a comparison with ▶ combinatorial molecular phenotypes (CMPs), without the need of analyzing single images which eases the knowledge discovery process. If interesting prototypes are found, the associated data items can be analyzed in a subsequent step following the Shneiderman mantra of "Overview first, zoom in, and filter, details on demand." However, suitable prototype visualization is not as straightforward as it seems. One way to display multivariate data are glyph or icon displays. According to Colin Ware "A glyph is a graphical object designed to convey multiple data values" (Ware 2004, p.145). Each data feature is mapped to a different graphical attribute of the glyph such as size, shape, or color. We apply a glyph-based visualization approach, which has been developed to suit the needs of non-binarized signal co-location analysis (Loyek et al. 2011a, b; Kölling et al. 2012). This glyph approach combines visualization aspects known from bar charts and star glyphs and is to some extent inspired by the sequence logo display, which represents patterns in nucleotide or amino acid sequences. In a sequence logo, for each position of a set of aligned sequences, e.g., nucleotide sequences, the four nucleotides are arranged on top of each other sorted according to their frequency at that position. The character height represents the frequency of the according nucleotide. Through this visualization, a rapid identification of prominent sequence patterns can be achieved as high frequent nucleotides can directly be "read" from the logo. To construct a glyph for one signal co-location $\mathbf{c}^k$ (k = 1,. . ., K), a horizontal box is drawn for each data feature (see Fig. 1a). The height, as well as the length, of each box is scaled according to the feature's value. To increase differentiation between neighboring boxes, they are alternatingly colored in black and light shaded grey. This follows Ware's suggestion for star glyphs or whisker plots to increase the number of dimensions by changing length and width of the bars as well as using different luminance levels. Furthermore, by employing



Cluster Analysis, Fig. 1 (**a**) Generation of a glyph for one cluster prototype $\mathbf{c}^{(k)}$: Each box is scaled in height and width according to the features' values. The prototype names (P1, . . ., PN) are written in the *boxes* for fast association of the proteins to the boxes. The relative and absolute abundance of the prototype, i.e., how many pixel are associated to that prototype, as well as the assigned prototype color, are given. For comparison, two classic bar plot displays for two 15-dimensional cluster prototypes are displayed. Since this glyph display is much more compact compared to a standard box plot (**b**), more prototypes can be displayed on the same screen space and a comparative analysis is easier

length as an attribute for data representation, a well-suited graphical parameter to encode quantitative data has been chosen, as has already been discussed for the bar plot. To allow for a fast identification of prominent proteins, the protein names are directly incorporated into the visualization. To this end, the associated protein name is written in each bar and scaled in height and length analogue to the bar itself. With this strategy, prominent protein co-localization can easily be identified by "reading" the glyph analogous to the reading of a sequence logo. Figure 1a displays the construction of the glyph display and shows two glyph examples on the right. In addition to the special box plot display of the feature vector components, the top bar of the glyph shows the color which is assigned to the glyph, and meta-information as the associated data set and the prototype number can be displayed here as well. How color will be assigned to each prototype will be explained below. On the right, the glyph displays information about the abundance of the feature combination in one selected TIS image (or in a set of TIS images of course). The grey box with the overlaid number shows the relative abundance, below the absolute abundance of pixels is displayed, which are assigned to this prototype according to the best match criterion. This meta-information is very valuable for prototype analysis. If this glyph visualization is compared to a classic bar graph (see Fig. 1b), it is evident that in the glyph display the association of proteins to individual graphical attributes is much easier. Furthermore, besides being able to rapidly identify the dominant proteins, an advantage of the glyph display is that only features with high values allocate space, whereas low value features are squeezed in contrast to Fig. 1a). Thereby, space is only allocated proportional to the importance of the protein and the total size of the glyph reflects the amount of information provided by the prototype. In some applications, this might not be a desirable feature so that bar graphs, or glyphs with constant bar width would better be suited.

Different strategies to assign color to the prototypes and the corresponding glyphs can be proposed. If one

wants to achieve preservation of topologies between the cluster positions in the *N*- dimensional co-location space and the color space, i.e., similar clusters get similar colors, dimensional reduction can be applied. Using for instance Sammon mapping or ▶ principal component analysis (PCA), the *N*-dimensional prototypes are mapped into a 3D color space where each prototype is associated with a three-dimensional color vector $\mathbf{h} = (h1, h2, h3)$.. A whole pseudo-color image can be rendered by finding for each feature vector $\mathbf{g}^{(x,y)}$ its best matching prototype Nearest Neighbor Method and its color. The pixel position x, y is plotted in the corresponding color. The most common color space is the RGB space, where color is produced by additively mixing red, green, and blue primaries. Another widely used space is the HSV space, where color is described by means of hue (H), saturation (S) and value (V).

Clustering Algorithms and Tools

In general, there is a large variety of clustering algorithms and tools; open source tools widely used in computational biology include Weka (Weka, Machine Learning Tool) and R (R, Data Analysis Tool) ▶ R, Programming Language.

Cross-References

- ▶ Clinical Aspects of the Toponome Imaging System (TIS)
- ▶ Clustering
- ▶ Clustering, k-Means
- ▶ Data Mining
- ▶ Information Theory and Toponomics
- ▶ Knowledge Acquisition
- ▶ Learning, Supervised
- ▶ Model Testing, Machine Learning
- ▶ Model Training, Machine Learning
- ▶ Principal Component Analysis (PCA)
- ▶ Subset Surprisology and Toponomics
- ▶ TIS Robot
- ▶ Topology and Toponomics
- ▶ Toponome Analysis
- ▶ Toponomics
- ▶ Unsupervised Learning Methods

References

Calinski T, Harabasz J (1974) A dendrite method for cluster analysis. Commun Stat Theory Methods 3:1–27. doi:10.1080/03610927408827101. http://dx.doi.org/10.1080/03610927408827101

Davies DL, Bouldin DW (1979) A cluster separation measure. IEEE Trans Pattern Anal Mach Intell PAMI-1:224–227

Handl J, Knowles J, Kell D (2005) Computational cluster validation in post-genomic data analysis. Bioinformatics 21:3201–3212

Hastie T, Tibshirani R, Friedman JH (2001) The elements of statistical learning. Springer, New York

Kölling J, Langenkämper D, Abouna S, Khan M, Nattkemper TW (2012) WHIDE—a web tool for visual data mining colocation patterns in multivariate bioimages. Bioinformatics 28(8):1143–1150. doi:10.1093/bioinformatics/bts104.

Lloyd S (1982) Least squares quantization in PCM. IEEE Trans Inf Theory 28:129–137

Loyek C, Bungkowski A, Nattkemper TW (2011a) Web2.0 paves new ways for collaborative and exploratory analysis of chemical compounds in spectrometry data. J Integr Bioinform 8(2):158. doi:10.2390/biecoll-jib-2011-158

Loyek C, Langenkämper D, Kölling J, Niehaus K, Nattkemper TW (2011b) A Web2.0 strategy for the collaborative analysis of complex bioimages. In: Advances in intelligent data analysis X: 10th international symposium, IDA 2011, Porto, October 29–31, 2011, Proceedings. Lecture notes in computer science (Applications, incl. Internet/Web, and HCI). Springer, Berlin/Heidelberg, pp 258–270

Maulik U, Bandyopadhyay S (2002) Performance evaluation of some clustering algorithms and validity indices. IEEE Trans Pattern Anal Mach Intell 24:1650–1654

Ware C (2004) Information visualization: perception for design. Morgan Kaufmann, San Francisco

Wu X, Kumar V, Ross GJ, Yang Q, Motoda H, Mclachlan G, Ng A, Liu B, Yu P, Zhou ZH, Steinbach M, Hand D, Steinberg D (2008) Top 10 algorithms in data mining. Knowl Inf Syst 14:1–37

Yeung K, Haynor D, Ruzzo W (2001) Validating clustering for gene expression data. Bioinformatics 17:309–318

Cluster of Differentiation

Rohan Dhiman
Center for Pulmonary and Infectious Disease Control, University of Texas Health Science Center at Tyler, Tyler, TX, USA

Definition

It stands for cluster of differentiation. It is a method conceptualized in 1st international

workshop and conference on human leukocyte differentiation antigens to provide nomenclature to all the cell surface molecules present on white blood cells. It includes various molecules some which either act as ligands or receptors (Zola et al. 2005; Bernard and Boumsell 1984; Fiebig et al. 1984).

Cross-References

▶ Fluorescent Markers

References

Bernard A, Boumsell L (1984) Human leukocyte differentiation antigens. Presse Méd 13(38):2311–2316

Fiebig H, Behn I, Gruhn R, Typlt H, Kupper H, Ambrosius H (1984) Characterization of a series of monoclonal antibodies against human T cells. Allerg Immunol (Leipz) 30(4):242–250

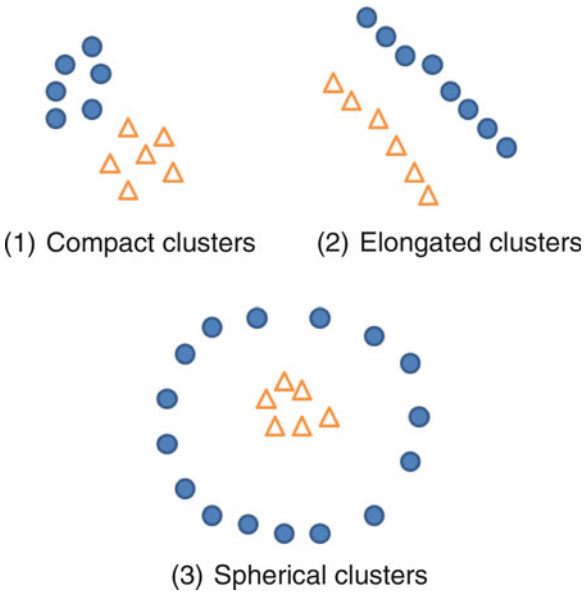
Zola H, Swart B, Nicholson I, Aasted B, Bensussan A, Boumsell L, Buckley C, Clark G, Drbal K, Engel P, Hart D, Horejsí V, Isacke C, Macardle P, Malavasi F, Mason D, Olive D, Saalmueller A, Schlossman SF, Schwartz-Albiez R, Simmons P, Tedder TF, Uguccioni M, Warren H (2005) CD molecules 2005: human cell differentiation molecules. Blood 106(9):3123–3126

Clustering

Kejia Xu
Institute of Systems Biology, Shanghai University, Shanghai, China

Definition

A loose definition of clustering could be "the process of organizing objects into groups whose members are similar in some way." Each cluster is then characterized by the common attributes of the entities it contains. Usually, the objects in a cluster are "more similar" to each other and "less similar" to the objects of the other clusters. From different clustering types, there are three clustering patterns as below.



References

Cornish (2007) Cluster analysis. Mathematics learning support chapter 3.1.

Clustering Coefficient

Guilhem Chalancon, Kai Kruse and M. Madan Babu
MRC Laboratory of Molecular Biology, University of Cambridge, Cambridge, UK

Synonyms

Cliquishness; Density of a subgraph; Transitivity; Watts-Strogatz local clustering coefficient

Definition

Many problems in network analysis converge to the question of the cohesion of a graph, aiming to determine

the extent to which the nodes of a network are closely connected with one another. Numerous strategies exist to measure the cohesion of a network and therefore several metrics can be used (Kolaczyk 2009). In this respect, the clustering coefficient of a graph is widely used in network analysis. One can distinguish between local measurements of the clustering of nodes in a graph and global measurements of the clustering coefficient of an entire graph.

Characteristics

Local Clustering Coefficient

In a graph G composed of a set of vertices (nodes) V and a set of edges (links) E, the local clustering coefficient, denoted C_v, measures the local density of edges in the neighborhood of a node $v \in V$. C_v corresponds to the proportion of links within the immediate neighborhood of v, among all possible links connecting v and its neighbors.

Let $n_e(v)$ denote the number of edges that connect the immediate neighbors of a node v. Let k_v denote the node degree of v, that is, its number of immediate neighbors. The maximal number of possible edges between the neighbors of v is therefore $k_v(k_v - 1)/2$ for an undirected graph. Therefore, in an undirected graph with no self-loops, C_v is given as:

$$C_v = \frac{2n_e(v)}{k_v(k_v - 1)}$$

With directed edges, the number of possible links in the neighborhood of v becomes $k_v(k_v - 1)$.

Therefore:

$$C_v = \frac{n_e(v)}{k_v(k_v - 1)}$$

In both cases, a clustering coefficient of $C_v = 1$ will indicate that v and its neighbors form a clique, that is, a fully connected cluster of nodes (see Fig. 1). Inversely, a value of C_v close to zero indicates that v is part of a loosely connected set of nodes.

Global Clustering Coefficient

The global clustering coefficient cL corresponds to the number of triangles τ_Δ found in a set of nodes divided by the total number of connected triplets τ_3 found in this given set. A connected triplet corresponds to three nodes connected by at least two edges. Triangles are cliques of three nodes, in which each node is directly connected to the two others.

Therefore, the intuition behind the global clustering coefficient is to measure the density of connected triplets that form triangles in a graph, which indicates how much edges are clustered together. This is also referred to as the density of a subgraph (Kolaczyk 2009).

For a subset of nodes $v \in V$ in a graph G,

$$\forall v \in V \; s.t. \; \tau_3(v) > 0, \; cL(v \in V) = \frac{\tau_\Delta(v)}{\tau_3(v)}$$

For the nodes v being part of connected triplets (i.e., $\tau_3(v) > 0,$) $C_v = cL(v)$. Therefore, the extension of this definition to a whole graph G, which defines the global clustering coefficient $cL(G)$, is equal to the average of the local clustering coefficient of nodes such that $\tau_3(v) > 0$. Defining $\|V\|$as the size of such set of nodes, then $cL(G)$ is obtained by:

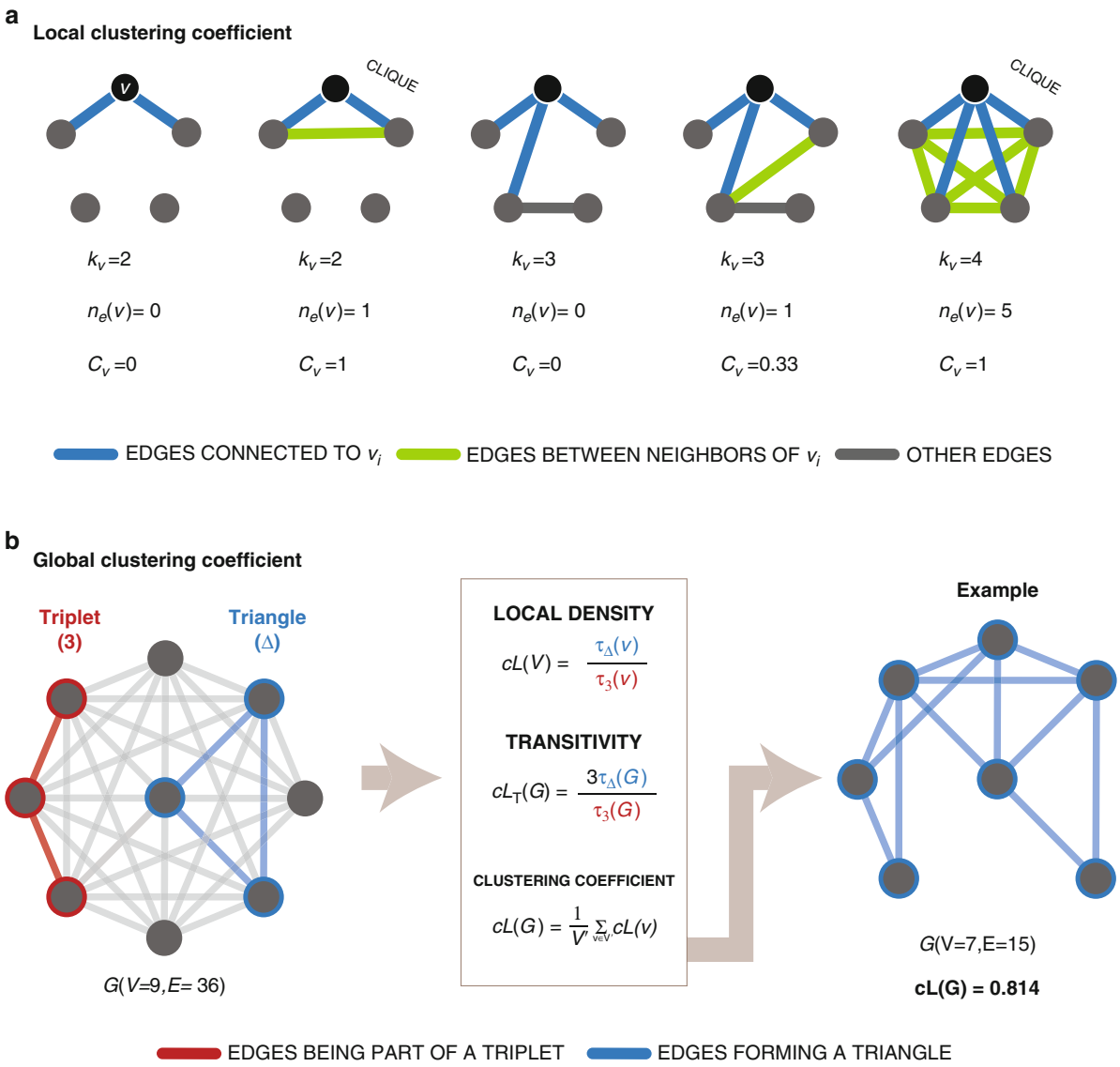
$$cL(G) = \frac{\sum_{v \in V} \tau_\Delta(v)}{\tau_3(G)} = \frac{1}{\|V\|} \sum_{v \in V} C_v$$

This definition is commonly used in network biology (Dong and Horvath 2007), for instance, to measure the modular organization of metabolic networks (▶ Metabolic Networks, Structure and Dynamics).

$cL_T(G)$, also referred to as the transitivity of a graph, is given by:

$$cL_T(G) = \frac{\sum_{v \in V} \tau_3(v) cL(v)}{\tau_3(G)} = \frac{3\tau_\Delta(v)}{\tau_3(G)}$$

As for the local definition of the clustering coefficient, a value of 1 indicates that the graph is fully connected. Clustering coefficients are often used in network biology to measure the cohesion



Clustering Coefficient, Fig. 1 The density of connections in a graph can be approached by local and global definitions of the clustering coefficient. (**a**) The local clustering coefficient represents the density of connections among the neighbors of a node, and ranges from 0 to 1. The higher the value, the more the node is part of a densely connected cluster of nodes. A value of 1 indicates that the node is part of a clique. (**b**) The Global definition of the clustering coefficient relates to the proportion of triplets that form triangles in a graph. The clustering coefficient corresponds to the mean value of the local clustering coefficient (referred to as local density), while the transitivity of a graph gives the probability that the direct neighbors of a node are connected

into modules of protein–protein interactions or metabolic pathways.

Cross-References

► Metabolic Networks, Structure and Dynamics

References

Dong J, Horvath S (2007) Understanding network concepts in modules. BMC Syst Biol 1:24

Kolaczyk ED (2009) Statistical analysis of network data. Number XII. Springer Series in Statistics, pp 94–97

Clustering Methods

Jiguang Wang
Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, Beijing, China

Synonyms

Automatic classification methods; Numerical taxonomy methods; Typological analysis methods; Unsupervised learning methods

Definition

Clustering methods are unsupervised data mining methods that assign objects into groups based on similarity. Clustering methods such as *k*-means clustering, hierarchical clustering, graph partitioning, spectral clustering, etc., have been widely used in marketing, library, image processing, medicine, biology, and other fields.

Characteristics

▶ Clustering or ▶ clustering analysis is the process of putting objects into groups whose members are similar, so an important step in clustering methods is to define similarity or distance.

Similarity Measure

Similarity or distance (dissimilarity) that determines the probability of two objects to be in one group is important in clustering. There are several frequently used methods to measure similarity or distance. In Euclidean n-space, if $x = (x_1, x_2, \cdots, x_n)$, and $y = (y_1, y_2, \cdots, y_n)$, then

- The Euclidean distance is defined as $\sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$
- The Manhattan distance is defined as $\sum_{i=1}^{n} |x_i - y_i|$
- The ▶ Pearson's correlation coefficient is defined as $\frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2}}$

Besides, there are also some other distances, such as Hamming distance, Mahalanobis distance, Chebyshev distance, Minkowski distance, Euclidean distance, Spearman distance, Jaccard coefficient, city block metric, and so on.

Based on similarity, there are many types of clustering methods with different advantages and weaknesses.

K-Means Clustering

Given n numeric points in d dimensional space and fixed integer k, k-means clustering (▶ Clustering, k-Means) attempts to partition these points into k groups so as to minimize the sum of within-cluster sum of deviation

$$\sum_{i=1}^{k} \sum_{x_j \in G_i} \left\| x_j - \mu_i \right\|,$$

where G_i is the ith group, and μ_i is the mean of points in G_i

This problem has been proved to be ▶ NP-hard, and it can be solved in time $O(n^{dk+1} \log n)$ (Inaba et al. 1994), so it is usually solved with heuristic algorithm (▶ Heuristic optimization) as follows:

Step 1: Assign all points to a group by random;
Step 2: Repeat step 2.1 and step 2.2 until stable:
Step 2.1: Compute centroid of each group;
Step 2.2: Reassign each point to its nearest centroid.

k-Means is simply and easy to understand, so it has been widely used and works well in most applications, but the weaknesses are obvious too. The number of groups k should be artificially chosen. k-Means tends to identify spherical clusters and is sensitive to outliers. Centroid positions are easily to be distorted by outliers.

Hierarchical Clustering

Instead of direct partition, hierarchical clustering aims to build a dendrogram (▶ hierarchy) of groups. According to the strategy, hierarchical clustering can be divided into two types, "agglomerative" and "divisive" (Fowlkes and Mallows 1983). "Agglomerative" is a "bottom up" method following two steps:

Step 1: Treat each object as a group.
Step 2: Merge the closest pair of groups until there is only one group left.

By contraries, "divisive" starts from one big group, and follows a "top down" rule to divide groups.

In hierarchical clustering, it is necessary to measure the similarity or dissimilarity of two groups to decide which two groups should be merged or how a group should be divided. Several methods have been developed. Single linkage clustering defines distance between groups as the distance between the closest pair of objects, one from each group, while complete linkage is opposite, defining that as the distance between the most distant pair of objects. Average linkage clustering defines that as the average of distances between all pairs of objects, where each pair is made up of one object from each group. Besides, there are also some other approaches like average group linkage, Ward's hierarchical clustering method, and so on.

Hierarchical clustering is also one of the most commonly used clustering methods. It has the advantage that any valid measure of distance can be used. Comparing with k-means, the observations themselves are not necessary. Hierarchical clustering does not really produce groups, and the user should artificially decide where to split the tree into groups. Besides, hierarchical clustering is also sensitive to noise and outliers.

Spectral Clustering

As described by Luxburg (2007), spectral clustering is one of the most popular modern clustering algorithms. The basic idea of spectral clustering is to make use of the spectrum of the graph Laplacian of data to perform dimension reduction for clustering in low dimension.

The first step in spectral clustering is to transform similarities or distances into a similarity graph to model the local neighborhood relationships between objects. There are several methods to construct the graph. ε-neighborhood graph connects two objects when their distance is smaller than given ε; k-nearest neighbor connects an object with k nearest neighbors, and kernel-based fully connected graph connects each pair of objects with a weighted edge. Finally, a similarity graph matrix W is obtained to represent local relationships between objects. The graph Laplacian is defined as

$$L = D - W,$$

where D is a diagonal degree matrix with $d_{ii} = \sum_{j=1}^{n} w_{ij}$.

Based on the graph Laplacian, given number k of clusters, unnormalized spectral clustering algorithms compute the first k eigenvectors to represent the original n objects with k-dimension vectors. Then use k-means algorithm to partition all these vectors into k clusters. Similarly, there are also revised algorithms like normalized spectral clustering according to Shi and Malik (2000) and normalized spectral clustering according to Ng et al. (2002).

Spectral clustering is successful mainly based on the fact that it does not make strong assumptions on the form of the clusters, but at the same time spectral clustering can be quite unstable under different choices of the parameters or neighborhood graphs. It should be used with care.

In addition to these classical clustering methods, there are also some other related methods including fuzzy c-means clustering (Bezdek and Ehrlich 1984), graph-theoretic methods (Sharan and Shamir 2000), biclustering (Madeira and Oliveira 2004), probabilistic clustering (Nikhil et al. 2005), Markov clustering (Dongen 2000), support vector clustering (Ben-Hur et al. 2001), and so on. Clustering methods are useful in exploring data, but most methods are still very ad hoc, depending on choosing proper similarity metric and parameters.

Cross-References

- Clustering
- Clustering, k-Means
- Heuristic Optimization
- Heuristic Optimization
- Hierarchy
- NP-Hard
- Pearson Correlation Coefficient

References

Ben-Hur A, Horn D, Siegelmann HT, Vapnik V (2001) Support vector clustering. J Mach Learn Res 2:125–137

Bezdek JC, Ehrlich R (1984) FCM: the fuzzy c-means clustering algorithm. Comput Geosci 10(22):191–203

Dongen SV (2000) Graph clustering by flow simulation. Ph.D. thesis. University of Utrecht

Fowlkes EB, Mallows CL (1983) A method for comparing two hierarchical clusterings. J Am Stat Assoc 78(384):553–584

Inaba M, Katoh N, Imai H (1994) Applications of weighted voronoi diagrams and randomization to variance-based k-clustering. Proceedings of 10th ACM symposium on computational geometry. 332–339. doi:10.1145/177424.178042

Luxburg UV (2007) A tutorial on spectral clustering. Stat Comput 17(4):395–416
Madeira SC, Oliveira AL (2004) Biclustering algorithms for biological data analysis: a survey. IEEE Trans Comput Biol Bioinform 1(1):24–45
Ng A, Jordan M, Weiss Y (2002) On spectral clustering: analysis and an algorithm. In: Dietterich T, Becker S, Ghahramani Z (eds) Advances in neural information processing, systems 14. MIT, Cambridge, pp 849–856
Pal NR, Pal K, Keller JM, Bezdek JC (2005) A possibilistic fuzzy c-means clustering algorithm. IEEE Trans Fuzzy Syst 13(4):517–530
Sharan R, Shamir R (2000) CLICK: a clustering algorithm with applications to gene expression analysis. In: Proceedings of ISMB'00, pp 307–316
Shi J, Malik J (2000) Normalized cuts and image segmentation. IEEE Trans Pattern Anal Mach Intell 22(8):888–905

Clustering, Hierarchical

Larissa Stanberry
Bioinformatics and High-throughput Analysis Laboratory, Seattle Children's Research Institute, Seattle, WA, USA

Definition

Hierarchical clustering represents a collection of methods which seeks to partition the data into groups by building a hierarchy of clusters.

Characteristics

The goal of a clustering method is to partition the data into distinct groups where each group contains objects with similar characteristics. Hierarchical clustering refers to the collection of methods which builds a hierarchy of clusters. The hierarchical clustering methods are subdivided into two categories

- Agglomerative, where each datum initially represents a cluster and the two clusters are merged together at each step of the hierarchy. Hence, the hierarchy is built from the bottom up.
- Divisive, where initially all data represents a single cluster that is split as one moves down the hierarchy.

Agglomerative clustering has been extensively studied and applied in numerous studies. Hierarchical clustering is simple to apply and in comparison to other clustering methods makes no parametric assumptions about the size, shape, or quantity of clusters (Jain and Dubes 1988).

Dissimilarity Measure

Unlike k-means clustering (▶ Clustering, k-means), hierarchical clustering does not require one to specify the number of clusters a priori. In contrast, it uses a dissimilarity measure.

The choice of the dissimilarity measure depends on the data type (nominal, ordinal, ratio etc.). Common examples include Euclidean distance, Mahalanobis distance, Manhattan distance, and others.

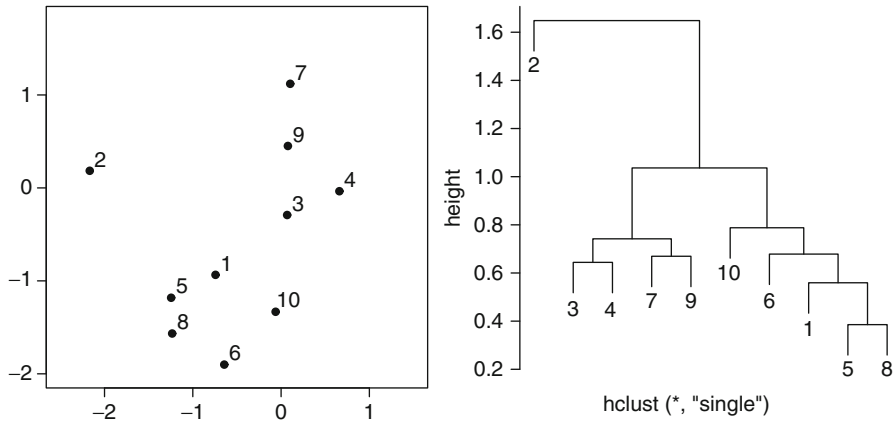
Linkage Method

Recursive splitting or merging of the clusters can be represented by a binary tree known as dendrogram. The tree provides a simple and convenient visualization of data hierarchy. Based on the dissimilarity measure, the data are grouped into a hierarchical tree called dendrogram using a linkage method. The three most popular linkage methods in hierarchical agglomerative clustering are

- Single linkage, or the nearest neighbor, method. Here, the dissimilarity between two groups is defined as the minimum pairwise distance of points in the two groups.
- Complete linkage, or the furthest neighbor, where the dissimilarity between two clusters is taken as the maximum distance between points in the clusters.
- Average linkage uses the average dissimilarity between the groups.

If the grouping structure is compact with well-separated clusters, all of the three clustering methods produce a similar result.

Single linkage has certain advantages as compared to other methods. In particular, it is known to correctly identify the underlying structure of the data. This is not necessarily the case for the other linkage methods, unless points of high density are sufficiently separated. For example, when sets of original points become close or overlap, the complete, average, and K-means clustering (▶ Clustering, k-means) yield several large clusters, giving an impression of distinct grouping in the data regardless of the density. It has been shown that the large clusters produced by these algorithms depend on the range, but not on the true density of the data set. Single linkage behaves differently, producing



Clustering, Hierarchical, Fig. 1 The binary tree for the simulated example

long-chained clusters that are difficult to interpret. The "chaining" effect indicates the lack of spatial separation in case of touching or overlapping clusters.

In addition, the single linkage is fully consistent for separating two disjoint, high-density clusters in one dimension, and is fractionally consistent for data in higher dimensions. The fractional consistency means that if two disjoint population groups exist, there will be two distinct single linkage clusters containing a positive fraction of the sample points from the corresponding population groups. Hence, the single linkage is in a sense conservative, as it will not necessarily detect all clusters, but will identify modal regions separated by a sufficiently deep valley (Hartigan 1975, 1977).

Cluster Identification

The clusters are typically defined by setting an appropriate inconsistency threshold, which is equivalent to cutting the dendrogram at a certain level. For that the length of each link in a cluster tree is compared with the length of neighboring links below it. If the length of the link does not differ significantly from the length of the neighboring links, it means that objects, joined at this level of the hierarchy, have similar characteristics. Large differences would indicate some dissimilarity between the objects, and the link is said to be inconsistent. Alternatively, one can consider an adaptive rule that determines inconsistency threshold for each link.

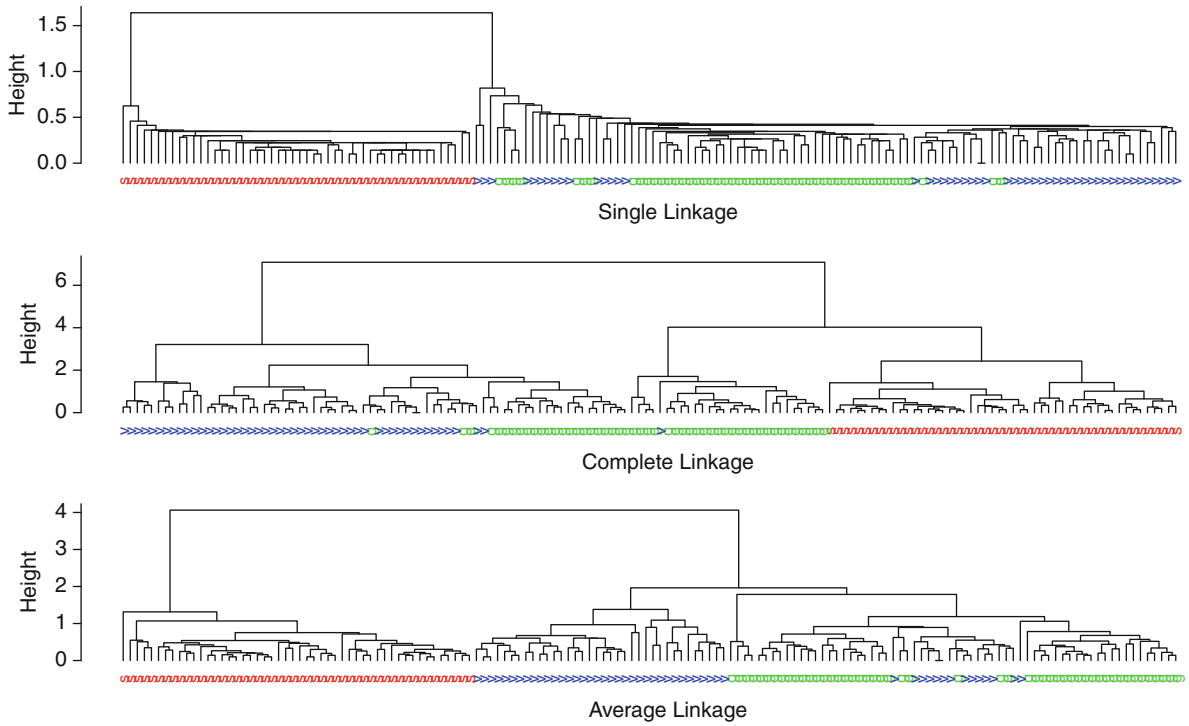
Example

To gain a better understanding, consider the following example of the single linkage algorithm for a small data set consisting of ten two-dimensional points labeled from 1 to 10 (Fig. 1). Initially every point is a cluster. The first step is to compute all pairwise distances between the points. Here, we use the Euclidean. The two closest points are 5 and 8 and will be merged to form a new cluster**.** The next smallest distance is between points 1 and 5, but 5 has already been merged with 8, thus 1 is added to a cluster $\{5, 8\}$. Next, points 3 and 4 are merged together and so on. The process continues until all the nodes form a single cluster. The result is the single linkage dendrogram shown in the left plot. The original points are marked on the tree.

The single linkage clusters can be obtained by cutting the tree at a certain level. In the example, deleting the largest edge will yield to two clusters a singleton $\{2\}$ and a cluster containing all the remaining nine points. Cutting, for example, at 0.9 would give three clusters $\{2\}$, $\{3,4,7,9\}$, and $\{10,6,1,5,8\}$.

To demonstrate the performance of the hierarchical clustering, we classify the famous iris data set that gives the measurements (in cm) of the four variables, including sepal length and widths, and petal length and width. The data was collected on the three species of Iris setosa, versicolor, and virginica and is freely available as a part of R package.

Figure 2 shows the dendrogram trees for the single, complete, and average linkage. The three iris species are color coded to show their relative positions in the tree. The average and single-linkage trees look similar misplacing only a few versicolor species into the virginica branch. The complete linkage performs the worst as it identifies half of the virginica species to be closer to the setosa and half to the versicolor. The



Clustering, Hierarchical, Fig. 2 Binary trees for the iris data for the single-, complete and average linkage methods. The three iris species were color-coded

misclassification error is about 9% for the average and single linkage and is 16% for the complete linkage method.

Divisive Clustering

Divisive clustering was studied and used considerably less extensively. The divisive method can be accomplished by recursively subdividing groups into two clusters. The splitting continues until every cluster is a singleton. The divisive clustering does not necessarily leads to the convenient representation of the data as a binary tree, although some methods were proposed to address this issue.

Clustering Software and Tools

There are a number of existing software that offer various types of hierarchical clustering. The R Project for Statistical Computing offers a number of clustering options; for details see the CRAN task view on Cluster Analysis and Mixture Models. The statistical package in Matlab (Mathworks) includes functions for agglomerative hierarchical clustering. ELKI package implements a single-linkage clustering among others. Orange is a free data-mining software that can be used for clustering. Hierarchical Clustering Explorer is another useful tool for data visualization and cluster analysis.

Cross-References

- ▶ Clustering, k-Means
- ▶ Clustering, Model-Based
- ▶ Model Testing, Machine Learning

References

Hartigan JA (1977) Distribution problems in clustering. In: van Ryzin J (ed) Classification and clustering. Academic, New York

Hartigan JA (1975) Clustering algorithm. Wiley, New York

Jain AK, Dubes RC (1988) Algorithms for clustering data. Prentice-Hall, Englewood Cliffs

Clustering, k-Means

Larissa Stanberry
Bioinformatics and High-throughput Analysis Laboratory, Seattle Children's Research Institute, Seattle, WA, USA

Synonyms

k-means

Definition

K-means is an unsupervised clustering algorithm for classifying data points into *K* distinct groups.

Characteristics

K-means algorithm

The *K*-means algorithm is based on the dissimilarity measure given by squared Euclidean distance, i.e., for *p*-dimensional data vectors $x_1 = (x_{11} \ldots,,x_{1p})$ and $x_2 = (x_{21}, \ldots, x_{2p})$

$$d(x_1, x_2) = \sum_{i=1}^{p} (x_{i1} - x_{2i})^2$$

In *K*-means, the clusters are determined by minimizing the loss function

$$L(C) = \sum_{k=1}^{K} N_k \sum_{C(i)=k} d(x_i, \bar{x}_k),$$

where $k = 1, \ldots, K$ indexes clusters, $\bar{x}_k = (\bar{x}_{1k}, \ldots, x_{pk})$ is the mean vector of cluster k, and N_k is the total number of points in cluster k. The loss function $L(C)$ is minimized when, within each cluster, the average distance from any point to the cluster mean is minimized.

In *K*-means, the cluster labels are determined by an iterative algorithm that alternates between the two steps:

1. Given the set of *K* means, assign each observation to the cluster with the closest mean.
2. Update the *K* means by computing an average of the observations within each cluster.

The algorithm converges when the cluster assignments no longer change. The iteration steps reduce the value of the loss function thus assuring convergence. To initialize, one can either randomly choose *K* points to represent cluster means or randomly assign *K* labels to *N* observations.

Note that the cluster assignment may be suboptimal corresponding to the local rather than global minimum of the loss function. To elevate this problem, one can do multiple cluster runs with random initializations and choosing the final partition with the smallest value of the loss function (Hastie et al. 2009).

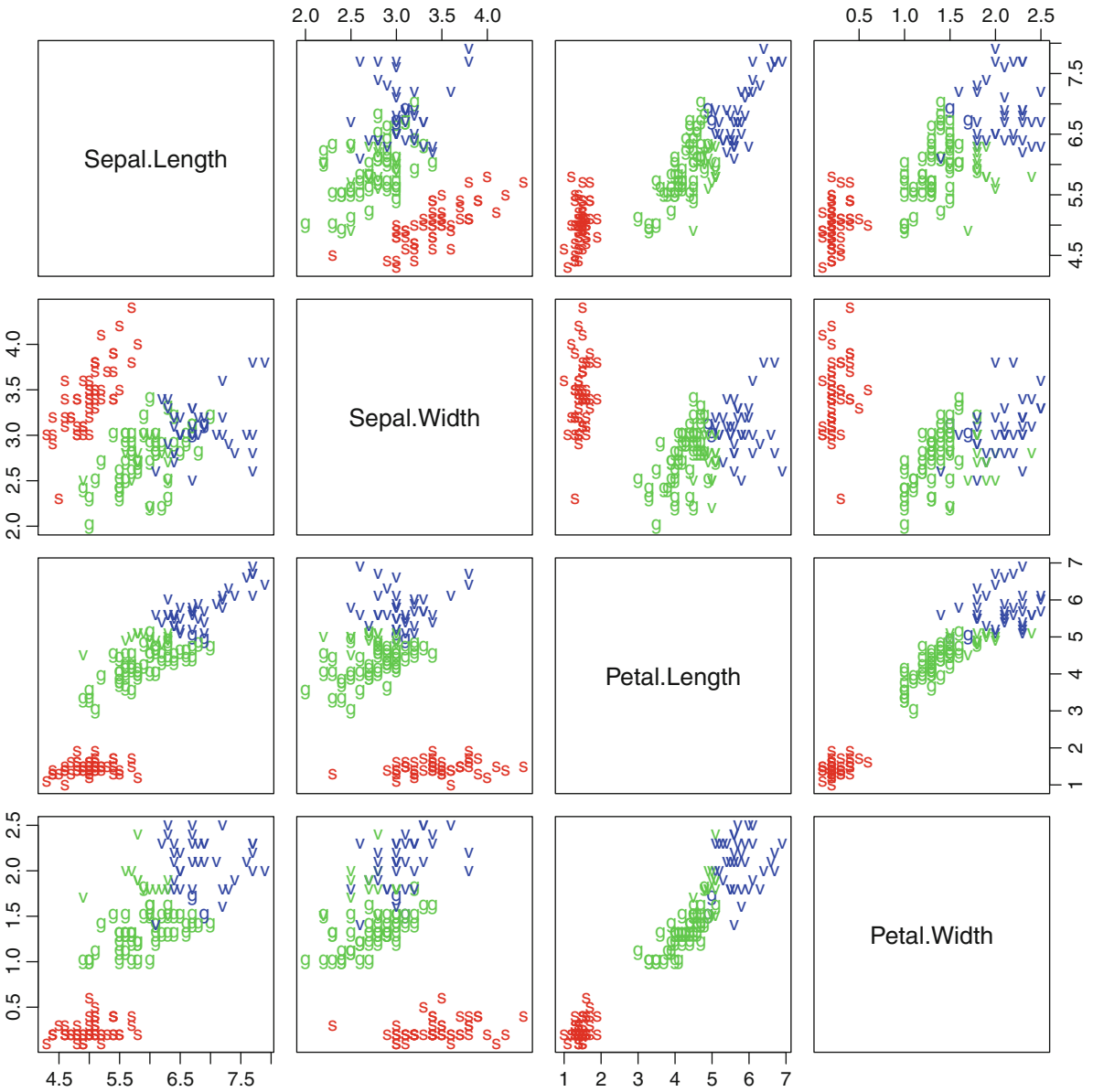
Example

Here we demonstrate the performance of the *K*-means clustering algorithm on the famous iris data set that gives the measurements (in cm) of the four variables including sepal length and widths, and petal length and width. The data was collected on the three species of *Iris setosa*, *Iris versicolor*, and *Iris virginica*. The data set is freely available as a part of R package.

Figure 1 shows the results of clustering the data into four groups using *K*-means method with 25 random initializations. Here, *s, v, g* stand for the three different species setosa, versicolor, and virginica. The cluster assignments are marked by the three different color. Clearly, *K*-means exactly identifies setosa species, but makes some misassignments by classifying some versicolor with virginica and vice versa. This is expected, since the setosa species are well separated, while the other two have some overlap across different measurements. The misclassification error for the *K*-means classification is about 11%.

Software

The *K*-means clustering algorithm is available in a number of software packages both free and commercial including R, S-plus, SAS, Apache Mahout, Matlab, and Mathematica.



Clustering, k-Means, Fig. 1 Color-coded *K*-means cluster assignments for the three different iris species setosa, *s*, versicolor, *v*, and virginica, *g*

Conclusion

The *K*-means clustering is an efficient algorithm for data classification. The method is intuitively simple and is easy to implement. The *K*-means clustering, however, has a number of assumptions and performs poorly when these assumptions are violated. In particular, the method fails when clusters are nonspherical or differ considerably in size. Furthermore, the algorithm requires a prior knowledge on the number of clusters which is not readily available in practice. Numerous variations have been proposed to improve the performance of the algorithm.

References

Hastie T, Tibshirani R, Friedman J (2009) The elements of statistical learning, 2nd edn. Springer, New York, pp 509–510

Clustering, Model-based

Larissa Stanberry
Bioinformatics and High-throughput Analysis Laboratory, Seattle Children's Research Institute, Seattle, WA, USA

Synonyms

Mixture model

Definition

Model-based clustering is a classification technique where the data is viewed as arising from the underlying mixture of probability distributions with each mixture component representing a cluster.

Characteristics

Overview

Model-based clustering treats data as arising from the mixture of probability distributions. The mixture components are assumed to correspond to distinct clusters. The goal of the clustering approach is to classify data into distinct groups by assigning every data point to a mixture component.

More specifically, if $y = (y_1, y_2, y_n)$ is the observed data, the finite mixture model has a form

$$L(y; \Theta, K) = \prod_{i=1}^{n} \sum_{k=1}^{K} p_k f_k(y_i; \theta_k)$$

where K is the number of mixture components, p_k is the probability that an observation belongs to the kth component, $f_k(\cdot; \theta_k)$ is the density of the kth component with parameter θ_k, and $\Theta = (\theta 1, \theta_2, \ldots, \theta_K)$ are the parameters of the K components. Note that $p_k \geq 0$ and $\sum_{i=1}^{K} p_k = 1$.

The component densities f_k are chosen to best reflect the data-generating mechanism. The most popular approach is the Gaussian mixture model, where the data is viewed as a sample from the mixture of normal distributions. Other examples include beta mixtures, gamma mixtures, etc.. The K-means clustering algorithm is implicitly based on the Gaussian mixture model.

Implementation

For estimation purposes, the data is specified as (y_i, z_i) where $z_i = (z_{i1}, z_{i2}, \ldots, z_{iK})$ is the vector of binary variables such that $z_{ik} = 1$, if y_i is in cluster k, and zero otherwise. The membership vector z_i is considered as a sample from the multinomial distribution of a single draw from K groups with probabilities $(p_1, p_2 \ldots, p_k)$. By maximizing the likelihood function for the data (y_i, z_i), one estimates $\hat{z}_{ik}$, which is the conditional probability of observation y_i being in cluster k. In hard clustering, point y_i is ultimately assigned to a cluster with the largest value of $\hat{z}_{ik}$. In fuzzy clustering, the data point is not assigned to any particular cluster, rather the results are returned as a vector $(\hat{z}_{i1}, \hat{z}_{i2}, \ldots, \hat{z}_{iK})$ which gives the probabilities of the ith point being in any one of the K clusters.

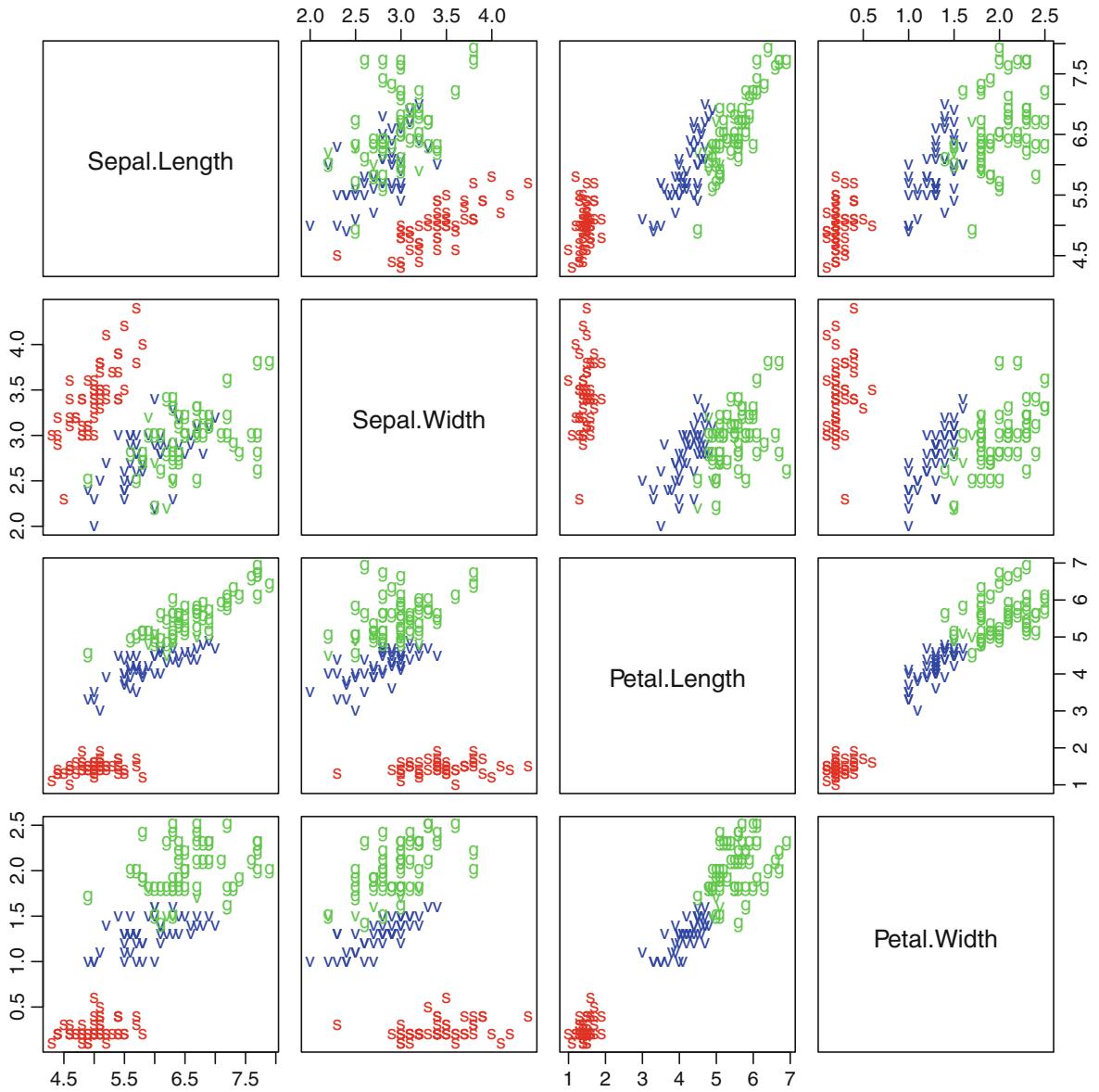
The model-based clustering can be solved using the expectation maximization (EM) algorithm which iterates between computing the estimates $\hat{z}_i$ given the current parameter estimates and estimating the parameters by maximizing the likelihood given the current estimates $\hat{z}_i$ (Fraley and Raftery 2002).

Estimating Number of Clusters

In real data, the number of clusters is typically unknown. One can estimate K by comparing models with different number of mixture components. The idea is to select a criterion to assess the model fit and choose the partition that corresponds to the model with the highest value of the criterion. The most popular criterion choices include Bayesian information criterion (BIC), Akaike information criterion, and the minimum description length.

Software

An R package for normal mixture modeling MCLUST can be used for model-based clustering. The MCLUST package includes model-based clustering, the implementation of the EM algorithm for four different covariance mixture models, and the BIC approach to estimating the model and the number of clusters.



Clustering, Model-based, Fig. 1 Scatter plot of the three clusters identified in the iris data using model-based clustering. Clusters are color-coded and the three species are labeled, i.e., setosa, *s*, versicolor, *v*, and virginica, *g*

The package is well documented and provides enhanced display and visualization tools.

Example

Figure 1 shows the classification of the iris data using the model-based clustering approach with unconstrained covariance model as implemented in MCLUST. The species are labeled with letters and the three clusters are color-coded. Clearly, there is a good agreement between the true data and clustering results. The misclassification rate is 3.3%.

Cross-References

- Bayesian Information Criterion (BIC)
- Clustering
- K-Means

► Model Testing, Machine Learning
► Model Validation

References

Fraley C, Raftery A (2002) Model-based clustering, discriminant analysis and density estimation. J Am Stat Assoc 97:611–631

CMP Motif

► Combinatorial Molecular Phenotypes (CMPs)

c-Myc

Bernhard Schmierer
Department of Biochemistry, Oxford Centre for Integrative Systems Biology (OCISB), University of Oxford, Oxford, UK

Definition

c-Myc is a proto-oncogene and member of the Myc family of transcription factors. c-Myc is a key activator of cell proliferation in response to growth factors and, when activated by mutation or overexpression, can drive unrestricted proliferation. Hyperactivation of c-Myc is observed in many cancers.

Cross-References

► Cell Cycle Signaling, Hypoxia

Coalgebra

Baltasar Trancón y Widemann
Ecological Modelling, University of Bayreuth, Bayreuth, Germany

Definition

Coalgebra is the field of study of mathematical structures that are dual to structures studied in the field of algebra. The term *dual* is used in the formal sense of category (► Mathematical Structure, Category) theory, namely, that the defining operations are reversed.

Characteristics

Coalgebra for Classical Algebra

Classical algebra studies structures endowed with binary operations that obey various properties: associativity in semigroups, existence of neutral and inverse elements in groups, distributivity in rings, commutativity in Abelian algebras, absorption in Boolean algebras, etc. The corresponding coalgebraic structures are endowed not with operations that produce a single element as a combination of two but conversely with operations that produce a pair of elements as a decomposition of one. Applications of coalgebraic structures in science are still extremely scarce, compared with their algebraic counterparts. The most important instance is the use of Hopf coalgebras alongside Hopf algebras in standard model particle physics (de Wild Propitius and Bais 1995). In biological sciences, coalgebra has been used for formal accounts of genetics (Tian and Li 2004). There, the operation of genetic combination from parents to offspring is reversed to a deduction of genetic traits of the parents from those of their offspring.

Coalgebra for Universal Algebra

Universal algebra, more abstractly, studies properties common to all the aforementioned structures and many others beside. It has been recognized only at the end of the twentieth century that, when these universal properties are dualized rather than their specific manifestations, many structures of systems theory and computer science come into scope, most notably transition systems (Aczel 1988) and automata (Adámek and Trnková 1990). This insight has attracted theoretical computer scientists to the field, culminating in the proposal by Rutten (2000) of universal coalgebra as a general theory of system dynamics, with applications to stochastic dynamics (► Markov Chain) and control.

In the language of category theory, both algebras and coalgebras are studied relative to a ► functor F that specifies an expression language.

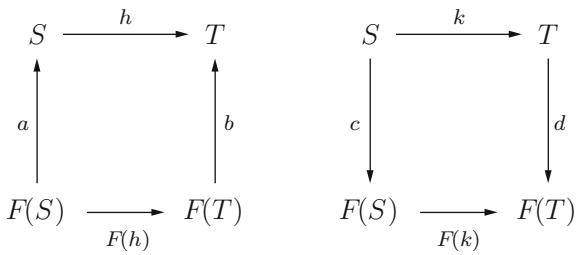
An F-algebra is a carrier set S together with an operation a: $F(S) \to S$; that is, single elements in S

are produced by combination according to composite expressions in $F(S)$. Dually, an F-coalgebra is a carrier set S together with an operation c: $S \to F(S)$; that is, single elements in S produce decompositions into composite expressions in $F(S)$.

The appearance of the carrier set S on the left or right hand side of a mapping makes a crucial difference for the interpretation of the operation: Algebras formalize the construction and hence the causal explanation of elements (▶ Causality; ▶ Reduction). Coalgebras, on the other hand, formalize the characterization and hence the behavioral explanation of elements; they may take existence as granted and are hence indifferent to infinitely regressing and circular dependencies (▶ Holism).

Mappings and Relations

The notion of *homomorphisms* as structure-respecting mappings, classically defined individually for each class of algebras, can be generalized in the categorical language to all functors F and to both algebra and coalgebra as commuting diagrams. The diagrams for a homomorphism h between two F-algebras and for a homomorphism k between two F-coalgebras are as follows:



(1)

The same statements can be rephrased in equational form as

$$h \circ a = b \circ F(h) \qquad F(k) \circ c = d \circ k \qquad (2)$$

respectively, where $\circ$ denotes the composition of mappings.

Distinguished (Co)algebras

F-algebras and their homomorphisms form a category, and so do F-coalgebras and their homomorphisms.

For many functors F, the category of F-algebras has an *initial* object: There exists an F-algebra that admits a unique homomorphism *to* any other F-algebra and is isomorphic to the *term algebra* that consists of finitely nested composite F-expressions. Thus, universal algebra is equipped with a universal syntactic structure for constructions, and the corresponding unique homomorphism is their systematic bottom-up interpretation.

Dually, for many functors F, the category of F-coalgebras has a *final* object: There exists an F-coalgebra that admits a unique homomorphism *from* any other F-coalgebra. In contrast to the term algebra, it contains also infinite and circular expressions in the spirit of the non-well-founded set theory of Aczel (1988). It can be thought of as the space of all possible ways of global system behavior (Jacobs and Rutten 1997), and the corresponding unique homomorphism is their systematic top-down representation.

Application Examples

As an example, consider the functor $F(X) = 1 + X$ on the category of sets, which adds an extra element, say *, to each set. Its initial algebra consists of the set $\mathbf{N}$ of nonnegative integers and an operation that assigns zero to * and the successor to each number. Every deterministic discrete dynamical system can be represented as an F-algebra, consisting of the state space S and an operation that assigns a distinguished initial state x_0 to * and the successor to each state. The unique homomorphism $h : \mathbf{N} \to S$ then computes the *trajectory of* x_0.

Dually, the final coalgebra for the same functor F consists of the set $\mathbf{N} \cup \{\infty\}$ of extended nonnegative integers and the predecessor operation that maps ∞ to itself and zero to *. Every subset $U \subset S$ of states of a dynamical system can be represented as an F-coalgebra, consisting of S and the evolution function of the system, but restricted to U and mapping all other states to *. The unique homomorphism $h : S \to \mathbf{N} \cup \{\infty\}$ then assigns to each state the number of steps the system may take before leaving the region U. U is *stationary* if and only if $h(x) = \infty$ for all $x \in U$.

Equivalence

Elements of different F-algebras may be related by originating from the same element of the initial F-algebra via the respective unique homomorphisms. In that case, one may speak of them as *constructively*

equivalent. Elements of different F-coalgebras may be related by being mapped to the same element of the final F-coalgebra via the respective unique homomorphisms. In that case, one may speak of them as *behaviorally equivalent*.

The two relationships are logically similar to the concepts of *homology* and *analogy* of comparative biology. The genuinely coalgebraic notion of formal behavioral equivalence has proven very useful in the abstract modeling of systems and processes in computer science (Milner 1989). The same abstraction can be used to circumvent the issue of ▶ multiple realization of biotic phenomena, by modeling and formalizing the essential behavior rather than the accidental structures that cause it.

Cross-References

- ▶ Category Map
- ▶ Causality
- ▶ Functor
- ▶ Holism
- ▶ Markov Chain
- ▶ Multiple Realization
- ▶ Reduction

References

Aczel P (1988) Non-well-founded sets. CSLI, Stanford University Press, Stanford
Adámek J, Trnková V (1990) Automata and algebras in categories. Kluwer, Dordrecht/Boston
de Wild Propitius M, Bais FA (1995) Discrete gauge theories. Report 95-46, LPTHE, Paris, arXiv:hep-th/9511201v2
Jacobs B, Rutten J (1997) A tutorial on (co)algebras and (co) induction. EATCS Bull 62:222–259
Milner R (1989) Communication and concurrency. Prentice Hall, New York
Rutten J (2000) Universal coalgebra: a theory of systems. Theor Comput Sci 249(1):3–80. doi:10.1016/S0304-3975(00)00056-6
Tian J, Li B (2004) Coalgebraic structure of genetic inheritance. Math Biosci Eng 1(2):243–266

Coefficient of Correlation

▶ Correlation Coefficient

Co-expression

Jiguang Wang
Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, Beijing, China

Definition

Co-expression means the simultaneous expression of two or more genes. Identifying co-expression genes is an important method in exploring microarray data. Yu et al. (2003) show that genes targeted by the same transcription factors tend to show similar expression profiles. Also, interacting proteins are often significantly co-expressed (Jansen et al. 2003).

References

Jansen R, Yu HY, Greenbaum D, Kluger Y, Krogan NJ, Chung S, Emili A, Snyder M, Greenblatt JF, Gerstein M (2003) A Bayesian networks approach for predicting protein-protein interactions from genomic data. Science 302(5644):449–453
Yu HY, Luscombe NM, Qian J, Gerstein M (2003) Genomic analysis of gene expression relationships in transcriptional regulatory networks. Trends Genet 19(8):422–427

Co-expression Network

▶ Functional/Signature Network Module for Target Pathway/Gene Discovery

Cofactors

Tetsuro Kokubo
Department of Supramolecular Biology, Graduate School of Nanobioscience, Yokohama City University, Yokohama, Kanagawa, Japan

Synonyms

NC1; NC2; PC1; PC2; PC3; PC4

Definition

USA (*U*pstream factor *S*timulatory *A*ctivity) was originally identified as an activity promoting a greater level of transcriptional activation by its dual function of repressing basal transcription and potentiating activator-dependent transcription (Kaiser and Meisterernst 1996; Roeder 1998; Thomas and Chiang 2006). Further fractionation of USA revealed that it contained four positive cofactors (PC1, PC2, PC3, and PC4) and one negative cofactor (NC1). PC1, PC2, PC3, and NC1 were later found to be equivalent to poly(ADP-ribose) polymerase-1, Mediator, DNA topoisomerase I and HMG1, respectively. PC1 stimulates PIC formation at a post-TFIID binding step, and PC3 enhances TFIID-TFIIA-DNA complex formation. PC4 has a single-stranded DNA binding activity and plays a general role in the stabilization of various protein-DNA interactions. NC1 binds to the TBP-TATA complex and prevents the incorporation of TFIIB into PIC. Notably, each component of USA shows dual functions (positive and negative), depending on whether activators are present or absent, similar to that observed in the original mixed fraction of USA.

NC2 was isolated from a chromatographic fraction different from that of USA, as a factor that could associate with the TBP-TATA complex. This factor comprises two subunits, NC2α and NC2β, which interact with each other via their HFD (*h*istone *f*old *d*omain). Interestingly, NC2 inhibits TATA-dependent transcription while it promotes DPE-dependent transcription. The negative function may be due to its inhibitory effect on the incorporation of TFIIA and TFIIB into the PIC, and the positive function may be related to its remarkable activity that mobilizes TBP on DNA.

Cross-References

► Transcription in Eukaryote

References

Kaiser K, Meisterernst M (1996) The human general co-factors. Trends Biochem Sci 21(9):342–345

Roeder RG (1998) Role of general and gene-specific cofactors in the regulation of eukaryotic transcription. Cold Spring Harb Symp Quant Biol 63:201–218

Thomas MC, Chiang CM (2006) The general transcription machinery and general cofactors. Crit Rev Biochem Mol Biol 41(3):105–178

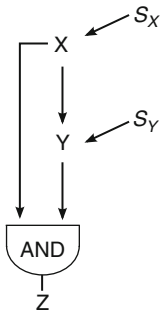
Coherent Type I FFL

Jinzhi Lei
Zhou Pei-Yuan Center for Applied Mathematics, Tsinghua University of Beijing, Beijing, China

Definition

In the coherent type 1 FFL (C1-FFL), both transcription factors X and Y are transcriptional activators (Fig. 1). The C1-FFL is a "sign-sensitive delay" element and a persistence detector. This means that this motif can provide pulse filtration in which short pulses of signal will not generate a response but persistent signals will generate a response after short delay (Mangan and Alon 2003).

The C1-FFL has opposite effect in the OR input function comparing with those in the AND input function. With an AND input function, the C1-FFL shows a delay after stimulation, but no delay when stimulation stops, while with an OR input function, the C1-FFL shows no delay after stimulation, but does show a delay when stimulation stops (Mangan and Alon 2003; Alon 2007).



Coherent Type I FFL, Fig. 1 The coherent type-1 FFL with an AND input function at the Z promoter. S_X and S_Y are input signals for X and Y

References

Alon U (2007) Network motifs: theory and experimental approaches. Nat Rev Genet 8:450–461
Mangan S, Alon U (2003) Structure and function of the feed-forward loop network motif. Proc Natl Acad Sci USA 100:11980–11985

Cohesin

Rosella Visintin
IEO, European Institute of Oncology,
Milan, Italy

Definition

Cohesin is a hetero-tetrameric complex organized in a ring-like shape whose main function is to hold sister chromatids together from DNA replication up to anaphase.

Cross-References

► Mitosis

Cold and Flu-Like Illness

► Viral Respiratory Tract Infections

Collaborative and Distributed Biomedical Applications

Tim Clark
Department of Neurology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA

Synonyms

Biomedical web communities; Collaboratories; Virtual laboratories

Definition

Collaborative and distributed biomedical applications on the *World Wide Web* or biomedical "collaboratories" (Olsen et al. 2008) enable and socialize multimodal activities critical to biomedical research, using social media technologies. They enable near-real-time group discussion of methods and findings, sharing of reagents, cooperative construction of databases, and publication and annotation of new scientific communications.

Characteristics

The use of the Internet to support scientific communication is one of the major shifts in the practice of science in this era (Kling 1999).

Collaborative and distributed web applications (or "collaboratories") provide biomedical researchers and clinicians with online facilities for very-large-scale cross-disciplinary biomedical research communication. They support essential scientific activities of reporting, documenting, searching, sharing, validating, combining, collectively witnessing, critiquing, reviewing, reproducing, integrating, and extending the results and methods of experiment, observation, and theoretical endeavor. As such they are part of an ongoing transition in science generally from print-based to web-based scientific communications (Borgman 2007).

Many biomedical web communities are disease-focused, such as the pioneering Alzforum (Kinoshita and Clark 2007) as well as newer offerings such as Pain Research Forum (http://painresearchforum.org) and SFARI (https://sfari.org/).

Other collaboratories – many of them *wikis* – focus on specific biological entities, such as WikiProteins (http://conceptwiki.org/index.php/WikiProteins); domain *ontologies*, such as NeuroLex (http://neurolex.org/wiki); or protocols, such as OpenWetWare (http://openwetware.org/) (Waldrop 2008).

Collaborative workflow repositories such as myExperiment (http://myexperiment.org) support archival, retrieval, and discussion of analytical workflows used in computational systems biology and other data-intensive disciplines (Goble and

DeRoure 2007). They provide distributed support for detailed scientific discourse on the computational methods used to arrive at published results (Gil et al. 2008).

Collaboratories markedly reduce time and distance constraints upon participants, while providing essential computational and communications support for large-scale cross-disciplinary integration of biological and clinical findings on a global scale.

Collaboratories are thus core infrastructure for what Weston and Hood (Weston and Hood 2004) term "studying biological systems on a global level," ideally allowing "as many levels of information as possible to be integrated," and are thus fundamental enabling technology for the development of systems biology.

Scientific blogs, such as *Science in the Open* (http://cameronneylon.net/), *Open Reading Frame* (http://www.sennoma.net/) or *RRResearch* (http://rrresearch.fieldofscience.com/), are another important form of scientific social media. However, unlike collaboratories, which orient strongly toward collective establishment and refinement of scientific methods, hypotheses, models, and findings, scientific blogs typically support group discussion of the periodic publications of a single individual.

An important feature of biomedical and other collaboratories is the open-access nature of their content licensing. Creative commons or "CC" (http://creativecommons.org/) licenses are standardized and widely recognized legal models for open-access content sharing. These licenses are frequently used for collaboratory content, with "CC-BY" (free to share with attribution to the original author) a popular option. CC licenses enable legally sound, free of charge, and end-to-end transactions in copyrighted works on the web (Carroll 2006).

Collaboratories are increasingly supported by standard web content management systems (WCMS), providing reusable and customizable features, such as Drupal (http://drupal.org); by wiki platforms such as MediaWiki (http://www.mediawiki.org) and Semantic Mediawiki (http://semantic-mediawiki.org/); or by web application frameworks such as Ruby on Rails (http://rubyonrails.org).

WCMS used to support collaboratories in bioscience are, as a rule, "open source software" – that is, their license terms provide for free use, reuse, modification, and redistribution. In an open source WCMS, the community of its software developers is typically organized online using its own software development collaboratory (e.g., http://drupal.org).

Informatics to support collaboratories can now be considered a distinct research agenda in itself (Lee et al. 2003).

Cross-References

- ▶ Ontologies
- ▶ Wiki
- ▶ World Wide Web

References

Borgman CL (2007) Scholarship in the digital age: information, infrastructure, and the internet. MIT Press, Cambridge, MA

Carroll MW (2006) Creative commons and the new intermediaries. Mich St L Rev 45:55–59

Gil Y, Deelman E, Ellisman M, Fahringer T, Fox G, Gannon D, Myers J (2008) Examining the challenges of scientific workflows. IEEE Computer 40(12):24–32

Goble CA, DeRoure DC (2007) myExperiment: social networking for workflow-using e-scientists. Paper presented at the proceedings of the 2nd workshop on workflows in support of large-scale science, Monterey, CA

Kinoshita J, Clark T (2007) Alzforum. Methods Mol Biol 401:365–381. doi:10.1007/978-1-59745-520-6_19

Kling R (1999) What is social informatics and why does it matter? D-Lib Magazine 5(1)

Lee FSL, Vogel D, Limayem M (2003) Virtual community informatics: a review and research agenda. J Inf Technol Theory Appl 5(1):47–61

Olsen GM, Zimmerman A, Bos N (2008) Scientific collaboration on the internet. MIT Press, Cambridge, MA

Waldrop MM (2008) Science 2.0. Sci Am 298(5):68–73

Weston AD, Hood L (2004) Systems biology, proteomics, and the future of health care: toward predictive, preventative, and personalized medicine. J Proteome Res 3:179–196. doi:10.1021/pr0499693

Collaboratories

- ▶ Collaborative and Distributed Biomedical Applications

Collective Behavior

Andreas Deutsch
Center for Information Services and High Performance Computing (ZIH), Technical University Dresden, Dresden, Germany

Definition

The term collective behavior describes the macroscopic behavior observed in systems that consist of a large number of interacting components. Collective behavior can result from self-organization, that is, the emergence of patterns without external influence, and the formation of complex spatiotemporal patterns even from relatively simple interactions. Examples of collective behavior include the sorting of cells by differential adhesion, the formation of fruiting bodies by microorganisms, and organization of fish into huge swarms.

Colony-Forming Fibroblastic Cells

▶ Single Cell Assay, Mesenchymal Stem Cells

Combinatorial Diversity

▶ Immune Repertoire Diversity

Combinatorial Molecular Phenotypes (CMPs)

Reyk Hillert
Molecular Pattern Recognition Research (MPRR) Group, Otto-von-Guericke-University, Magdeburg, Germany

Synonyms

CMP motif; Multi-protein cluster; Three-symbol code; TIS code

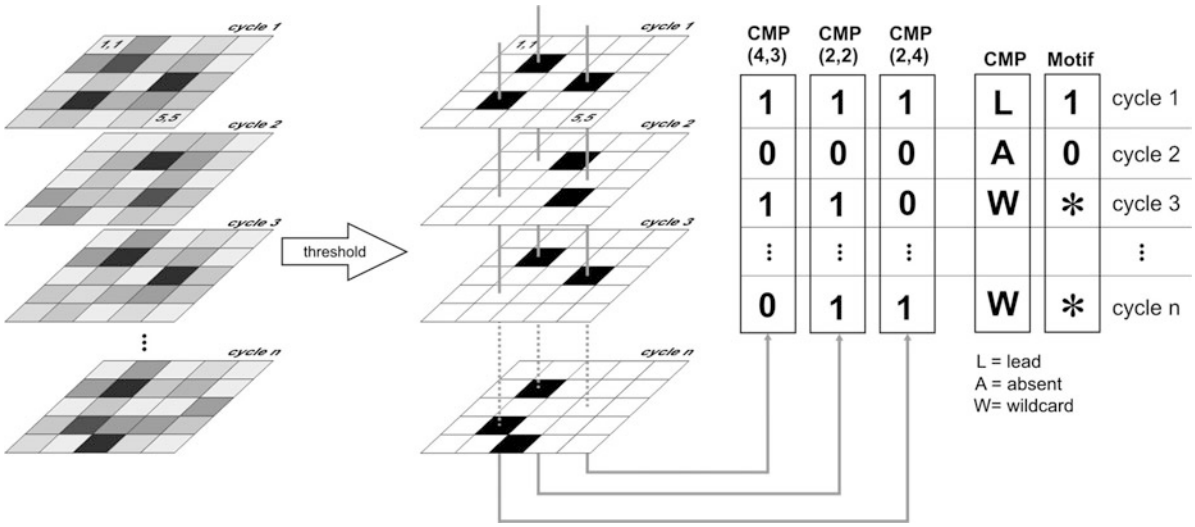
Definition

The combinatorial molecular phenotype (CMP) is calculated from the alignment of a stack of binary images originating from fluorescence signals (gray values) generated by a Toponome Imaging System (TIS) (▶ Clinical Aspects of the Toponome Imaging System (TIS)). A CMP is a bijective binary vector denoting the presence or absence (1 or 0) of any single out of N marker molecules co-mapped by TIS (▶ Clinical Aspects of the Toponome Imaging System (TIS)) at a certain pixel/voxel position or at multiple positions in the image. A CMP vector includes a certain number of pixel coordinates, at least one, which determine(s) the CMP frequency for statistical analysis. Several distinct CMPs can be grouped as a so-called CMP motif (Fig. 1). A motif describes a set of CMPs by using a three-symbol code (Schubert 2007). The symbol L (or 1) describes a marker or protein that all CMPs of the group have in common. Symbol A (or 0) is used when a certain marker is always absent (anti-colocated) in the CMPs of the group. If a marker is variably present within a CMP group, the symbol in the corresponding motif is a wild card W (or *).

A CMP at pixel position x,y is defined as binary vector as follows:

$$CMP(x,y) = (s_1, s_2, s_3, \ldots, s_i)^T_{x,y}$$
$$(s_i \text{ is a binary signal}, s \in \{0,1\};\ i = \{1,2,3,\ \ldots, N\}$$
$$|N = \text{number of markers};\ x,y = \text{pixel coordinates})$$

CMPs and resulting CMP motifs with their lead protein(s) are direct quantitative measures of the hierarchy of proteins interlocked as protein cluster networks. Lead proteins appear to exert control over the topology and function of protein cluster networks (Schubert et al. 2006; Friedenberger et al. 2007; Schubert et al. 2011; Schubert 2010). Hence, their detection by hierarchical toponome/CMP analysis is a new way in medical systems biology and drug target discovery (Schubert et al. 2008).



Combinatorial Molecular Phenotypes (CMPs), Fig. 1 The figure shows the stepwise generation of binary images, as well as the calculation of CMP vectors and the three-symbol code of the resulting CMP motif

References

Friedenberger M, Bode M, Krusche A, Schubert W (2007) Fluorescence detection of protein clusters in individual cells and tissue sections by using toponome imaging system: sample preparation and measuring procedures. Nat Protoc 2(9):2285–2294

Schubert W (2007) A three-symbol code for organized proteomes based on cyclical imaging of protein locations. Cytometry A 71(6):352–360

Schubert W (2010) On the origin of cell functions encoded in the toponome. J Biotechnol 149(4):252–259

Schubert W, Bonnekoh B, Pommer AJ, Philipsen L, Böckelmann R, Malykh Y, Gollnick H, Friedenberger M, Bode M, Dress AW (2006) Analyzing proteome topology and function by automated multidimensional fluorescence microscopy. Nat Biotechnol 24(10):1270–1278

Schubert W, Bode M, Hillert R, Krusche A, Friedenberger M (2008) Toponomics and neurotoponomics: a new way to medical systems biology. Expert Rev Proteomics 5(2):361–369

Schubert W, Gieseler A, Krusche A, Serocka P, Hillert R (2011) Next-generation biomarkers based on 100- parameter functional super-resolution microscopy TIS. N Biotechnol. doi:10.1016/j.nbt.2011.12.004

Combinatorial Regulation

► Combinatorial Transcription Regulatory Network

Combinatorial Transcription Regulatory Network

Yong Wang
Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China

Synonyms

Combinatorial regulation; Transcriptional factor cooperativity

Definition

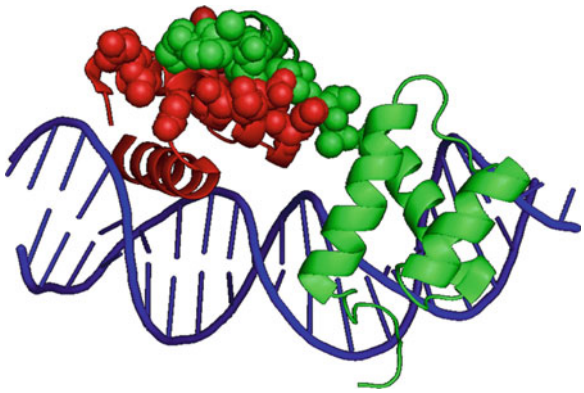
A regulatory network consists of regulators such as transcription factors or kinases that control the expression or activity of their target genes (Chen et al. 2009). Compared to more commonplace contexts, these regulators can be thought of as managers in a social or corporate setting controlling their common subordinates. Usually, these regulators are not working alone and, almost always, these multiple regulators are partnering together to control their targets. Combinatorial regulation is not a well-defined term in Biology and

we refer to the combinatorial regulation generally when the regulators (both TFs and kinases) perform their function mostly in combination with other regulators under different spatial and/or temporal conditions.

Specifically, combinatorial transcriptional regulatory network describes the combinatory regulatory relationships among transcription factors. Transcription factors (TFs) are proteins that dynamically read and interpret the static genetic instructions in the DNA. As mentioned above, transcription factors usually cooperate with other TFs to facilitate (as an activator) or inhibit (as a repressor) the recruitment of RNA polymerase, using complex logic rules building upon simple ones (AND, OR, and NOT) to control the precise condition-dependent expression of target genes. For example, one TF can help another TF to stabilize onto regulatory DNA sequence and recruit RNAP to start transcription; several TFs may use diverse and complex logic rules to control the phased temporal expression of target genes; also, it has been demonstrated by computational model that, in yeast cell-cycle modeling, the inclusion of synergistic interactions can increase the prediction accuracy of transcriptional regulation to as much as 1.5–3.5-fold.

Below, we show a picture for TF combinatorial regulation, which is the structure of the a1/α2 complex and their binding DNA fragment. The TF a1 is shown in red and the TFα2 is green and the DNA fragment is blue (Fig. 1). The interaction interface between two transcription factors is highlighted by spheres representation, which includes residues whose distance is within 7 Å.

It should be noted that the TF combinatorial regulation is not uniquely defined and can be interpreted in different ways (transcriptional combinatorial regulation, TF synergism, TF cooperation, TF interaction, and TF cooperativity). It can refer to TF–TF physical/genetic/regulatory interaction, co-occurrence of DNA-binding sites close to each other in the same promoter region, and TF spatial and temporal combinatorial co-regulation.



Combinatorial Transcription Regulatory Network, Fig. 1 TF combinatorial regulation, structure of the a1/a2 complex and their binding DNA fragment. Red: TF a1, green: TFa2, blue: DNA fragment

Characteristics

Significance of Combinatorial Regulation

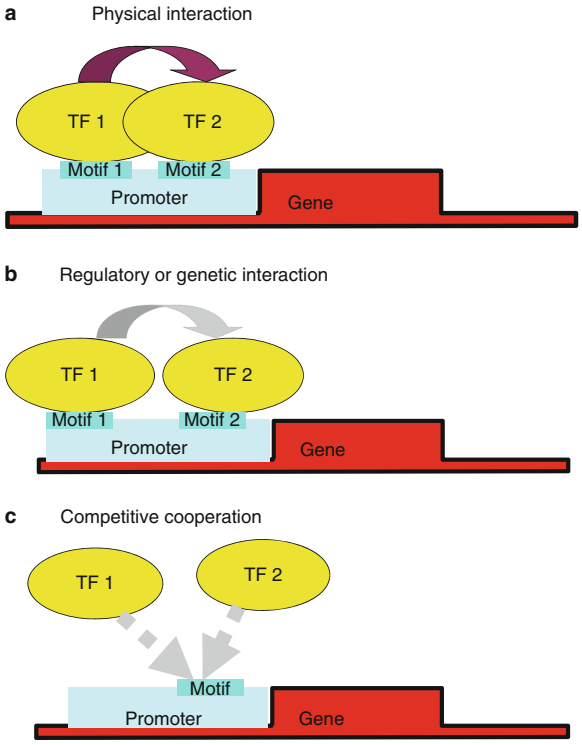
Combinatory regulation is very important since it allows for a sophisticated response to multiple conditions in the environment, integration of multiple signaling inputs, and generation of highly specific outputs with the help of a relatively small number of regulators. Overall, transcriptional cooperativity among several TFs is believed to be the main mechanism of complexity and precision in transcriptional regulatory programs.

Combinatorial regulation is a common theme in eukaryotic transcriptional circuits, as transcription factors often work in different combinations to regulate different sets of genes under different conditions. Many combinatorial interactions are due to direct protein–protein contacts between sequence-specific DNA-binding proteins. These interactions are often much weaker than the protein–DNA interactions. It is therefore not surprising that changes in the interactions between transcription factors play an important role in transcriptional rewiring.

Patterns of Combinatory Regulation

There are three different patterns of TF combinatorial regulation that suggest three possible types of TF cooperativity mechanisms in transcriptional control (Fig. 2):

In type-I cooperativity (Subfigure A), TFs cooperate through physical interactions in a transcriptional complex, and jointly regulate many target genes. Upon the formation of the complex, the binding probability of RNA polymerase onto the promoter sequence will either increase or decrease, thus affecting the



Combinatorial Transcription Regulatory Network, Fig. 2 Different patterns of TF combinatorial regulation. (a) type-I cooperativity, (b) type-II cooperativity, (c) type-III cooperativity

subsequent transcriptional efficiency. Here, protein physical interaction is the dominant predictor for this kind of cooperative relationships. Examples of this type of cooperativity include the Met4 Complex, CCAAT-binding factor complex, MBF complex, and SBF complex in yeast. Since the two TFs cooperate in a transcriptional complex and they stay together most of the time, we expect that other features are good predictors as well. For example, the co-evolution relationship is a good predictor since interacting TFs tend to co-evolve in different genomes.

In type-II cooperativity (Subfigure B), TFs cooperate largely through genetic or regulatory interaction, and jointly regulate many target genes. Different from the first case, TFs interact in a genetic or regulatory module that may or may not be explained by the existence of a physical complex. For example, one TF can regulate a secondary TF, and they jointly regulate the target genes. The well-known network motifs, such as the feed-forward loop and the regulator cascade, belong to this type. Alternatively, one TF can genetically interact with another TF which leads to a joint phenotype. Here, TF regulatory or genetic interactions are the dominant predictors for type-II cooperativity. Examples of this type of cooperativity include Sok2-Phd1, Yap6-Cup9, and Skn7-Yap1 pairs in yeast, which are all detected by ChIP-chip experiments as belonging to regulatory feed-forward motifs and supported by literature evidences.

In type-III cooperativity (Subfigure C), TFs often cooperate with each other in a competitive way (TF competitive regulation), or in a redundant manner. In this case, the DNA-binding sites of these TFs share high sequence identity or similarity. Their binding motifs often overlap with each other, leading to competition between TFs for transcriptional control. In this case, the motif-occurrence data-based features are the dominant predictors. Some representative cases of type-III cooperativity are Pdr1-Pdr3, Cat8-Sip4, and Msn2-Msn4 in yeast.

Methods to Detect or Predict TF Combinatorial Regulations

Experimental methods for detecting TF interaction include co-immunoprecipitation and super-gel shift. One difficulty for these methods is that the TF interaction is usually transient and hard to be captured. In addition, these methods are generally time consuming, and it is difficult to apply them to mapping the whole-genome TF cooperativity network in the living cell.

Complementarily, a wide variety of computational approaches have been proposed to predict TF cooperativity. Intuitive idea is that combinatorial regulation and higher-order regulatory logic can be revealed by examining overlaps of target genes and binding sites between transcription modules. Such quantitative modeling may provide insight into underlying mechanisms and design principles, which can be tested by experiment.

In addition to some case studies of combinatorial regulation (Wang et al. 2009; Parisi et al. 2007), existing methods can be roughly classified into three catalogs. The first class is the binding motif–based methods. For example, Wagner (1999) employed a statistical technique to identity significant homotypic or heterotypic TF-binding site clusters. (Pilpel et al. 2001) used TF-binding motifs with gene expression data to look for cooperatively binding TFs.

(Hannenhalli and Levy 2002) predicted TF synergism by the co-localization of transcriptional factor–binding site (TFBS) on the genome at the specific distance apart. (Das et al. 2004) created a multivariate adaptive regression splines model to correlate the occurrences and interactions of TFs to the logarithm of the ratio of gene expression levels. (Yu et al. 2006) identified interacting binding motif pairs considering their orientation and distance. Similarly, (Chang et al. 2006) used promoter sequences and gene expression data to detect cooperativity. Recently, (Datta and Zhao 2008) used a log-linear model to infer cooperative binding. The second class is the target gene (TG)-based methods. (Banerjee and Zhang 2003) assumed that a TF pair is cooperative if their TGs are significantly co-expressed while (Nagamine et al. 2005) considered TG interactions instead of TG co-expression. (Tsai et al. 2005) directly selected TF pairs with significant number of common TGs and further applied the ANOVA test to determine synergy. (Balaji et al. 2006) applied a network transformation procedure to ChIP-chip data and analyzed combinatorial regulation among multiple TFs. The third class is the transcription factor activity (TFA)-based methods. (Yang et al. 2005) inferred the TFAs from gene expression data then assessed TF cooperativity by their TFA correlation. Wang (2007) used the similar framework to study combinatorial regulation in yeast cell cycle.

In addition to the above case studies and the unsupervised framework using information from a single data source such as TF-binding motif, target gene, and TF activity, Wang (2009) used Bayesian networks, a supervised learning approach, to systematically integrate diverse data sources at different levels to predict and analyze TF cooperativity. Specifically, they design a Bayesian network structure to capture the dominant correlations among features and TF cooperativity, and introduce a supervised learning framework with a well-constructed gold standard dataset. This framework allows us to assess the predictive power of each genomic feature, validate the superior performance of our Bayesian network compared to alternative methods, and integrate genomic features. Data integration reveals 159 high-confidence predicted cooperative relationships among 105 TFs, most of which are subsequently validated by literature search. The existing and predicted transcriptional cooperativities can be further grouped into three categories based on the combination patterns of the genomic features, which provide further biological insights into the different types of TF cooperativity.

Cross-References

- Gene Regulatory Networks
- Regulation

References

Balaji S, Babu MM, Iyer LM, Luscombe NM, Aravind L (2006) Comprehensive analysis of combinatorial regulation using the transcriptional regulatory network of yeast. J Mol Biol 360:213–227

Banerjee N, Zhang MQ (2003) Identifying cooperativity among transcription factors controlling the cell cycle in yeast. Nucleic Acids Res 31:7024–7031

Chang Y-H, Wang Y-C, Chen B-S (2006) Identification of transcription factor cooperativity via stochastic system model. Bioinformatics 22:2276–2282

Chen L, Wang RS, Zhang XS (2009) Biomolecular networks: methods and applications in systems biology. Wiley, Hoboken

Das D, Banerjee N, Zhang MQ (2004) Interacting models of cooperative gene regulation. Proc Natl Acad Sci 101:16234–16239

Datta D, Zhao H (2008) Statistical methods to infer cooperative binding among transcription factors in Saccharomyces cerevisiae. Bioinformatics 24:545–552

Hannenhalli S, Levy S (2002) Predicting transcription factor synergism. Nucleic Acids Res 30:4278–4284

Nagamine N, Kawada Y, Sakakibara Y (2005) Identifying cooperative transcriptional regulations using protein-protein interactions. Nucleic Acids Res 33:4828–4837

Parisi F, Wirapati P, Naef F (2007) Identifying synergistic regulation involving c-Myc and sp1 in human tissues. Nucleic Acids Res 35:1098–1107

Pilpel Y, Sudarsanam P, Church GM (2001) Identifying regulatory networks by combinatorial analysis of promoter elements. Nat Genet 29:153–159

Tsai H-K, Lu HH-S, Li W-H (2005) Statistical methods for identifying yeast cell cycle transcription factors. Proc Natl Acad Sci 102:13532–13537

Wang J (2007) A new framework for identifying combinatorial regulation of transcription factors: a case study of the yeast cell cycle. J Biomed Inform 40:707–725

Wang Y, Zhang X-S, Chen L (2009) Predicting eukaryotic transcriptional cooperativity by Bayesian network integration of genome-wide data. Nucleic Acids Res 37(18):5943–5958

Yu X, Lin J, Masuda T, Esumi N, Zack DJ, Qian J (2006) Genome-wide prediction and characterization of interactions between transcription factors in Saccharomyces cerevisiae. Nucleic Acids Res 34:917–927

Yang Y-L, Suen J, Brynildsen M, Galbraith S, Liao J (2005) Inferring yeast cell cycle regulators and interactions using transcription factor activities. BMC Genomics 6:90

Combinatorics of Sets

▶ Subset Surprisology and Toponomics

Committee

▶ Ensemble

Common Cold

▶ Viral Respiratory Tract Infections

Communication Theory

▶ Information Theory and Toponomics

Community

▶ Dynamic Modularity
▶ Module Network

Community Database

Sabina Leonelli
ESRC Centre for Genomics in Society, University of Exeter, Exeter, Devon, UK

Synonyms

Model organism database

Definition

Database collecting diverse types of data documenting the same (set of) organisms, with the purpose of (1) facilitating integration across datasets and disciplinary approaches to the study of the same organism and (2) facilitating cross-species comparison across available datasets (Howe and Rhee 2008; Leonelli and Ankeny 2012). Examples of community databases are The Arabidopsis Information Resource, gathering data about *Arabidopsis thaliana* (Koornneef and Meinke 2010); and FlyBase, gathering data about *Drosophila melanogaster* (Drysdale and FlyBase Consortium 2008).

Cross-References

▶ Data-Intensive Research

References

Drysdale R, FlyBase Consortium (2008) FlyBase: a database for the Drosophila research community. Methods Mol Biol 420:45–59
Howe D, Rhee SY (2008) The future of biocuration. Nature 455:47–50
Koornneef M, Meinke D (2010) The development of Arabidopsis as a model plant. Plant Journal 61(6):909–921
Leonelli S, Ankeny RA (2012) Re-thinking organisms: the epistemic impact of databases on model organism biology. Stud Hist Philos Biol Biomed Sci 43(1):29–36

Community Detection

▶ Modules, Identification Methods and Biological Function

Community Genome

▶ Metagenome

Community Genomics

▶ Metagenomics

Community Metabolomics

▶ Metametabolomics

Community Proteomics

▶ Metaproteomics

Community Structure

▶ Modular Organization of Gene Regulatory Networks

Community Transcriptomics

▶ Metatranscriptomics

Compact Chromatin

▶ Heterochromatin

Compact DNA

▶ Chromatin

Comparative Analysis of Molecular Networks

Shihua Zhang[1] and Zhenping Li[2]
[1]National Center for Mathematics and Interdisciplinary Sciences, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China
[2]School of Information, Beijing Wuzi University, Beijing, China

Synonyms

Network alignment; Network querying

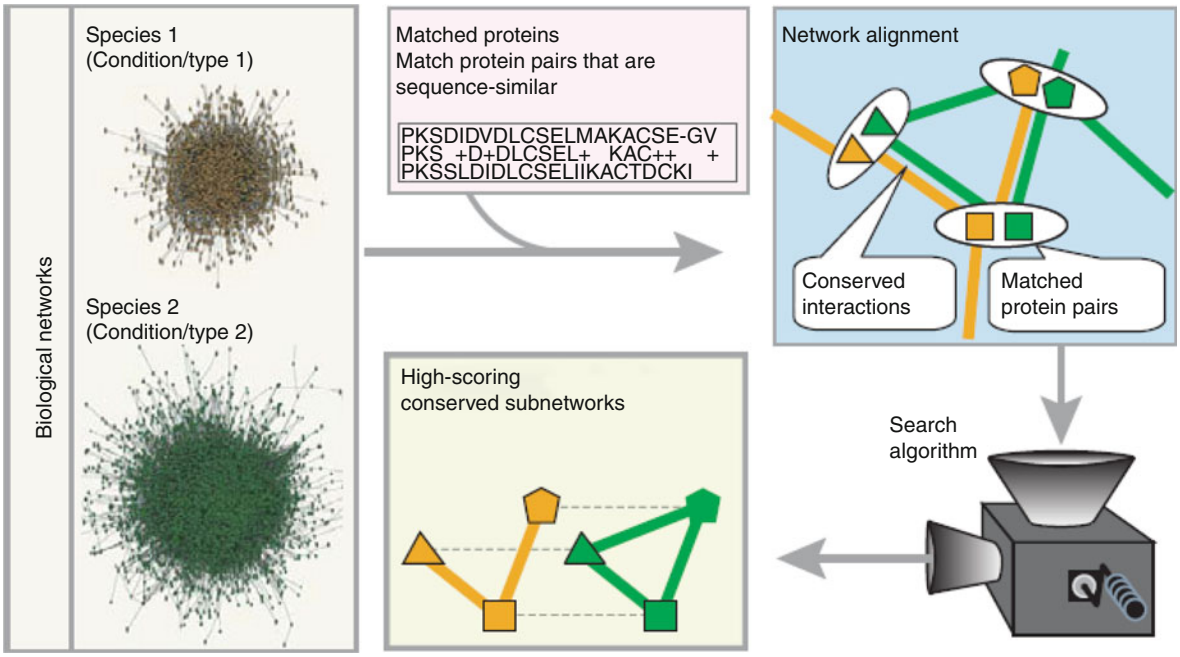
Definition

A fundamental goal of biology is to understand the cell as a system of interacting components which can be represented as a network (Chen et al. 2009). The discovery, analysis, and understanding of interactions between molecules as well as the networked system have received significant attentions in recent years. In a network, each node corresponds to a molecule and an edge indicates a direct/indirect interaction between the molecules. As the number of available molecular networks for various species rapidly increases, comparative analysis of them across species is proving to be increasingly important. The comparative study of molecular networks can be topologically coarse-level comparison such as the comparative analysis of degree distribution, clustering coefficient, diameter, and relative graphlet frequency distribution, or can be the discovery of conserved subgraphs among networks. Actually, the later one which is known as network alignment problem in bioinformatics field is a well-known concept for molecular network studies and will be the key of this entry. We should note here that the problems similar to network alignment have also been studied in other fields like graph theory, data mining, and computer vision. For example, the problem of matching a query image to an existing image in the database has been formulated as a graph-matching problem in computer vision field, where each image is represented as a graph/network.

Characteristics

The network comparative analysis especially network alignment is similar in spirit to traditional sequence-based comparative genomic analyses or structure-based comparative analysis. It also offers a function-oriented perspective that complements traditional sequence-based and structure-based methods. Comparative network analysis also enables us to uncover the network (dis)similarity, identify conserved functional modules across species, perform accurate ortholog prediction and function prediction as well as transfer insights and information across species.

In general, the goal of network alignment problem is to find a common and approximative subgraph with a set of conserved edges across the input networks (Fig. 1). There exists a mapping between the nodes of



Comparative Analysis of Molecular Networks, Fig. 1 A general framework for pairwise network alignment, with permission from Sharan and Ideker (2006)

the subgraph which may correspond to the sequence similarity or function similarity of molecules (Ogata et al. 2000). The goal is to maximize the overlap between the aligned networks and ensure that the proteins mapped to each other are evolutionarily or functionally related. Moreover, according to the number of input networks, the network alignment problem can be formulated as pairwise or multiple alignments. While according to the scope of node mapping desired, the alignment can be local or global which is analogous to sequence alignment or structure alignment problem. We may also want to align a known or even unknown small network to a large network or network database to find the conserved ones which is named as the network querying problem.

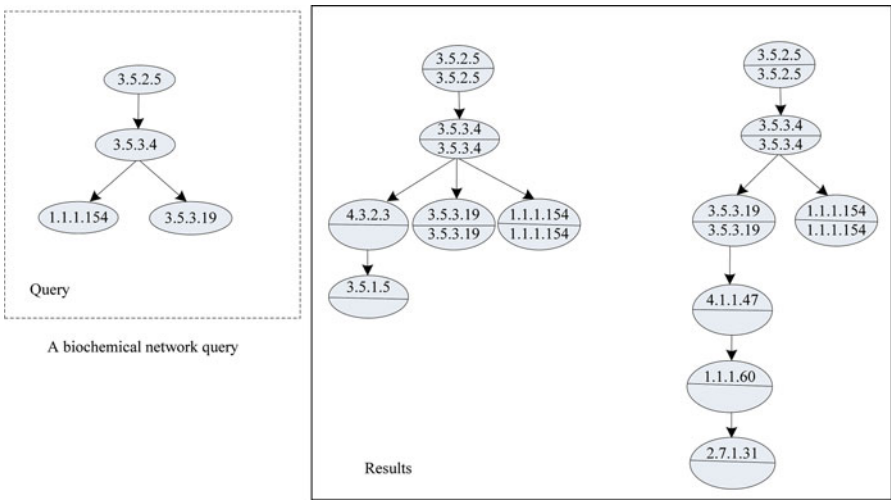
The majority of methods used for alignment of molecular networks have focused on local alignments (Berg and Lässig 2004, 2006; Kelley et al. 2004; Flannick et al. 2006; Liang et al. 2006) which attempts to find local region similarity and find node mappings independently for different regions. There have been many algorithms for local alignment in the literature. The PathBLAST tool searches for high-scoring linear or tree substructures between two large networks, by taking into account both the sequence homology between the aligned molecules and the probabilities that interactions in the subnetworks are true or not (Kelley et al. 2004), while the NetworkBLAST tool detects conserved protein clusters rather than only paths across pairwise or multiple networks, by deploying a likelihood-based scoring scheme that weighs the denseness of a subnetwork versus the chance of observing such subnetwork at random (Sharan et al. 2005). By using this tool to compare multiple networks across different species, Suthram et al. (2005) explored whether the divergence of *Plasmodium* at the sequence level can be embodied at the level of the structure of its protein interaction network. They found that *Plasmodium* has only three conserved complexes versus yeast, and no conserved complexes against fly, worm, and bacteria. But yeast, fly, and worm share an abundance of conserved complexes with each other. In another method MaWISh, the authors formulated network alignment as a maximum weight induced subgraph problem and implemented an evolution-based scoring scheme to discover conserved modules. The method extends the concepts of evolutionary events in sequence alignments to that of

duplication, match, and mismatch in network alignments and evaluates the similarity between network structures through a scoring function that accounts for these evolutionary events (Koyutürk et al. 2005), while the Graemlin is the first method which can identify dense conserved subnetworks of arbitrary structure. This method scores a module by computing the log-ratio of the probability that it is subject to evolutionary constraints and the probability that it is under no constraints, while taking into account phylogenetic relationships between species whose networks are being aligned (Flannick et al. 2006). Another example is MNAligner (Li et al. 2007), which is designed for general molecular networks and combines both molecular similarity and topological similarity. This method can detect conserved subnetworks in an efficient mathematical programming manner without requiring special structures on the networks.

Local alignments can be ambiguous, in which one node can have different counterparts in different conserved modules. By contrast, a global network alignment provides a unique alignment for each node in the smaller network to exactly one node in the larger network. We have to note that this may lead to suboptimal alignments in some local regions. Generally, the local network alignment algorithms have not been able to identify large subnetworks that have been conserved during evolution due to the algorithmic complexity or modeling aspect (Berg and Lässig 2004; Kelley et al. 2004). Recently, global network alignment methods have been studied and designed for aligning molecular networks (Singh et al. 2007; Flannick et al. 2008; Zaslavskiy et al. 2009). Unlike the local alignment algorithms, the IsoRank method (Singh et al. 2008) aims to maximize the overall match between the two networks. The method is based on spectral graph theory which computes scores of aligning pairs of nodes according to the score of their respective neighbors from different networks. In other words, the score of a protein pair depends on the score of their neighbors. And then sequence scores are incorporated into these "topological" scores in the pairwise alignment scores. Finally, IsoRank constructs the node alignment with the repetitive greedy strategy of identifying among all protein pairs the highest scoring pair (Singh et al. 2008). Recently, the IsoRankN has extended based on the notion of node-specific rankings with other ranking algorithms (Liao et al. 2009). The Graemlin method has also been extended to allow global network alignment which is based on a learning algorithm that uses a training set of known network alignments and their phylogenetic relationships to learn parameters for its scoring function, and by automatically adapting the learned objective function to any set of networks (Flannick et al. 2008).

Another aspect of network alignment studies is alignment for multiple networks which has gained great progress recently. Several of above methods have also been developed for multiple cases. For example, Graemlin developed by Flannick et al. (2006), which uses a probabilistic function for topology matching, and can be applied to search for conserved functional modules among multiple protein interaction networks. The C3Part-M method (Deniélou et al. 2009) addressed the problem of multiple network alignment with an exact and generic approach by avoiding the explicit construction of the network alignment multigraph, and then can deal with a large number of networks. By using microarray data from multiple conditions and species, various comparative studies have been conducted so as to reveal transcriptional regulatory modules, predict gene functions, and uncover evolutionary mechanisms (Zhou and Gibson 2004). For example, Yan et al. (2007) have developed a graph-based data-mining algorithm NeMo to detect frequent co-expression modules among a large number of gene co-expression networks across various conditions. They found a large number of potential transcriptional modules, which are activated under multiple conditions.

In addition to the network alignment method discussed above, the network querying technique is becoming a new major network comparative analysis tool. The goal of network querying is to find a small network against a large-scale network or a database of large-scale networks which is a NP-hard problem (Fig. 2). The network querying problem has been studied in the past few years and a few search tools have been developed, such as PathBLAST (Kelley et al. 2003, 2004), QPath (Shlomi et al. 2006), TORQUE (Bruckner et al. 2009), and QNet (Dost et al. 2007). PathBLAST (Kelley et al. 2003, 2004) developed by Kelley et al. can query a small pathways with length no



Comparative Analysis of Molecular Networks, Fig. 2 Illustration of metabolic pathways querying in a pathway database, redrawn from (Pinter et al. 2005)

more than 5 nodes. MetaPathwayHunter (Pinter et al. 2005) developed by Pinter et al. can fast query for smaller pathways or trees. QPath (Shlomi et al. 2006) developed by Shlomi et al. can search for linear pathways. NetMatch (Ferro et al. 2007) based on a graph-matching algorithm can find the subgraphs of the original graph connected in the same way as the querying graph. The results of NetMatch can be viewed as candidate network motifs as a result of their similar topological features (Alon 2007). Besides the network querying tools discussed above, many querying tools, such as BLAST for sequence querying and DALI for structure querying, have been developed by researchers in other areas of computational biology, and have had a tremendous impact on the development of biological science (Zhang et al. 2008).

To benefit from the accumulation of networked data for different species, it will be important to develop user-friendly network comparison tool. Recent advances in the field demonstrate the great potential of network comparison in elucidating network organization, function, and evolution. Advances in computational methods and powerful software tools are being made by interdisciplinary cooperation across different fields. With the development of powerful and sophisticated network comparison tools, we expect to gain deep insight into essential mechanisms of biological systems at the network level.

Cross-References

- Global Network Alignment
- Local Network Alignment
- Network Alignment
- Network Querying
- Networks Comparison

References

Alon U (2007) Network motifs: theory and experimental approaches. Nat Rev Genet 8:450–461

Berg J, Lässig M (2004) Local graph alignment and motif search in biological networks. Proc Natl Acad Sci USA 101:14689–14694

Berg J, Lässig M (2006) Cross-species analysis of biological networks by bayesian alignment. Proc Natl Acad Sci USA 103:10967–10972

Bruckner S, Huffner F, Karp RM, Shamir R, Sharan R (2009) Topology-free querying of protein interaction networks. Nucleic Acids Res 37(suppl 2):W106–W108

Chen L, Wang RS, Zhang XS (2009) Biomolecular networks: methods and applications in systems biology. Wiley, Hoboken

Deniélou Y, Boyer F, Viari A, Sagot M (2009) Multiple alignment of biological networks: a flexible approach. Combinatorial pattern matching. LNCS 5577:263–273

Dost B, Shlomi T, Gupta N, Ruppin E, Bafna V, Sharan R (2007) QNet: a tool for querying protein interaction networks. Research in computational molecular biology. LNCS 4453:1–15

Ferro A, Giugno R, Pigola1 G, Pulvirenti A, Skripin D, Bader GD, Shasha D (2007) NetMatch: a cytoscape plugin for searching biological networks. Bioinformatics 23:910–912
Flannick J et al (2006) Graemlin: general and robust alignment of multiple large interaction networks. Genome Res 16(9):1169–1181
Flannick J et al (2008) Automatic parameter learning for multiple network alignment. RECOMB LNBI 4955:214–231
Kelley BP et al (2003) Conserved pathways within bacteria and yeast as revealed by global protein network alignment. Proc Natl Acad Sci USA 100:11394–11399
Kelley BP, Yuan B, Lewitter F, Sharan R, Stockwell BR, Ideker T (2004) PathBLAST: a tool for alignment of protein interaction networks. Nucleic Acids Res 32 (Web Server issue):W83–W88
Koyutürk M, Grama A, Szpankowski W (2005) Pairwise local alignment of protein interaction network guided by models of evolution. RECOMB LNBI 3500:48–65
Li Z, Zhang S, Wang Y, Zhang X, Chen L (2007) Alignment of molecular networks by integer quadratic programming. Bioinformatics 24(4):594–596
Liang Z, Xu M, Teng M, Niu L (2006) NetAlign: a web-based tool for comparison of protein interaction networks. Bioinformatics 22:2175–2177
Liao CS, Lu K, Baym M, Singh R, Berger B (2009) IsoRankN: spectral methods for global alignment of multiple protein networks. Bioinformatics 25:i253–i258
Ogata H, Fujibuchi W, Goto S, Kanehisa M (2000) A heuristic graph comparison algorithm and its application to detect functionally related enzyme clusters. Nucleic Acids Res 28:4021–4028
Pinter RY, Rokhlenko O, Yeger-Lotem E, Ziv-Ukelson M (2005) Alignment of metabolic pathways. Bioinformatics 21:3401–3408
Sharan R, Ideker T (2006) Modeling cellular machinery through biological network comparision. Nature Biotechnol 24:427–433
Sharan R, Suthram S, Kelley RM, Kuhn T, McCuine S, Uetz P, Sittler T, Karp RM, Ideker T (2005) Conserved patterns of protein interaction in multiple species. Proc Natl Acad Sci USA 102:1974–1979
Shlomi T, Segal D, Ruppin E, Sharan R (2006) QPath: a method for querying pathways in a protein-protein interaction network. BMC Bioinformatics 7:199
Singh R, Xu J, Berger B (2008) Global alignment of multiple protein interaction networks with application to functional orthology detection. PNAS 105(35):12763–12768
Suthram S, Sittler T, Ideker T (2005) The plasmodium protein network diverges from those of other eukaryotes. Nature 438:108–112
Yan X, Mehan M, Huang Y, Waterman MS, Yu PS, Zhou XJ (2007) A graph based approach to systematically reconstruct human transcriptional regulatory modules. Bioinformatics 23(13):i577–i586
Zaslavskiy M, Bach F, Vert JP (2009) Global alignment of protein-protein interaction networks by graph matching methods. Bioinformatics 25(12):i259–i267
Zhang S, Zhang XS, Chen L (2008) Biomolecular network querying: a promising approach in systems biology. BMC Syst Biol 2:5, Jan 18
Zhou XJ, Gibson G (2004) Cross-species comparison of genome-wide expression patterns. Genome Biol 5(7):232

Competitive Repopulation

▶ Single Cell Assay, Hematopoietic Stem Cell

Compiler

José Román Bilbao Castro
Supercomputing-Algorithms Group, University of Almería, Almería, Spain

Definition

A compiler is a computer program that translates a programmer-written code into something a processor can "understand." The programmer writes what is called a "source code" in a specific language. Such a language is easier to understand to the programmer than a "machine language" composed of 1s and 0s. The compiler then facilitates programming by allowing a much higher level of abstraction. Compiler development is a very important field of study as compiled programs' performances will strongly depend on the compiler ability to interpret the high level code and optimize it for the underlying machine it will be executed on.

Cross-References

▶ Multicore Computing

Complementarity Determining Region

▶ Complementarity Determining Region (CDR-IMGT)

Complementarity Determining Region (CDR-IMGT)

Marie-Paule Lefranc
Laboratoire d'ImmunoGénétique Moléculaire, Institut de Génétique Humaine UPR 1142, Université Montpellier 2, Montpellier, France

Synonyms

CDR-IMGT; Complementarity determining region

Definition

A Complementarity determining region (CDR-IMGT) is a loop region of a V-DOMAIN (▶ Variable (V) Domain), delimited according to the IMGT unique numbering for V domain (Lefranc et al. 2003). There are three CDR-IMGT in a V-DOMAIN: CDR1-IMGT (loop BC), CDR2-IMGT (loop C′C″), and CDR3-IMGT (loop FG).

In a V-DOMAIN (V domain of the immunoglobulins (IG) or antibodies and T cell receptors (TR)), the amino acids of the CDR-IMGT bind to an antigen (▶ Epitope) and confer the specificity to the IG and TR (▶ Paratope, ▶ IMGT-ONTOLOGY, SpecificityType). The first two CDR-IMGT are part of the V-REGION (encoded by a ▶ variable (V) gene), whereas the CDR3-IMGT corresponds to the junction and results from the rearrangement between a V gene and a ▶ joining (J) gene (V-J rearrangement) or between a V gene, a ▶ diversity (D) gene, and a J gene (V-D-J rearrangement) (▶ Immunoglobulin Synthesis). The two anchors, 2nd-CYS 104 and J-TRP or J-PHE 118 (▶ Framework region (FR-IMGT)), are part of the JUNCTION but do not belong to the CDR3-IMGT.

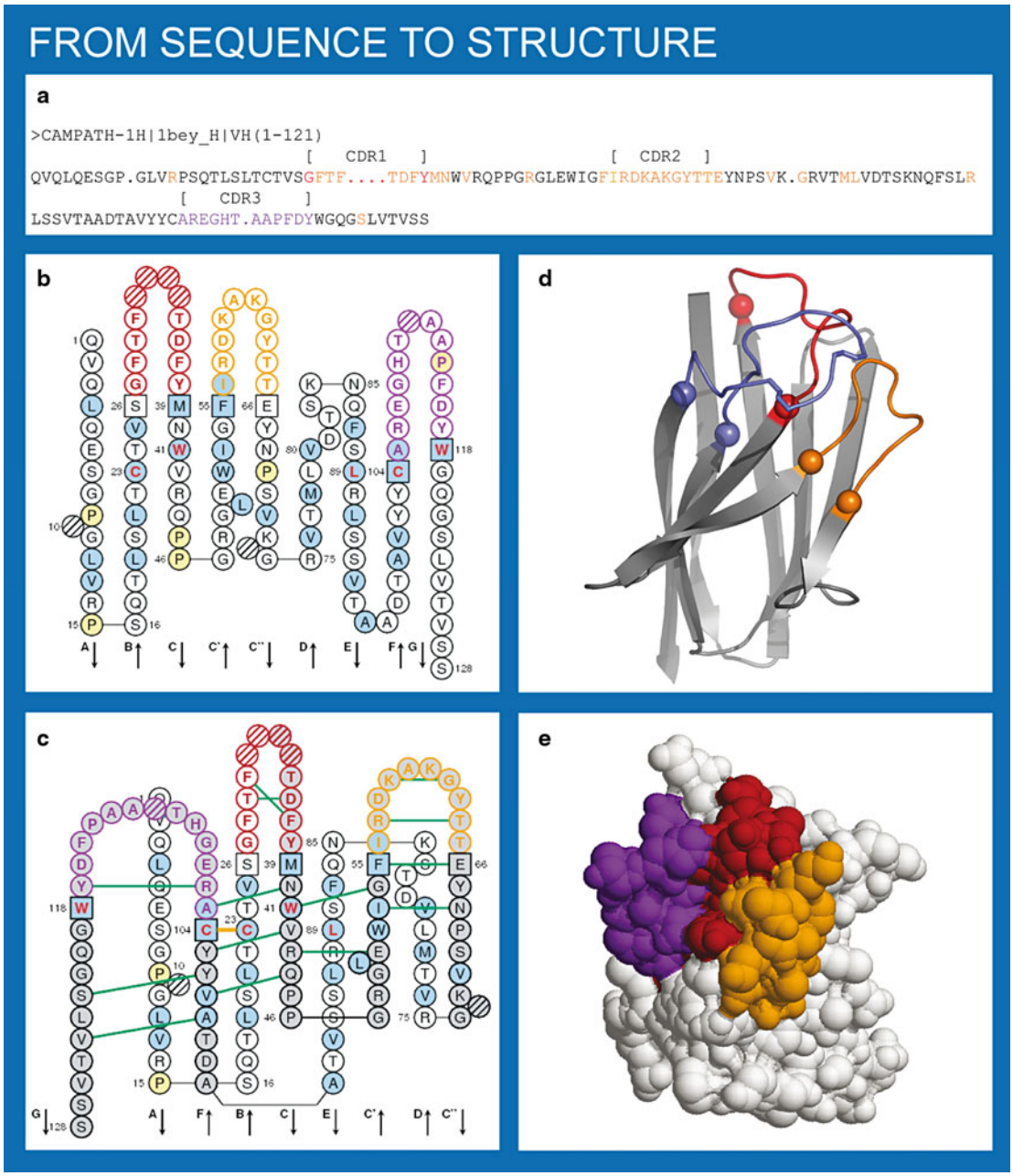
Analysis of the JUNCTION and of the CDR3-IMGT of V-DOMAIN nucleotide sequences is performed by IMGT/JunctionAnalysis, integrated in IMGT/V-QUEST, the nucleotide sequence analysis tool (Lefranc 2009; Ehrenmann et al. 2010) of IMGT®, the international ImMunoGeneTics information system® (http://www.imgt.org) (▶ IMGT® Information system).

In a V-LIKE-DOMAIN (V domain of ▶ Immunoglobulin Superfamily (IgSF) other than IG and TR), the three loops BC, C'C" and FG correspond, structurally, to the three CDR-IMGT of a V-DOMAIN, and are delimited by anchors at the same positions, 26 and 39, 55 and 66, 104 and 118, respectively.

Amino acid positions of the CDR-IMGT of a V-DOMAIN have always the same number according to the ▶ IMGT unique numbering for V domain (Lefranc et al. 2003). This allows to define in an ▶ IMGT Collier de Perles the lengths of the CDR-IMGT. These characteristics, based on the ▶ IMGT-ONTOLOGY concepts, are managed in the ▶ IMGT® information system databases and tools. Starting from amino acid sequences, the CDR-IMGT lengths are obtained using the IMGT/DomainGapAlign tool (http:www.imgt.org) (Lefranc 2009; Ehrenmann et al. 2010).

The lengths of the three CDR-IMGT lengths of a V-DOMAIN are shown between brackets and separated with dots. For example, the CDR-IMGT lengths of the V-DOMAIN displayed in Fig. 1 are [8.10.12]. The CDR-IMGT lengths of a basic V domain without gaps are [12.10.13]. For CDR3-IMGT with more than 13 amino acids, additional positions are added at the top pf the loop (Lefranc et al. 2003).

The CDR-IMGT lengths of therapeutic antibodies are required for the World Health Organization/International Nonproprietary Names (WHO/INN) and are included in the INN definitions (Lefranc 2011). The standardized delimitation of the CDR-IMGT has a crucial importance in antibody engineering and antibody humanization by CDR grafting: it allows to precisely delimit the CDR amino acids from the original antibody (murine, rat, etc.) that need to be grafted on an human antibody framework, in order to preserve the specificity of the original antibody in the therapeutic monoclonal antibody, while decreasing its immunogenicity (Lefranc 2009, 2011; Ehrenmann et al. 2010).



Complementarity Determining Region (CDR-IMGT), Fig. 1 Complementarity determining regions (CDR-IMGT) of a V-DOMAIN. The CDR-IMGT correspond to three loops in IMGT Colliers de Perles (on one layer, and on two layers with hydrogen bonds) and in three-dimensional (3D) structures. (**a**) Amino acid sequence (http://www.imgt.org) of an antibody V-DOMAIN (VH of the IG-Heavy chain) with delimitation of the three CDR-IMGT. (**b**) IMGT Collier de Perles on one layer. (**c**) IMGT Collier de Perles on two layers. Hydrogen bonds between the amino acids of the C, C′, C″, and F and G strands and those of the CDR-IMGT are shown. (**d**) Ribbon 3D representation. (**e**) Spacefill 3D representation. The CDR1-IMGT, CDR2-IMGT and CDR3-IMGT regions are colored in red, orange and purple, respectively (IMGT Color menu). The CDR-IMGT lengths are [8.10.12]. The antiparallel beta strands (A to G) correspond to the ▶ Framework Region (FR-IMGT). Anchors positions, shown as squares in B and C (26 and 39, 55 and 66, 104 and 118), and as spheres in D, belong to the FR-IMGT. Hydrophobic amino acids (hydropathy index with positive value) and tryptophan (W) found at a given position in more than 50% of analysed IG and TR sequences are shown in blue in B and C

Cross-References

- Diversity (D) Gene
- Epitope
- Framework Region (FR-IMGT)
- IMGT Collier de Perles
- IMGT Unique Numbering
- IMGT® Information System
- IMGT-ONTOLOGY
- Immunoglobulin Superfamily (IgSF)
- Immunoglobulin Synthesis
- Joining (J) Gene
- Paratope
- IMGT-ONTOLOGY, SpecificityType
- Variable (V) Domain
- Variable (V) Gene

References

Ehrenmann F, Duroux P, Giudicelli V, Lefranc M-P (2010) Standardized sequence and structure analysis of antibody using IMGT®. In: Kontermann R, Dübel S (eds) Antibody engineering, vol 2, 2nd edn. Springer-Verlag Berlin, Heidelberg, pp 11–31, chapter 2

Lefranc M-P (2009) Antibody database and tools: the IMGT® experience. In: Zhiqiang An (ed) Therapeutic monoclonal antibodies: from bench to clinic. Wiley, Hoboken, New Jersey, pp 91–114, chapter 4

Lefranc M-P (2011) Antibody nomenclature: from IMGT-ONTOLOGY to INN definition Mabs 3:1–2

Lefranc M-P, Pommié C, Ruiz M, Giudicelli V, Foulquier E, Truong L, Thouvenin-Contet V, Lefranc G (2003) IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. Dev Comp Immunol 27:55–77

Complex Behavior

Niall Palfreyman
Biotechnology and Bioinformatics, Weihenstephan-Triesdorf University of Applied Science, Freising, Germany

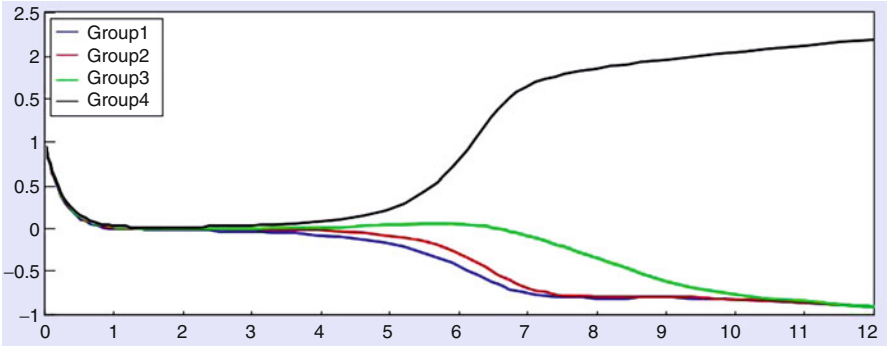
Definition

Since the 1990s, the term *complexity* has played an increasingly prominent role in our attempts to understand living systems as being somehow both lawful and yet also unpredictable. Yet in her book of the same name, Mitchell (2009) notes the current lack of a consensual definition of what precisely constitutes a complex system. She cites definitions of complexity as the *algorithmic information content* of a system (i.e., the minimum coding length of an algorithm capable of reconstructing the system), as the *fractal dimension* of a system, and as the level of *compositional hierarchy* possessed by a system's structure. She concludes from this potpourri of definitions that "notions of complexity [...] have many interacting dimensions and probably can't be captured by a single measurement scale."

These definitions reflect a marked ambiguity in what precisely we mean when we describe life as being complex. Definitions in terms of algorithmic information content regard complexity as *complicatedness*: Humans are more complex than bacteria because they possess a more complicated behavioral repertoire. On the other hand, definitions in terms of compositional hierarchy tackle a more elusive quality of living systems as *organized*: Humans are more complex than bacteria because their wide behavioral repertoire is organized around overarching purposes.

A synthesis of these quite differing criteria has emerged from Kauffman's (1993) analysis of the behavior of a family of dynamical systems called *NK Boolean networks*. When the connectivity K of a Boolean network is below a certain threshold, the network exhibits *ordered* behavior, converging to a temporally stable, globally coordinated attractor state, while at higher values of K, the behavior becomes *chaotic*, veering unpredictably between topologically uncoordinated states. At intermediate values of K the network enters the *complex* regime, in which the network divides into connected regions of comparatively stable, topologically coordinated states, yet intermittently redistributing these coordinated states and regions.

Such complex behavior is certainly visually reminiscent of living systems – it is neither random nor predictably ordered, but is characterized by *dynamic metastability*. A dynamical system is metastable with respect to some parameter λ if its dynamics remain relatively stable over large variational regions of λ, but bifurcate abruptly across specific critical values of λ.



Complex Behavior, Fig. 1 Complex speciation of birds under a symmetrical evolutionary rule

According to this account, complex behavior is therefore behavior which is metastable with respect to one or more parameters which may vary either spontaneously or in response to environmental inputs. To put it differently: Complex behavior is the ability to switch between different, reliably reproducible, varieties of behavior in response to varying environmental conditions.

For example, the orbiting of the Moon round the Earth is not complex – it is too predictable. The motion under gravity of three stars of equal mass about each other is also not complex, being in general chaotic and completely unpredictable. By contrast, the self-organization exhibited by Golubitsky and Stewart (2002, pp. 3–6: see Fig. 1) model of sympatric speciation is complex. Four distinct interacting groups of organisms evolve according to a single symmetrical rule parameterized by an exogenous parameter λ. Although initially the groups have almost identical traits, as the value of λ varies smoothly over time, this symmetry collapses, and the four trait groups self-organize into two distinct species with clearly distinguishable traits. While the final trait value of any particular group depends sensitively on initial conditions, the bifurcation into the two species is predictable.

References

Golubitsky M, Stewart I (2002) The symmetry perspective. Birkhäuser, Basel

Kauffman SA (1993) The origins of order. Oxford University Press, New York

Mitchell M (2009) Complexity: a guided tour. Oxford University Press, Oxford

Complex EGI and Enriched Complex EGI

Virginio Cantoni[1,2], Alessandro Gaggia[1] and Luca Lombardi[1]
[1]Department of Computer Engineering and Systems Science, University of Pavia, Pavia, Italy
[2]Computational Biology, KTH Royal Institute of Technology, Stockholm, Sweden

Definition

The most important advantage of the ▶ Extended Gaussian Image, the position invariance, is also its principal drawback: as a consequence, the translation of a 3D target object cannot be recovered. A solution to this problem is to represent object surfaces by their EGI adding a support function: the signed distance of the oriented tangent plane from a predefined origin. In this new representation, called Complex EGI (CEGI) (Kang and Ikeuchi 1991) (Kang and Ikeuchi 1993), the weight at each patch, representing a discretized cell, is a complex number that encodes both area and distance (Fig. 1). The magnitude of the complex number is the surface area of the object associated with that surface normal, while the phase, which supports the displacement information, is the signed distance of the surface patch from a predefined origin in the direction of its normal.

The complex weight associated with the surface patch A_i is $A_{i,nk}e^{jdk}$, where $A_{i,nk}$ is the area of patch A_i with the outward normal n_k, and d_k is the normal distance of the plane within which $A_{i,nk}$ lies to an assigned origin. For any given point in the CEGI

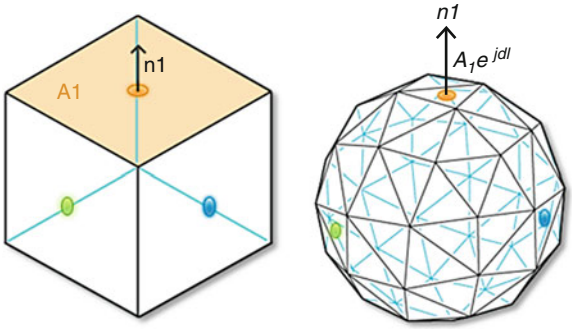
corresponding to normal n_k, the magnitude of the point's weight is $IA_{nk}e^{jdk}I$. A_{nk} is independent of the normal distance, and if the object is convex, the distribution of A_{nk} corresponds to the conventional EGI representation. Equation 1 represents the CEGI complex weight:

$$W_{n\,\vec{k}} = \sum_{l=1}^{N\,\vec{k}} A_{l,nk}\, e^{jd_{l,k}} \tag{1}$$

Also in the Enriched Complex Extended Gaussian Image (ECEGI) (Hu et al. 2010) each patch of the 3D object surface contributes with a complex weight to the associated point orientation of the Gaussian sphere. However, while the CEGI uses only a scalar complex number, the ECEGI uses a vector of three complex numbers. The resultant weight is the sum of the contributions of all surface patches having the normal in common, where the exponent is given by the distance from each one from the coordinate planes. The magnitude of the ECEGI representation is translation-invariant. The ECEGI can be viewed as three independent complex Gaussian spheres, each corresponding to the axis (x, y, z) (Fig. 2).

The weight is, in this case, represented by three complex numbers given by (2):

$$\begin{aligned} W_{x,n\,\vec{k}} &= \sum_{i=1}^{N\,\vec{k}} A_{i,nk}\, e^{jX_{i,k}} \\ W_{y,n\,\vec{k}} &= \sum_{i=1}^{N\,\vec{k}} A_{i,nk}\, e^{jY_{i,k}} \\ W_{z,n\,\vec{k}} &= \sum_{i=1}^{N\,\vec{k}} A_{i,nk}\, e^{jZ_{i,k}} \end{aligned} \tag{2}$$



Complex EGI and Enriched Complex EGI, Fig. 1 Complex extended Gaussian image

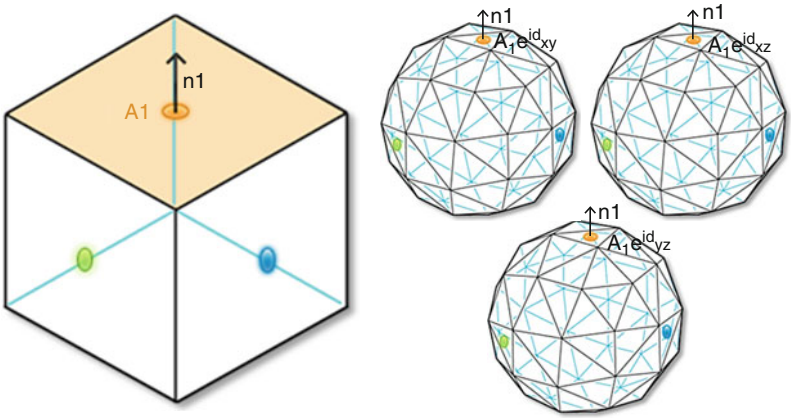
Cross-References

- Extended Gaussian Image
- Extended Gaussian Image for Pocket-Ligand Matching

References

Hu Z, Chung R, Fung KSM (2010) EC-EGI: enriched complex EGI for 3D shape registration. Machine vision and applications vol 21, pp 177–188

Kang SB, Ikeuchi K (1991) Determining 3-D object pose using the complex extended Gaussian image. In: IEEE Computer



Complex EGI and Enriched Complex EGI, Fig. 2 Enriched complex extended Gaussian image

Society conference on computer vision and pattern recognition, Maui, HI, pp 580–585
Kang S, Ikeuchi K (1993) The complex EGI, a new representation for 3D pose determination. IEEE Trans Pattern Anal Machine Intell 15(7):707–721

Complex System

Roberta Alfieri and Luciano Milanesi
Institute for Biomedical Technologies – CNR (Consiglio Nazionale delle Ricerche), Segrate, Milan, Italy

Synonyms

Highly structured system

Definition

A complex system is a system composed of interconnected parts which as a whole exhibit one or more properties (behavior among the possible properties) not obvious from the properties of the individual parts. This characteristic of every system is called emergence and is true of any system, not just complex ones.

Complex systems are usually open systems since they exist in a thermodynamic gradient and dissipate energy. In other words, complex systems are frequently far from energetic equilibrium: but despite this flux, there may be pattern stability (Rocha 1991). Complex systems may have a memory: the history of a complex system may be important. Because complex systems are dynamical systems, they change over time, and prior states may have an influence on present states. More formally, complex systems often exhibit hysteresis. Complex systems may be nested since the components of a complex system may themselves be complex systems.

Generally relationships in a complex system are nonlinear. In practical terms, this means a small perturbation may cause a large effect, a proportional effect, or even no effect at all. In linear systems, effect is always directly proportional to cause. Moreover, relationships contain feedback loops: Both negative (damping) and positive (amplifying) feedback are often found in complex systems. The effects of an element's behavior are fed back to in such a way that the element itself is altered (Kitano 2007). Emergent behavior in complex systems arises if all constituents of the system observed on one level cannot explain the system properties on a coarser or higher level (Walleczek 2000).

References

Kitano H (2007) Towards a theory of biological robustness. Mol Syst Biol 3:137
Rocha LM (1991) Systems modeling: using metaphors from nature in simulation and scientific models, Los Alamos National Laboratory, Nov 1999
Walleczek J (2000) Self-organized biological dynamics and nonlinear control. Cambridge University Press, Cambridge

Complexity

Marie I. Kaiser
Department of Philosophy, University of Cologne, Cologne, Germany

Synonyms

Complication; Intricacy

Definition

Complex systems are systems which show a range of different characteristics like the multiplicity of their parts, the nonlinearity and nonadditivity of the interactions between their parts, the sensitivity of their behavior to initial conditions, their hierarchical organization, their self-organization, and the robustness and the emergence of their behavior. Some of these characteristics are necessary conditions for a system to be complex; others are just typical for many complex systems. The latter is due to the fact that nature exhibits a variety of different kinds of complexity.

Characteristics

In recent decades, the focus of scientific research has shifted more and more to trying to understand and

handle the complexity of nature. In biology, for instance, the reductionistic assumption (▶ Reduction) that there is a neat, simple gene-protein-trait-fitness relationship and that the behavior of a biological system can be understood merely by studying the parts of the system in isolation has been rejected. Instead, contemporary biologists try to account for the complexity of nature by recognizing the "wholeness" (▶ Holism) of biological systems (e.g., Chong and Ray 2002), that is, by paying attention to the various interactions between the parts of a system and to how these parts are integrated to the system as a whole. For example, rather than examining the isolated functions of genes, biologists study the dynamics of entire ▶ gene regulatory networks. This focus shift toward complexity issues is not restricted to the biological sciences, but rather, it is a quite general trend. This is why some authors speak of a "complex system revolution" (Hooker 2011, 6) that has taken place and still goes on in contemporary science.

What Is a Complex System?

Despite the fact that complexity issues more and more gain center stage in several research fields, there exists neither a unified science of complex systems nor a consensus about what complexity is and what makes a system complex. Rather, the characterizations of complexity partially vary from field to field and from author to author. There are two different ways of how one could react to this situation. On the one hand, one could argue that there is still much empirical and conceptual work left to do, and that sometime in the near future, scientists and philosophers would have figured out how to specify what a complex system is. On the other hand, one could point out that the actual disagreement on how to characterize complexity is not due to our insufficient knowledge. Rather, it arises from the actual variety of ways that systems are complex. In other words, nature exhibits different kinds of complexity that cannot be captured by a single definition.

Depending on which claim an author subscribes to, he will favor one of two strategies for characterizing complexity: first, several authors try to specify what complexity is by proposing a list of features, each of which is necessary and all together are sufficient for a system to be complex (e.g., Hooker 2011; Ladyman et al. 2012); the second strategy consists in distinguishing different kinds of complexity (e.g., Mitchell 2003; Kuhlmann 2011). In this section, some crucial insights of pursuing the first strategy will be revealed, whereas section "Different Kinds of Complexity" introduces some classifications of complexity (second strategy).

Before doing so, it should be stressed that the two strategies are neither opposite nor incompatible. What distinguishes them is the emphasis they put on the diversity of complexity. Those who aim at a definition of complexity which identifies necessary and jointly sufficient conditions for a system to be complex (first strategy) focus on the similarities among different cases of complexity. On the contrary, those who pursue the second strategy and distinguish different kinds of complexity put more emphasis on the actual variety of ways that systems are complex. However, it is also possible to seek after a list of core features of complex systems and at the same time account for their diversity. For instance, one could abandon the requirement that these features must be necessary and allow also features on the list that are only typical for many (but not for all) complex systems. Furthermore, one could also combine the first and the second strategy by at first distinguishing different kinds of complexity, and then specifying these types of complexity by identifying different sets of features that are associated with these different kinds.

What are the features that are said to be necessary for complexity or that are at least typical for many complex systems? The following main features are widely associated with complex systems (which is not to say that this list is exhaustive; for alternative approaches, see Hooker 2011 or Ladyman et al. 2012).

Multiplicity of Parts

Complex systems typically consist of a *large number of parts*. In some cases, many of these parts are of the same or of similar kind (e.g., a swarm of birds is composed of birds of the same species). Other complex systems are made of components that belong to several different kinds (e.g., organ systems like the cardiovascular system, which consist of different kinds of tissues and cells).

Nonlinearity and Nonadditivity of Interactions

The parts of a complex system causally interact (▶ Causality) with each other in order to bring about a particular behavior of the overall system. It is characteristic for many complex systems that these interactions are *nonlinear*, more precisely, that the behavior of the system is described by a mathematical function

that is nonlinear (e.g., because the variable of interest is squared). Nonlinear interactions frequently involve positive or negative ▶ feedback.

Nonlinear dynamical equations are characterized by *nonadditivity*, which means that numerical combinations of solutions are in general not solutions. The feature of nonadditivity is one reason why complex systems are said to be "more than the sums of their parts." To put it another way, complex systems are not *aggregative systems* (Wimsatt 2007) because their behavior does not remain invariant under interchanging their parts, under changes in the number of parts, and under decomposing and rearranging their parts, and the interactions among their parts are not linear.

Sensitivity to Initial Conditions

The behavior of some complex systems is *sensitive to initial conditions*, that is, their behavior differs largely in the long run even if there are only minuscule differences at the beginning. This is due to the fact that the nonlinearity of the interactions between the system's parts allows small differences in the system state to be amplified (▶ Amplification) into large differences in the subsequent system trajectory.

Hierarchical Organization

Complex systems typically possess a *hierarchical nature* (▶ Hierarchy), that is, they are parts of higher-level systems and they consist of parts that are themselves (lower-level) systems which, in turn, may also be composed of subsystems, and so on. For instance, organisms consist of organ systems that are composed of tissues which consist of cells, etc. This feature is also referred to as the *multilevel character* of complex systems. Recognizing the hierarchical nature of complex systems is important for understanding how complexity can evolve (Simon 1962).

Self-organization

Complex systems are *self-organizing systems*. ▶ Self-organization means that an initially disordered system becomes more organized or ordered because of the interactions of the parts of the system. The process of self-organization is spontaneous, that is, it is not centrally controlled by any agent or subsystem.

Robustness

The organization or order of a complex system is said to be *robust* (▶ Robustness), that is, it remains stable under a certain range of disturbances of the parts of the system.

Emergence

The behavior of complex systems is often said to be emergent (▶ Emergence) in the sense of being unexpected, unpredictable, or unexplainable on basis of the knowledge about the parts of the system in isolation (more on this in section "Emergence and the Limits of Reductionism").

Different Kinds of Complexity

A second way to characterize the phenomenon of complexity is to distinguish different kinds of complexity. At least two ▶ classifications of complexity are worth being mentioned here. The first one has been introduced by Mitchell (2003, 2009). She distinguishes three different kinds of complexity in biology: constitutive complexity, dynamic complexity, and evolved complexity. "*Constitutive complexity*" refers to the complexity of the structure that biological systems (like organisms) display. The structure of biological systems is complex if the system as a whole is being formed of numerous parts in nonrandom, non-simple ▶ organization (see also Simon 1962; Wimsatt 2007). "*Dynamic complexity*" concerns the complexity of the processes that biological systems are engaged in, for instance, the developmental or evolutionary processes that organisms undergo. Finally, "*evolving complexity*" refers to the domain of alternative adaptive solutions that are available for certain adaptive problems. If there exists a wide diversity of forms in life which have evolved as solutions to the same adaptive problem, there is said to be much evolving complexity.

More recently, Kuhlmann (2011) has argued that it is important to distinguish compositional complexity from dynamical complexity. Although he uses similar words as Mitchell, the two kinds of complexity he identifies are different from hers. This difference might (at least partly) be due to the fact that Kuhlmann is more interested in complex systems from physics and socio-economics, rather than from biology. What Kuhlmann means by *compositional complexity* of a system is the complicated organization of the setup conditions of a system, that is, the fact that a system consists of many parts and that the individual behaviors of the parts as well as their organization determine the overall behavior of a system. Somehow surprisingly, Kuhlmann emphasizes that, in this case, the

parts of compositionally complex systems interact with each other in a linear fashion, which is why the behavior of the system is a summation of the behaviors of its parts. Contrary to this, in the case of *dynamical complexity*, most facts about the nature of the parts of the system as well as their initial arrangement have no bearing on the behavior of the system. Rather, what makes these kinds of systems complex is that although they are compositionally simple and the rules that determine their dynamics are simple (but nonlinear), they show patterns that are factually unpredictable and qualitatively unexpectable.

Further Philosophical Issues

Explaining Complexity

One important philosophical question that arises in the context of complex system research concerns the nature of scientific ▶ explanation. Do the explanations of the behavior of complex systems constitute a *special kind of explanations*, as, for instance, Mitchell (2009) and Kuhlmann (2011) argue? Or do they belong to one or more of the established kinds of explanation (like causal-mechanistic explanation, covering-law explanation, functional explanation, mathematical explanation, etc.)? A good starting point for addressing this question is the widespread claim of biologists that reductionistic-mechanistic research strategies (▶ Reduction; ▶ Mechanism) are inappropriate or, at least, insufficient to investigate and explain the behavior of complex systems (e.g., Gallagher and Appenzeller 1999). This suggests that explanations that are developed in sciences of complex systems are *non-reductive* (▶ Explanation, Reductive) and *non-mechanistic*. However, it is far from clear what exactly this claim amounts to and whether it is true for all explanations in this field (e.g., whether it also applies to computational explanations that can be found, for instance, in systems biology). One reason why the explanation of the behavior of complex systems might be non-reductive is that these explanations frequently appeal not only to the parts of the system, but also to contextual factors and higher-level factors (which is why they are said to be multilevel explanations; Mitchell 2009). Moreover, the behavior of complex systems cannot be explained by referring only to the parts of the system in isolation (which is, in turn, typical for reductive explanations; Kaiser 2011). However, in the last 15 years, several accounts of mechanistic explanation have been developed which decouple the concept of a mechanistic explanation from the concept of a reductive explanation because they allow higher-level and contextual factors to figure center stage in a mechanistic explanation (e.g., Bechtel 2008).

Causality and Complex Systems

A second philosophical question concerns the *causal structure* of complex systems (▶ Causality). Do any challenges and constraints for a philosophical theory of causation arise from the peculiarities of the causal structure of complex systems?

A possible challenge might be the existence of downward causation, which is a topic that is also frequently discussed in science itself. Downward causation encompasses cases, in which entities from a higher level of organization causally affect entities on a lower level (▶ Interlevel Causation). Downward causation between a whole (i.e., the higher-level entity) and its parts (i.e., the lower-level entities) is regarded as particularly problematic because a central assumption in most theories of causation is that cause and effect must not be identical. However, if it is true that the relation between a whole and the set of its organized parts is one of identity, as one could argue, the whole cannot be causally related to its parts.

Other challenges and constraints of a theory of causation that may arise from the investigation of complex systems concern the context-sensitivity of their behavior and the nonlinearity of the interactions between their parts. The latter often involve cyclic causal relations like ▶ feedback, which constitute a challenge for some theories of causation (e.g., for causal graph theories).

Emergence and the Limits of Reductionism

One of the features that are characteristic of complex systems is that their behavior is said to be *emergent* (see section "What Is a Complex System"). However, despite its ubiquity, the notion of ▶ emergence is left notoriously vague. Most scientists use the term "emergence" in an epistemic sense, that is, they call a behavior (or property) of a system emergent if it is unpredictable, unexpectable, or unexplainable on the basis of present knowledge about the behavior (or properties) of its parts in isolation.

Others want to understand emergence ontologically and link it to irreducibility. According to them, studying emergent behaviors of complex systems reveals the

limits of reductionism (▶ Reduction). More precisely, it discloses the conditions under which the (exclusive) application of reductive research strategies (like decomposition, simplification of the system's context, and investigating parts in isolation; Wimsatt 2007; Kaiser 2011) is not adequate any more. In other words, the emergent behavior of a complex system cannot be understood by decomposing the system into its parts, by examining the parts in isolation (i.e., separated from the original system), and by ignoring the context of the system. This is, for instance, due to the fact that the parts of a complex system are organized or "integrated" (Bechtel and Richardson 2010) in such a complicated way that their behavior is co-determined by the system's organization. Furthermore, the behavior of several complex systems heavily depends on certain parts of their context (i.e., it is non-robust under variations of these contextual factors), which is why the context of the system cannot be ignored altogether or grossly simplified.

Cross-References

- ▶ Amplification
- ▶ Causality
- ▶ Classification
- ▶ Emergence
- ▶ Explanation in Biology
- ▶ Explanation, Reductive
- ▶ Feedback Regulation
- ▶ Gene Regulatory Networks
- ▶ Hierarchy
- ▶ Holism
- ▶ Interlevel Causation
- ▶ Mechanism
- ▶ Organization
- ▶ Reduction
- ▶ Robustness
- ▶ Self-Organization

References

Bechtel W (2008) Mental mechanisms. philosophical perspectives on cognitive neuroscience. Taylor and Francis, New York/London

Bechtel W, Richardson RC (2010) Discovering complexity. Decomposition and localization as strategies in scientific research. MIT Press, Cambridge, MA

Gallagher R, Appenzeller T (1999) Beyond reductionism. Science 284:7

Hooker C (ed) (2011) Philosophy of complex systems. Elsevier, Amsterdam (= Handbook of the philosophy of science; 10)

Kaiser MI (2011) Limits of reductionism in the life sciences. Hist Philos Life Sci 33:389–396

Kuhlmann M (2011) Mechanisms in dynamically complex systems. In: McKay Illari P, Russo F, Williamson J (eds) Causality in the sciences. Oxford University Press, Oxford, pp 880–906

Ladyman J, Lambert J, Wiesner K (2012) What is a complex system? Eur J Philos Sci (online first)

Mitchell SD (2003) Biological complexity and integrative pluralism. Cambridge University Press, Cambridge

Mitchell SD (2009) Unsimple truths. science, complexity, and policy. University of Chicago Press, Chicago/London

Simon HA (1962) The architecture of complexity. Proc Am Philos Soc 106:467–482

Wimsatt WC (2007) Re-engineering philosophy for limited beings. Piecewise approximations to reality. Harvard University Press, Cambridge, MA

Complication

- ▶ Complexity

Components of Variance

- ▶ Experimental Design, Variability

Composite Motifs

Jianhua Ruan
Department of Computer Science, University of Texas at San Antonio, San Antonio, TX, USA

Definition

A set of transcription factor binding motifs, usually located close to each other on the chromosome, associated with a cooperative set of transcription factors. It is also called *cis*-regulatory modules.

Computational Autopoiesis

Barry McMullin
The Rince Institute, Dublin City University, Dublin, Ireland

Definition

Autopoiesis ("self-production") denotes a particular class of "circular" system organization. A system is said to be autopoietic if its organization satisfies two separate, but intertwined, "closure" properties:

- Closure in *production:* The system is composed of components which give rise to (realize or instantiate) processes of production which, in turn, collectively produce more of those same components
- Closure in *space:* The self-construction of a boundary between the system and the world in which it is embedded, yet from which it distinguishes itself

Computational autopoiesis denotes the study of autopoietic systems realized in computational form – that is, embedded in artificial, computer-generated, virtual worlds.

Characteristics

Motivation

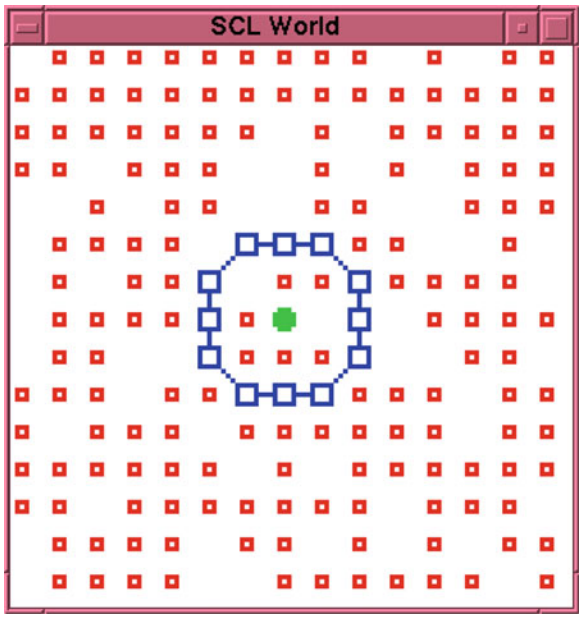
The term *autopoiesis* was coined in the early 1970s by Chilean biologists Humberto Maturana and Francisco Varela. They introduced it in an attempt to characterize what they saw as a decisive organizational distinction between living and nonliving systems – namely, that even the simplest living system (such as a bacterial cell) has the capacity both to regenerate all its own components and also to demarcate itself, as a distinct system, from its surrounding environment. In their view, this entails an essential self-referential circularity in the organization of living systems; and they conjectured that this self-referential circularity is at least a necessary (though hardly sufficient) condition for the distinctive phenomenology of biological systems. This conjecture is clearly relevant to what is now called systems biology insofar as if this demarcation were accepted, then autopoietic organization would be a required property of all properly *biological* systems, that is, of the entire domain of systems biology. Computational autopoiesis is concerned with testing the conjecture by building computational universes (virtual worlds) specifically designed to facilitate embedding of systems with autopoietic organization and investigating the extent to which these can then exhibit lifelike phenomenology within those universes (including dynamic self-maintenance, reproduction, and evolutionary growth of complexity). Computational autopoiesis is commonly classified as a subfield within ▶ Artificial Life.

The Minimal Model

Partly because the definition of autopoiesis was still informal and qualitative, Maturana and Varela offered what they called a concrete "minimal model" of autopoietic organization. With this they put forward a relatively abstract, but still "chemical-like," world, incorporating a minimal set of chemical species and reactions that would still suffice to allow their organization into an autopoietic system. With Ricardo Uribe, they took the highly original (for the time) further step of actually implementing this minimal model in the form of a computer simulation of their abstract chemical world (Varela et al. 1974).

The minimal (computer-simulated) chemistry takes places in a discrete, two-dimensional, space. Each position in the space is either empty or occupied by a single particle. Particles generally move in random walks in the space. There are three distinct particle types (chemical species), engaging in three distinct reactions:

- Production: Two substrate (S) particles may react, in the presence of a catalyst (K) particle, to form a link (L) particle.
- Bonding: L particles may bond to other L particles. Each L particle can form (at most) two bonds, thus allowing the formation of indefinitely long chains, which may close to form membranes. Bonded L particles become immobile.
- Disintegration: An L particle may spontaneously disintegrate, yielding two S particles. When this occurs, any bonds associated with the L particle are destroyed also.



Computational Autopoiesis, Fig. 1 Minimal autopoietic system embedded in 2D computational model. *Small hollow squares* are substrate (S); *solid* disk is catalyst (K); *large hollow squares* with bonds are membrane/link (L) particles

Chains of L particles are permeable to S particles but impermeable to K and L particles. Thus, a closed chain, or membrane, which encloses K or L particles effectively traps such particles.

The basic autopoietic system embedded in this world (Fig. 1) consists of a closed chain (membrane) of L particles enclosing one or more K particles. Because S particles can permeate through the membrane, there can be ongoing production of L particles. Since these cannot escape from the membrane, this will result in the buildup of a relatively high concentration of L particles. On an ongoing basis, the membrane will rupture as a result of disintegration of component L particles. Because of the high concentration of L particles inside the membrane, there should be a high probability that one of these will drift to the rupture site and effect a repair, *before* the K particle(s) escape, thus reestablishing precisely the conditions allowing the buildup and maintenance of that high concentration of L particles.

Practical computer simulation experiments demonstrated that such dynamic, self-maintaining, autopoietic systems can indeed both emerge and persist as discrete, self-demarcated, composite individuals embedded within this abstract world. In the original model, there was also an additional technical feature ("chain-based bond inhibition") which was not made explicit in the published algorithm but was, in fact, necessary to the claimed autopoietic operation. This oversight was subsequently identified through the failure of some independent attempts to reproduce the originally described results. The issue was eventually clarified and resolved (McMullin 2004).

Elaborations

The earliest systematic elaborations of the minimal model of computational autopoiesis were developed by Milan Zeleny (1978). In particular, he reported such phenomena as growth, change in shape, oscillation in chemical activity, and self-reproduction of autopoietic entities. Prima facie, these represented important demonstrations of progressively richer, life-like, phenomena within the framework of computational autopoiesis. However, the algorithmic ("chemical") basis for these phenomena was not fully detailed in the published reports. Technological limitations of the time meant that the source code (the definitive documentation) was not widely distributed, and, indeed, all copies are now believed to have been lost. The qualitative descriptions that are available suggest that, in at least some cases, the original constraints of local interaction, random motion of particles, and time-independent particle dynamics may have been substantially relaxed. If that were so, the interest of these results would be seriously diminished – in the sense that the more complex "macroscopic" phenomena may have been, in effect, implicitly programmed into individual microscopic particles. In any case, this particular line of elaboration was not pursued further.

Some years later, Breyer et al. (1998) described another series of developments beyond the minimal autopoiesis model. They first relaxed the original restrictions on the motion of bound L particles to allow the formation of flexible chains and, indeed, membranes. Allowing multiple, doubly bonded, L particles to occupy a single lattice site further enhanced membrane motion. Adopting more sophisticated "bond rearrangement" interactions then allows chain fragments formed within the cell to be dynamically integrated into the membrane. These mechanisms together obviate the need for the chain-based bond inhibition mechanisms – since now, even if L particles spontaneously bond in the interior of the

cells, the oligomers so formed remain mobile and, through bond rearrangement, can still successfully function to repair membrane ruptures. In this way, the model was reported as successfully supporting cell growth.

More recently, a variation of the minimal model has been presented which demonstrates systematic motion of the autopoietic cell in a manner reminiscent of chemotaxis (Suzuki and Ikegami 2008).

A specific criticism of the minimal model is that it provides no mechanism for production of further K particles (within a functioning autopoietic cell). This is argued to be both a logical and practical deficiency. Logical because, prima facie, it conflicts with the stated requirement of autopoiesis for closure of production. But perhaps more importantly, from a practical point of view, in the absence of ongoing production of K, then *self-reproduction* of this form of autopoietic entity is impossible; and consequently, there is also no possibility for spontaneous Darwinian evolution among variant lineages of such entities.

Breyer et al. (1998) did address this issue to the extent of presenting a number of specific mechanisms for production of further K particles but did not pursue that to the demonstration of autopoietic entities actually capable of sustained reproduction (i.e., which could continue up to the limit of available resources of space and particles).

Further substantive progress in demonstrating evolution of computational autopoietic systems has (to date) not generally focused on direct elaboration of the original minimal model but instead has tended to shift to computer modeling at a more coarse-grained level. This has been at least partially driven by reasons of computational cost in the simulations. In the original model, each lattice site can generally be occupied by at most one particle: It is fine grained to the single molecule level. In the so-called *lattice artificial chemistry* (LAC) models, by contrast, each single lattice site is permitted to host many particles, of many distinct molecular species. Using this approach, a variety of additional phenomena have been reported, including self-reproducing autopoietic cells, with (limited) heritable variation and Darwinian evolution; for example, selection between varieties of catalyst particles having different rates of membrane particle generation (Ono and Ikegami 2002). Some example phenomena, including self-reproduction, have also been reported in three-dimensional LAC models.

Related Concepts

Computational autopoiesis is distinct from, but clearly related to several other, parallel, developments.

The chemoton concept of Tibor Gánti (2003), though apparently developed quite independently, shares significant features with cellular autopoiesis. It is again a proposal for an abstract, "minimal" cell, consisting of a collectively autocatalytic network of reactions which is enclosed within a membrane, which is also generated and maintained by the reaction network. It differs from the minimal autopoiesis model in explicitly including a "genetic subsystem." It is also rather more detailed in its analysis of the required chemical dynamics (kinetics, etc.) and aims at supporting self-reproduction by growth and fission even in the minimal version. The chemoton has been subject to various computer simulation studies. However, these have been based on an ordinary differential equation approach rather than spatially distributed molecular (particle) models. This means that, in particular, the distributed, spatial, dynamics of membrane growth, fission, and individuation have not been substantively modeled in the chemoton context.

Another related line of work involves much more physicochemically realistic computer models of minimal cellular structures. This line of attack poses some scaling difficulties, as the computational demands of physicochemical realism rise very rapidly. On the other hand, the cost of computation continues to fall, so this may well be a very fruitful domain of further research in the near future.

Finally, there are specific conceptual connections between the idea of autopoiesis and Robert Rosen's *metabolism-repair* (*MR*) *systems* and "closure under efficient causation" (Rosen 1991). However, it must be noted that Rosen explicitly argued that his form of closure could *not*, even in principle, be embedded within a purely computational system. By contrast, computer realization has been an explicit *exemplar* of autopoietic closure from the very start. The relationship between autopoiesis (especially in its computational realizations) and closure under efficient causation is, accordingly, an issue of continuing active research and debate.

Cross-References

► Artificial Life

References

Breyer J, Ackermann J, McCaskill J (1998) Evolving reaction-diffusion ecosystems with self-assembling structures in thin films. Artif Life 4:25–40

Gánti T (2003) The principles of life. Oxford University Press, Oxford

McMullin B (2004) Thirty years of computational autopoiesis: a review. Artif Life 10:277–295. doi:10.1162/1064546041255548, http://alife.rince.ie/bmcm-alj-2004/

Ono N, Ikegami T (2002) Selection of catalysts through cellular reproduction. In: Standish R, Bedau MA, Abbass HA (eds) Proceedings of the 8th international conference on artificial life (ALIFE 8). MIT Press, Sydney, Australia, pp 57–64. http://web.archive.org/web/20040518183545/http://parallel.hpc.unsw.edu.au/complex/alife8/proceedings/sub2844.pdf

Rosen R (1991) Life itself. Columbia University Press, New York

Suzuki K, Ikegami T (2008) Shapes and self-movement in protocell systems. Artif Life 15(1):59–70. doi:10.1162/artl.2009.15.1.15104, http://dx.doi.org/10.1162/artl.2009.15.1.15104

Varela FJ, Maturana HR, Uribe R (1974) Autopoiesis: the organization of living systems, its characterization and a model. Biosystems 5:187–196

Zeleny M (1978) APL-AUTOPOIESIS: experiments in self-organization of complexity. In: Progress in cybernetics and systems research, vol III. Hemisphere, Washington, pp 65–84

Computational Creativity

Włodzisław Duch
Department of Informatics, Nicolaus Copernicus University, Toruń, Poland
School of Computer Engineering, Nanyang Technological University, Singapore, Singapore

Synonyms

Ideation; Ingenuity; Innovation

Definition

Computational creativity is the capacity to find solutions that are both novel and appropriate using computational means.

Characteristics

Understanding brain processes behind creativity and modeling them using computational means is one of the grand challenges for systems biology. Computational creativity is a new field, inspired by cognitive psychology and neuroscience. In many respects, human-level intelligence is far beyond what artificial intelligence can provide now, especially in regard to the high-level functions, involving thinking, reasoning, planning, and the use of language. Intuition, insight, imagery, and creativity are important aspects of all these functions. Computational models show great promise both in elucidating mechanisms behind such high-level mental functions, and in applications requiring intelligence (Duch 2007).

Creativity, defined by Sternberg (1998) as "*the capacity to create a solution that is both novel and appropriate,*" has often been understood in a narrow sense, with a focus on big discoveries, inventions, and creation of novel theories, arts, and music, but it also permeates everyday activity, thinking, understanding language, providing flexible solutions to everyday problems (R. Richards, in Runco and Pritzke 2005, pp. 683–688). Creativity research has been pursued mostly in the domain of philosophy, education, and psychology, with many research results published in two specialized journals: *Creativity Research Journal* and *Journal of Creative Behavior*. The "*Encyclopedia of Creativity*" (Runco and Pritzke 2005), written by 167 experts, does not contain any testable neurological or computational models of creativity. MIT Encyclopedia of Cognitive Sciences (Wilson and Keil 1999) contains only a single page (of about 1,100 pages) about creativity, does not mention intuition at all, but devotes six articles to logic, appearing in the index almost 100 times. Logic has never been too successful in modeling real-thinking processes that rely on intuition and creativity. The interest in research on computational and neuroscience approaches to creativity is thus quite recent.

Creativity from the Psychological and Neuroscientific Perspective

D.T. Campbell (1960) described creativity as a two-stage process of blind variation and selective retention (BVSR). This idea is the basis of combinatorial models of creative thinking (Simonton 2010). It is also the

basis of evolutionary biological processes, where the mechanisms of blind variations operate on many levels, with selective retention due to the increased fitness in a given context. In this sense, one can say that viruses, bacteria, and other living organisms exhibit a primitive form of creativity by solving collectively the problem of survival. However, the BVSR idea is more general as it does not have to rely on specific Darwinian mechanisms. It has applications in such diverse fields as immunology, psychiatry, neuroscience, cognitive sciences, memetics, linguistics, anthropology, philosophy, and computer science (Simonton 2010). Blind variation is never random: it is structured by specific interactions of basic elements, from molecular to social, determining probabilities of arising combinations.

Psychological research on creativity has focused on empirical research with gifted children, distinguishing creativity from general intelligence, testing fluency, flexibility, and originality of thought in both visual and verbal domains (Runco and Pritzke 2005). Successful intelligence theory separates creative and cognitive components of intelligence (Sternberg 1998), with creativity implying not only high quality but also novelty. Creativity is not reducible to cognitive thinking skills. The four basic stages of problem solving according to the widely used *Gestalt model* involve preparation, incubation (that may be followed by a period of frustration), illumination (insight), and verification of solution, including communication. These stages, not necessarily in the same sequence, were identified in creative problem solving by individuals and small groups of people.

Boden (1991) defined creativity as "*a matter of using one's computational resources to explore, and sometimes to break out of, familiar conceptual spaces*." Concepts are patterns of brain activations (Pulvermüller 2003; Duch et al. 2008), and exploration of conceptual spaces may be linked to transitions between brain activations. Processing remote, loose associations between ideas is responsible for the associative basis of creativity (Mednick 1962; Simonton 2010). Exploratory creativity is incremental and combinatorial in nature, usually restricted to personal discoveries (novel only for one person), binding diverse activity of brain areas in a new way. "Transformational creativity" leads to ideas that are new for the whole humanity, i.e., big paradigm shifts (Boden 1991). It is not clear whether brain mechanisms behind transformational creativity are really different, requiring a change of the rules that are used to define conceptual spaces, or is it rather due to the linking of many brain patterns that form a new, higher-level complex representing observations in a more coherent way.

Despite the limitations of the current knowledge of the neural processes that give rise to the cognitive processes in the brain, it is possible to propose a testable, neurocognitive model of creative processes. Although direct brain imaging of creative thinking has not been done yet, the "Aha!" phenomenon, or *insight experience* (Sternberg and Davidson 1995) during problem solving, understanding a joke or a metaphor, has been studied using functional MRI and EEG techniques. Brain states during insight were contrasted with analytical problem solving that does not require insight (Kounios and Jung-Beeman 2009). Although the insight experience is sudden, it is a culmination of a series of brain processes. A few seconds before the insight, an alpha burst (i.e., a sudden high amplitude oscillations in the range of 8–12 Hz) is seen in the right occipital cortex (i.e., the visual cortex behind the rearmost portion of the skull), and about 300 ms before the feeling of insight, a burst of gamma activity (in the 40 Hz range) is observed in the right hemisphere anterior superior temporal gyrus (i.e., the upper ridge of the temporal lobe cortex, on the side of the brain). Alpha activity helps to decrease activation of irrelevant cortex after the information stating the verbal problem has been taken in, while gamma burst reflects the connection of distantly related patterns. The right brain hemisphere is able to create more abstract associations based on meanings, avoiding close associations that the left hemisphere is routinely processing. The same neural structures are probably involved in creative thinking. This shows the need for multiple levels of representations of concepts that help to constraint the search for solutions of problems requiring creativity.

Intuition is also a concept difficult to grasp (Lieberman 2000; Myers 2002) but plays an important role in mathematics, science, and general decision making. It has been defined as "*knowing without being able to explain how we know*." Intuition relies on implicit learning, gaining tacit knowledge without being aware of learning. Insight into structural relations is usually not present, only fast judgment or response based on probability estimation. Social

intuition is the basis of nonverbal communication, and can be seen as the phenomenological and behavioral correlate of knowledge gained through implicit learning (Lieberman 2000).

The measurement of intuition is based on several tests and inventories (for example, the Myers-Briggs Type Inventory, or Accumulated Clues Task), but there is little correlation between them, so the concept of intuition is not well defined from an operational point of view. Significant correlations were found between the Rational-Experiential Inventory (REI) intuition scale and some measures of creativity (Raidl and Lubart 2001). Such tests reflect rather complex cognitive processes, and it is not clear which brain processes are behind these measures.

From a computational point of view, it is much easier to create predictive models of data than to provide explanations (Duch 2007). In particular, it is difficult to explain decisions made by neural networks or similarity-based systems. Using such systems for learning from partial observations can constrain the search for solutions, avoiding combinatorial explosion that is a main problem in AI, making the reasoning process feasible.

Creativity from the Computational Perspective

Psychology and neuroscience agree that creativity is a product of ordinary cognitive processes. The lack of understanding of detailed mechanisms involved in creative activity makes the development of creative computing rather difficult. The need for everyday creativity has been almost completely neglected by the artificial intelligence research community and may be credited for failures of many AI programs. Early attempts to model intuition, insight, and inspiration from the AI point of view have been summarized by Simon (1995). His work has mostly been directed at understanding historical discoveries of scientific laws, as well as the search for new scientific knowledge of this kind in astronomy, physics, chemistry, and biology. Simon made no attempts to connect search-based AI approaches to processes in the brain. Research in automatic genetic programming (Koza et al. 2003) can be credited for useful patentable inventions in automated synthesis of antennas, analog electrical circuits, controllers, metabolic pathways, genetic networks, and other areas. These inventions have been mostly restricted to optimized versions of known designs. Genetic programming may be capable of creative problem solving, although the problem of find a good way to represent knowledge domain in which genetic processes operate may be as hard as the original problem. Other approaches to insight include the "small world" network model of Schilling (2005) and the work on fluid concepts and creative analogies (Hofstadter 1995), with some applications to design.

A direct attempt to model creative processes in the brain is still not feasible, but inspirations from the BVSR models may be used in a number of ways. The results of the experimental and theoretical research in this domain can be summarized in three points:

1. *Space*: creativity involves neural processes that are realized in the space of quasi-stable neural activities, leading to patterns of activity that reflect relations between concepts in some domain.
2. *Blind variation*: priming by concepts that represent the problem leads to distributed fluctuating neural activity constrained by the strength of associations between patterns of neural activity coding concepts; this process is responsible for imagination and the flexible formation of transient novel associations.
3. *Selective retention*: filtering of interesting results, discovering partial solutions that may be useful to reach goals, amplifying or forming new associations; in biological systems, this may involve emotional arousal.

The blind-variation process may require some structuring to be effective. Brainstorming, free associations, random stimulation, or lateral thinking have not been very successful in the generation of creative ideas in advertising and product innovation (Goldenberg and Mazursky 2002). However, structured approaches, based on higher-order rules and templates, led to excellent results. Computer-generated ideas based on templates were rated significantly higher both for creativity and originality than the non-template human ideas. The associative processes may in this case have been guided by general rules. Connectionist models for the generation of ideas within the brainstorming context can successfully predict factors that enhance brainstorming productivity. The model of Iyer et al. (2009) is perhaps the most sophisticated, with features, concepts, and cognitive control components as separate neural layers. Ideas emerge in a multilevel, modular semantic space from itinerant attractor dynamics (i.e., activity of neuronal changes in the itinerant way, attracted toward quasi-stable states where it slows down) shaped by context, synaptic

learning, and ongoing evaluative feedback. This model generates novel ideas by multilevel dynamical search in various contexts, capturing the interplay between semantic representations, working memory, focusing attention on various features, and using reinforcement signals. The model is quite useful for the elucidation of the mechanisms of creativity. For example, it has found interesting associations for tourist activities such as "outing and vacation," depending on the information of the context.

The simplest domain in which creativity is frequently manifested and can be studied in experiments as well as computational models is the invention and understanding of neologisms. Poems by Lewis Carroll are full of neologisms, but novel words are also in great demand for products, web sites, or company names. In languages with rich morphological and phonological compositionality (Latin-based, Slavic, and other families), out-of-vocabulary words appear fairly often in conversations. In most cases, the morphology of these words gives sufficient information to understand their meaning. Given keywords or a short description from which keywords are extracted primes the brain at the phonetic and semantic level. The structure of the language is internalized in the neural space. Priming leads to blind variation at the level of word constituents (syllables, morphemes), creating a large number of transient resonant configurations of neural cell assemblies that remain unconscious. This process explores the space of possibilities that agree with internalized constraints on the statistical probabilities of phonological structure (phonotactics of the language) and morphological structure. Imagery processes are approximated in a better way by taking keywords, finding their synonyms to increase the pool of words, breaking words into syllables and morphemes, and combining the fragments in all possible ways. Words that share properties with many other words (i.e., patterns that code them in the brain overlap strongly with patterns for other words) have a higher chance to win the competition for access to awareness. Many variants of words are created around the same morphemes. The same word can be used with many meanings because the context creates specific brain activation patterns for this word. Creative brains spread the activation to more words associated with initial keywords, produce more combinations, selecting the most interesting ones using phonological, affective, and semantic filters.

In computational models, these cognitive processes may be implemented in large-scale neural models, but even the simplest approximations give interesting results (Duch 2006; Duch and Pilichowski 2007). The algorithm involves three major components:

1. An autoassociative memory (AM) structure, trained on a large lexicon to learn statistical properties at the morphological level, providing the model of a neural space and storing background knowledge that is modified (primed) by keywords.
2. Blind-variation process (imagery), forming new strings from combinations of substrings found in keywords and their synonyms, with probabilistic constraints provided by the AM to select only lexically plausible strings.
3. Selective retention ranking the quality of the strings representing neologisms from a phonological and semantic point of view.

Filters should estimate phonological plausibility and "semantic density," or the number of potential associations with commonly known words, calculating the number of substrings in the lexical tokens that may serve as morphemes in each new candidate string. Many factors may be included here: general similarity between morphemes, personal biases, tweaking phonology for neologisms with interesting phonetics. The implementation of this algorithm led to the generation of neologisms with highest ranks that have actually been used as company or domain names in about 75% of cases. For example, starting from an extended list of keywords, *portal, imagination, creativity, journey, discovery, travel, time, space, infinite*, the best neologisms included *creatival* (used by creatival.com), *creativery* (used by creativery.com). Novel neologisms (not found by the Google search engine) included *discoverity*, associated with discovery of something true (verity), and linked to many morphemes: *disc, disco, discover, verity, discovery, creativity, verity*, and through phonology to many others. Another interesting word found is *digventure*, with many associations to *dig* and *venture*.

These examples show that computational approaches to creativity can, at least in restricted domains, lead to results that are comparable with human ingenuity, and that blind-variation selective retention ideas based on the generalization of evolutionary processes may be as useful in cognitive science as they are in life sciences.

References

Boden M (1991) The creative mind: myths and mechanisms. Abacus, London
Campbell DT (1960) Blind variation and selective retention in creative thought as in other knowledge processes. Psychol Rev 67:380–400
Duch W (2006) Computational creativity. IEEE World congress on computational intelligence, 16–21 July. IEEE Press, Vancouver, pp 1162–1169
Duch W (2007) Towards comprehensive foundations of computational intelligence. In: Duch W, Mandziuk J (eds) Challenges for computational intelligence. Springer, Berlin/Heidelberg/New York, pp 261–316
Duch W, Pilichowski M (2007) Experiments with computational creativity. Neural Inform Process Lett Rev 11(4–6):123–133
Duch W, Matykiewicz P, Pestian J (2008) Neurolinguistic approach to natural language processing with applications to medical text analysis. Neural Netw 21(10):1500–1510
Goldenberg J, Mazursky D (2002) Creativity in product innovation. Cambridge University Press, Cambridge
Hofstadter DR (1995) Fluid concepts and creative analogies: computer models of the fundamental mechanisms of thought. Basic Books, New York
Iyer LR, Doboli S, Minai AA, Brown VR, Levine DS, Paulus PB (2009) Neural dynamics of idea generation and the effects of priming. Neural Netw 22:674–686
Kounios J, Jung-Beeman M (2009) Aha! The cognitive neuroscience of insight. Curr Dir Psychol Sci 18:210–216
Koza JR et al (2003) Genetic programming IV: routine human-competitive machine intelligence. Kluwer, New York
Lieberman MD (2000) Intuition: a social cognitive neuroscience approach. Psychol Bull 126(1):109–137
Mednick SA (1962) The associative basis of the creative process. Psychol Rev 69:220–232
Myers DG (2002) Intuition: its powers and perils. Yale University Press, New Haven
Pulvermüller F (2003) The neuroscience of language On brain circuits of words and serial order. Cambridge University Press, Cambridge
Raidl M-H, Lubart TI (2001) An empirical study of intuition and creativity. Imagination, Cognition Personality 20(3):217–230
Runco M, Pritzke S (eds) (2005) Encyclopedia of creativity, vol 1–2. Elsevier, Burlington
Schilling MA (2005) A "Small-orld" network model of cognitive insight. Creativ Res J 17(2–3):131–154
Simon HA (1995) Explaining the ineffable: AI on the topics of intuition, insight and inspiration. In: Proceedings of the 14th international joint conference on artificial intelligence, Montreal, pp 939–948
Simonton DK (2010) Creative thought as blind-variation and selective-retention: combinatorial models of exceptional creativity. Phys Life Rev 7:156–179
Sternberg RJ (ed) (1998) Handbook of human creativity. Cambridge University Press, Cambridge
Sternberg RJ, Davidson JE (1995) The nature of insight. MIT Press, Cambridge, MA
Wilson R, Keil F (eds) (1999) MIT encyclopedia of cognitive sciences. MIT Press, Cambridge, MA

Computational Data Analysis

▶ Data-intensive Research

Computational Identification of microRNA Targets

▶ MicroRNA, Target Prediction

Computational Linguistics

▶ Natural Language Processing

Computational Methods for Mapping B Cell Epitopes

▶ B Cell Epitope Prediction

Computational Methods for Transcriptional Regulatory Networks

Jianhua Ruan
Department of Computer Science, University of Texas at San Antonio, San Antonio, TX, USA

Definition

The transcriptional level of gene expression is controlled, to a large extent, by specific interactions between transcription factors (TFs) and the promoter sequences of their target genes. The interactions between TFs and target genes are combinatorial in nature, and are typically many-to-many, i.e., each TF controls many genes, and a gene can be controlled by many TFs, forming complex transcriptional regulatory networks (TRNs). To understand gene functions in different biological processes, it is necessary to reconstruct such regulatory networks from experimental data.

Characteristics

Data

Input data used for computationally reconstructing genome-scale TRNs may include gene expression data measured by DNA microarrays or RNA-seq, promoter sequences that the cis-regulatory elements may be located, and protein-DNA interaction data obtained from high-throughput experiments such as chromatin immunoprecipitation (ChIP) combined with microarrays (ChIP-chip), chromatin immunoprecipitation combined with next-generation sequencing (ChIP-Seq), and universal protein-binding microarrays (PBM).

These data sets provide different, orthogonal, information and should be combined whenever possible. ChIP-based protein-DNA interaction data provides the most direct evidence of physical interactions between regulators and DNA. However it is relatively costly and may be limited by the availability of antibodies specific to the regulators of interest. Furthermore, protein-DNA interaction is dynamic and is affected by many factors, while ChIP-based data only captures the interaction under a particular condition and at a specific time point. Gene expression data combined with promoter sequences can be combined to reveal dynamic changes of transcriptional regulation under different conditions. However, the simple assumption that such method usually relies on, i.e., co-expressed genes are transcriptionally co-regulated, does not always hold.

In addition, using promoter sequences one can only infer the binding preferences of a transcription factor. The identity of the actual transcription factor needs to be resolved, which is often not easy. The PBM technique may fill this gap by explicitly measuring the affinity between regulators and all possible DNA subsequences of a fixed length.

Unsupervised Methods

Identifying Putatively Co-regulated Genes

An unsupervised method for reconstructing transcriptional regulatory networks usually starts with identifying potentially co-regulated genes. Although co-regulated genes can be directly obtained from protein-DNA interaction data, the most common source is gene expression data analysis, which relies on the simple assumption that co-expressed genes are likely co-regulated. For example, genes that show similar transcriptional responses to the same treatment ("differentially expressed genes") are often considered co-regulated. A number of statistical methods have been developed to identify differentially expressed genes, for example, Statistical Analysis of Microarray (SAM) (Tusher et al. 2001), and Rank Products (Breitling et al. 2004). Statistically, genes exhibiting similar expression patterns across multiple experimental conditions are more likely co-regulated than genes co-expressed under very specific conditions. Therefore, unsupervised methods perform better when the data set contains a large number of experimental conditions. Genes can then be grouped into clusters, where genes in the same cluster have similar expression patterns. This clustering-based method is generally considered a more reliably way of identifying co-regulated genes than methods based on differentially expressed genes. Many clustering algorithms have been introduced from the data mining field, including hierarchical clustering (Eisen et al. 1998), k-means clustering (Tavazoie et al. 1999), self-organizing maps (Tamayo et al. 1999), and graph-theoretic algorithms (Xu et al. 2002) (see Belacel et al. (2006) for a survey). Bi-clustering methods have also been introduced to find clusters of genes that are co-expressed under subsets of conditions (Cheng and Church 2000; Tanay et al. 2002; Turner et al. 2005; Madeira and Oliveira 2004).

Identifying Common cis-Regulatory Elements

After obtaining a list of putatively co-regulated genes, the next logical step is to find the common regulators, or to find common cis-regulatory elements. The latter is more often used, simply because the former relies on protein-DNA interaction data which is still relatively rare and may not be available for the particular condition of interest. In contrast, finding cis-regulatory elements only needs genomic sequence data – it attempts to find common short subsequences from the promoter regions of the co-regulated genes ("motif finding"). Motif finding can also be classified into two categories: de novo motif finding and motif scanning. While de novo motif finding does not require input of known motifs, motif scanning relies on databases of known cis-regulatory elements. With motif scanning, one searches for the occurrence of a list of known cis-regulatory elements (with some tolerance of mismatches) from the promoters of the co-regulated genes, and then reports the ones with the highest number of occurrence (or more accurately, with the highest statistical significance). The de novo motif finding method does not rely on databases of

known cis-regulatory elements; instead, it attempts to identify a subsequence pattern that appears more frequently in the co-regulated genes than would be expected. Such subsequence patterns are often compared with the known cis-regulatory elements for validations. The actual algorithm usually depends on the representation of motifs.

Motif representation. A cis-regulatory element (motif) can be represented by either a consensus or a position-specific weight matrix. A consensus describes the most frequent nucleotide in each position of the binding motif, while a position-specific weight matrix depicts the frequency of each nucleotide occurring at each position of the binding motif.

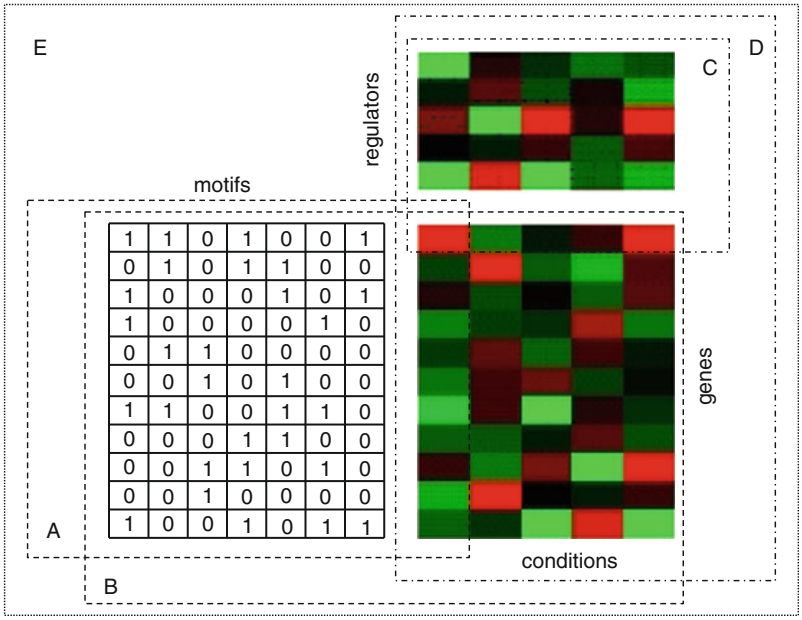
De novo motif finding. Many computational motif finding approaches have been developed. These methods differ from one another in their ways of defining motifs, the objective functions for calculating motif significance, and the search techniques used to find the optimal (or near optimal) motifs. Most of these algorithms can be classified into one of two broad categories: heuristic optimization algorithms based on position-specific weight matrices (PSWM), and enumerative algorithms based on consensuses. Examples of the first category include the well-known programs such as MEME (Bailey and Elkan 1994), AlignACE (Roth et al. 1998), Gibbs Sampler (Lawrence et al. 1993), and BioProspector (Liu et al. 2001), while the second category can be exemplified by Weeder (Pavesi et al. 2001), YMF (Sinha and Tompa 2002), MultiProfiler (Keich and Pevzner 2002), and Projection (Buhler and Tompa 2002).

Motif scanning. Motif scanning is relatively straightforward. Regardless of the motif representation, two issues need to be decided. First, how to score the matches between a known motif and a subsequence of the same length on a promoter sequence; second, what cutoff should be used to select true matches instead of random matches. In the case of consensuses, the simplest scoring function can be the number of matched nucleotides, and a cutoff is typically decided empirically based on the number of mismatches that can be allowed. Some programs can also take into consideration the nonuniform distribution of the different nucleotides and assign them different scores. In the case of position-specific weight matrices, the scoring function is usually based on the information score.

Supervised Methods

Another major direction to model transcriptional regulatory networks from high-throughput data relies on supervised machine learning. Unlike the unsupervised methods, the supervised methods do not assume knowledge of co-regulated genes. Instead, it builds quantitative or qualitative models to capture possible interactions between transcription factors and target genes. The assumption is that if a simple model can be used to explain the expression levels of one or more genes, then the variables used by the model are likely responsible for regulating the genes. Figure 1 shows the relationship between several types of methods under the supervised machine-learning framework.

The first type of supervised methods attempts to model gene expression levels/patterns with the regulatory motifs on promoter sequences (Fig. 1, Boxes A and B). In Box A, the expression levels of multiple genes under a single condition are modeled as a function of motif occurrences on their promoters. That is, each gene is treated as an instance, and attributes are the motifs. Gene expression levels under a particular condition are the response variables, which can be either discrete or continuous. Commonly used models include regression-based and rule-based. Rule-based models have relatively better interpretability over regression-based models. For examples, Bussemaker et al. (2001) and others (Keles et al. 2002; Conlon et al. 2003) modeled the expression levels of genes as a linear regression of putative binding motifs, and applied feature selection techniques to find the most significant motifs. Hu et al. (2000) used decision rules to find motif combinations that best separate two sets of genes. Simonis et al. (2004) combined a string-based motif finding method and linear discriminant analysis to identify motif combinations that can separate true regulons from false ones. In Box B, the expression patterns of multiple genes under multiple conditions are modeled by the motifs on their promoters. The difference between these methods and the ones represented by Box A is that each gene in Box B has multiple response variables. To solve this problem one can either pre-cluster the genes and then use gene cluster IDs as response variables, or conduct clustering and motif finding simultaneously. Probabilistic graphical models, e.g., Bayesian networks, were used to explain gene expression patterns from motifs (Beer and Tavazoie 2004). Phuong et al. (2004) applied multivariate regression trees to model the



Computational Methods for Transcriptional Regulatory Networks, Fig. 1 Schematic representation of existing supervised machine-learning methods for modeling transcriptional regulation. The *bottom right matrix* represents gene expression levels. The *bottom left matrix* represents motif occurrences on promoter sequences. The *top matrix* is the expression levels of regulators. *Box A*, the expression levels of multiple genes under a single condition are modeled by the motifs on their promoter. *Box B*, the expression levels of multiple genes under multiple conditions are modeled by the motifs on their promoters. *Box C*, the expression levels of a single gene under multiple conditions are modeled by the expression levels of putative regulators. *Box D*, the expression levels of multiple genes under multiple conditions are modeled by the expression levels of putative regulators. *Box E*, the expression levels of multiple genes under multiple conditions are modeled by the motifs on their promoters and the expression levels of putative regulators

transcriptional regulation of gene expressions over several time points simultaneously. In these two methods, motifs will be associated with a cluster of genes that have similar expression patterns; this is similar to the unsupervised methods. However, in these methods clustering of genes and identification of cluster-specific motifs are performed simultaneously.

The second type of supervised methods attempt to model the expression levels of target genes by the expression levels of other genes, e.g., TFs and other regulators (Fig. 1, Boxes C and D). In Box C, the expression levels of a single gene under multiple conditions are modeled as a function of the expression levels of putative regulators under the same set of conditions. That is, each condition is treated as an instance, and the expression levels of putative regulators under the condition are its attributes. Similar to the first type of methods, both rule-like and regression-based methods have been commonly applied. For example, Qian et al. (2003) applied support vector machines (SVMs) to predict the targets of 36 yeast transcription factors by identifying subtle relationships between their expression profiles. Soinov et al. (2003) and Segal et al. (2003a) used decision trees and regression trees for similar purposes. The method in Soinov et al. (2003) used decision tree to identify possible regulators for several cell-cycle genes individually, while the method in Segal et al. (2003a) proposed a more sophisticated procedure suitable for whole-genome analysis. The method first clusters genes according to their expression patterns, and then builds a regression tree for each cluster to represent their common regulation program. The procedure then iteratively refines the clusters and the trees.

Finally, the third type of supervised methods model gene expression levels using both putative binding motifs on promoter sequences and the expression levels of putative regulators (Fig. 1, Box E). For example, Middendorf et al. (2004) used an ensemble of

decision trees to model gene expression levels by combining putative binding motifs and the expression levels of putative TFs. The method by Ruan and Zhang (2006), Bi-Dimensional Regression Tree (BDTree), is an extension to the multivariate regression tree approach of Segal (1992) and Phuong et al. (2004). A multivariate regression tree recursively splits genes into groups, where the genes in each group have similar selected attributes (motifs) and responses (Segal 1992). Phuong et al. (2004) extended the method to handle multiple responses, so that the instances in each group have a similar pattern of responses across multiple conditions. The basic idea of BDTree, as suggested by its name, is to extend the multivariate regression tree approach to both dimensions of the expression matrix (Fig. 1, Box E). On one dimension, each gene is treated as an instance, where the attributes are the binding motifs on its promoter sequence, and the responses are its expression levels under different conditions. On this dimension, the approach recursively partitions genes so that those in the same subset have some common binding motifs and similar expression patterns across the conditions. On the other dimension, each condition is treated as an instance, while the attributes are the expression levels of candidate regulators under the condition, and the responses are the expression levels of genes under that condition. BDTree recursively partitions the conditions so that the expression levels of the regulators and target genes within each subset of conditions are similar. The final model, represented by a tree, can be considered as a set of rules, each of which has the form of "if a gene has binding motif A, it will be up-regulated when TF B is up-regulated."

Other Methods

Besides the supervised and unsupervised methods, several other approaches have been proposed to explicitly identify synergies among regulators or regulatory elements without clustering or classifying genes. For example, Pilpel et al. (2001) analyzed the combinatorial effects of motif pairs on gene expression profiles and identified many significant motif combinations. Also, several methods have been developed to combine gene expression clustering with additional information, such as promoter sequences, cellular functions, and genomic localization data (Ihmels et al. 2002; Bar-Joseph et al. 2003). Hybrid methods that combine gene clustering and classification iteratively have also been developed (Segal et al. 2003b; Beer and Tavazoie 2004).

References

Bailey TL, Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol 2:28–36

Bar-Joseph Z, Gerber GK, Lee TI, Rinaldi NJ, Yoo JY, Robert F, Gordon DB, Fraenkel E, Jaakkola TS, Young RA, Gifford DK (2003) Computational discovery of gene modules and regulatory networks. Nat Biotechnol 21:1337–1342

Beer MA, Tavazoie S (2004) Predicting gene expression from sequence. Cell 117(2):185–198

Belacel N, Wang Q, Cuperlovic-Culf M (2006) Clustering methods for microarray gene expression data. OMICS 10:507–531

Breitling R, Armengaud P, Amtmann A, Herzyk P (2004) Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. FEBS Lett 573(1–3):83–92

Buhler J, Tompa M (2002) Finding motifs using random projections. J Comput Biol 9:225–242

Bussemaker HJ, Li H, Siggia ED (2001) Regulatory element detection using correlation with expression. Nat Genet 27:167–171

Cheng Y, Church GM (2000) Biclustering of expression data. Proc Int Conf Intell Syst Mol Biol 8:93–103

Conlon EM, Liu XS, Lieb JD, Liu JS (2003) Integrating regulatory motif discovery and genome-wide expression analysis. Proc Natl Acad Sci USA 100:3339–3344

Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A 95:14863–14868

Hu YJ, Sandmeyer S, McLaughlin C, Kibler D (2000) Combinatorial motif analysis and hypothesis generation on a genomic scale. Bioinformatics 16:222–232

Ihmels J, Friedlander G, Bergmann S, Sarig O, Ziv Y, Barkai N (2002) Revealing modular organization in the yeast transcriptional network. Nat Genet 31:370–377

Keich U, Pevzner PA (2002) Finding motifs in the twilight zone. Bioinformatics 18:1374–1381

Keles S, van der Laan M, Eisen MB (2002) Identification of regulatory elements using a feature selection method. Bioinformatics 18:1167–1175

Lawrence CE, Altschul SF, Boguski MS, Liu JS, Neuwald AF, Wootton JC (1993) Detecting subtle sequence signals: a gibbs sampling strategy for multiple alignment. Science 262:208–214

Liu X, Brutlag DL, Liu JS (2001) Bioprospector: discovering conserved DNA motifs in upstream regulatory regions of co-expressed genes. Pac Symp Biocomput 2001:127–38

Madeira SC, Oliveira AL (2004) Biclustering algorithms for biological data analysis: a survey. IEEE/ACM Trans Comput Biol Bioinform 1:24–45

Middendorf M, Kundaje A, Wiggins C, Freund Y, Leslie C (2004) Predicting genetic regulatory response using classification. Bioinformatics 20(Suppl 1):I232–I240
Pavesi G, Mauri G, Pesole G (2001) An algorithm for finding signals of unknown length in DNA sequences. Bioinformatics 17:S207–S214
Phuong TM, Lee D, Lee KH (2004) Regression trees for regulatory element identification. Bioinformatics 20(5):750–757
Pilpel Y, Sudarsanam P, Church GM (2001) Identifying regulatory networks by combinatorial analysis of promoter elements. Nat Genet 29:153–159
Qian J, Lin J, Luscombe NM, Yu H, Gerstein M (2003) Prediction of regulatory networks: genome-wide identification of transcription factor targets from gene expression data. Bioinformatics 19:1917–1926
Roth FP, Hughes JD, Estep PW, Church GM (1998) Finding DNA regulatory motifs within unaligned noncoding sequences clustered by whole-genome mRNA quantitation. Nat Biotechnol 16:939–945
Ruan J, Zhang W (2006) A bi-dimensional regression tree approach to the modeling of gene expression regulation. Bioinformatics 22:332–340
Segal MR (1992) Tree-structured methods for longitudinal data. J Am Stat Assoc 87:407–418
Segal E, Shapira M, Regev A, Pe'er D, Botstein D, Koller D, Friedman N (2003a) Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. Nat Genet 34:166–176
Segal E, Yelensky R, Koller D (2003b) Genome-wide discovery of transcriptional modules from DNA sequence and gene expression. Bioinformatics 19(Suppl 1):i273–i282
Simonis N, Wodak SJ, Cohen GN, van Helden J (2004) Combining pattern discovery and discriminant analysis to predict gene co-regulation. Bioinformatics 20(15):2370–2379
Sinha S, Tompa M (2002) Discovery of novel transcription factor binding sites by statistical overrepresentation. Nucleic Acids Res 30:5549–5560
Soinov LA, Krestyaninova MA, Brazma A (2003) Towards reconstruction of gene networks from expression data by supervised learning. Genome Biol 4:R6
Tamayo P, Slonim D, Mesirov J, Zhu Q, Kitareewan S, Dmitrovsky E, Lander ES, Golub TR (1999) Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. Proc Natl Acad Sci USA 96:2907–2912
Tanay A, Sharan R, Shamir R (2002) Discovering statistically significant biclusters in gene expression data. Bioinformatics 18(Suppl 1):S136–S144
Tavazoie S, Hughes JD, Campbell MJ, Cho RJ, Church GM (1999) Systematic determination of genetic network architecture. Nat Genet 22:281–285
Turner HL, Bailey TC, Krzanowski WJ, Hemingway CA (2005) Biclustering models for structured microarray data. IEEE/ACM Trans Comput Biol Bioinform 2:316–329
Tusher VG, Tibshirani R, Chu G (2001) Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci USA 98(9):5116–5121
Xu Y, Olman V, Xu D (2002) Clustering gene expression data using a graph-theoretic approach: an application of minimum spanning trees. Bioinformatics 18:536–545

Computational microRNA Biology

Julio Vera and Ulf Schmitz
Department of Systems Biology and Bioinformatics, Institute of Computer Science, University of Rostock, Rostock, Germany

Synonyms

Kinetic modeling of microRNA regulation; Mathematical modeling of microRNA regulation; MicroRNA; MicroRNA clusters; MicroRNA databases and Web resources; MicroRNA gene prediction; MicroRNA synthesis; MicroRNA target hubs; MicroRNA target prediction; MicroRNA target regulation; Regulatory networks

Definition

MicroRNAs (miRNAs) are small regulatory RNAs of ~22nt length that regulate the activity and stability of a large number of messenger RNAs (Bartel 2004; Ambros 2004; Filipowicz et al. 2008). miRNAs bind to specific messenger RNAs and target them for cleavage, deadenylation, or translational repression. In all three processes, protein synthesis of the targeted messenger RNA is prevented. Until today, a vast amount of human miRNAs has been detected (1.524 human miRNAs are listed in the ▶ miRBase database, r18). Many of them are involved in the regulation of relevant cellular processes, including cell signaling, metabolism, apoptosis, and developmental processes (Ambros 2004).

In addition to their function in cancer, numerous studies have validated the role of miRNAs as key modulators in other high prevalence diseases, including cardiovascular disorders like fibrosis and atherosclerosis, neurodegenerative diseases, autoimmunity, and infectious processes including pulmonary infections (▶ microRNA, Disease and Therapy).

On the other hand, strategies have been established that use miRNAs as therapeutic agents by either downregulating those that are abnormally overexpressed or replacing silenced ones with stable nucleotide constructs.

Characteristics

The regulation of basic cell functions is controlled by complex intracellular biochemical networks involving interacting genes, proteins, and small molecules. Recently, an additional level of regulation has been discovered subsequent to the identification of a class of short non-coding RNAs called microRNAs. The first two miRNAs (lin-4 and let-7) were identified in *Caenorhabditis elegans* (Lee et al. 1993). Later, homologues of these were found in higher organisms like *D. melanogaster* and soon also in humans. This increased the interest in these small post-transcriptional regulators, and subsequently many more miRNAs have been found. Today, it is known that miRNAs are involved in the regulation of many critical cellular processes. For example, some of them post-transcriptionally repress genes that control critical features in developmental processes, including developmental timing and stem cell maintenance in both plants and animals (Carrington and Ambros 2003). Additionally, many studies associate miRNAs with a whole bunch of diseases either as mediator or, when deregulated, as causal factor. Here, of highest interest is the implication of miRNAs in tumourigenesis via regulation of target genes which play a role in tumor development and progression, like Ras, c-Myc, BCL2, and cell cycle-dependent kinases (Garzon et al. 2009). In this context, miRNAs can act as tumor suppressors when they downregulate oncogenes (Takamizawa et al. 2004), or like oncogenes (a.k.a. *oncomirs*) by inhibiting tumor suppressor expression (Iorio et al. 2005). Other miRNAs play a fundamental role during malignant transformation (the so-called metastamir, Hurst et al. 2009, ▶ MicroRNA Families, Cancer Progression).

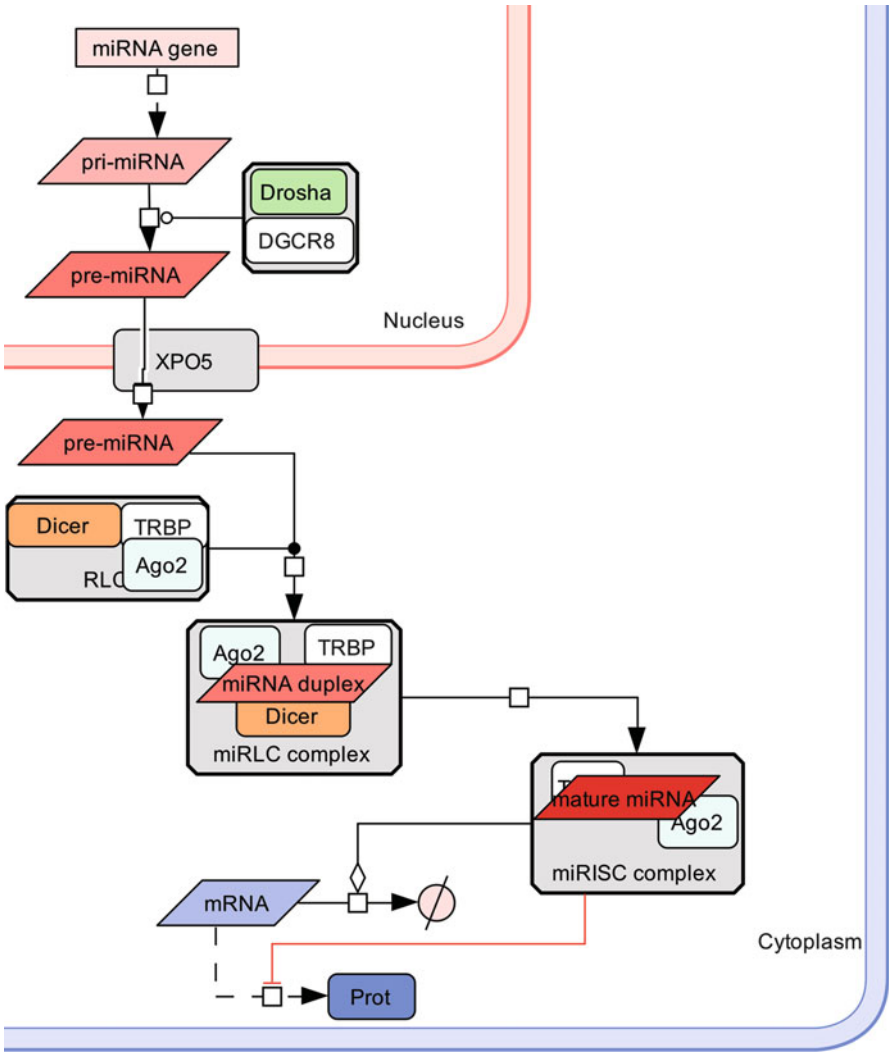
MiRNA Biogenesis and Target Regulation

MicroRNA Biogenesis. miRNAs go through a complex biogenesis process in which a maturated single-stranded functional miRNA is generated from a miRNA gene (Fig. 1). MiRNA biogenesis involves several successive steps: (1) transcription of the miRNA gene into a primary transcript, pri-miRNA (miRNA genes are often embedded in introns of protein-coding genes or clusters of several consecutive miRNA genes); (2) cleavage into a ~80-nt-long hairpin RNA (stem-loop structure) referred to as precursor miRNA (pre-miRNA) via a microprocessor complex composed of Drosha and DGCR8; (3) export from the nucleus to the cytoplasm via Exportin 5; (4) Dicer-mediated cleavage of the pre-miRNA into a short miRNA duplex; and (5) Argonaute2 controlled final processing, which generates a RNA strand that is incorporated into the RNA silencing complex (RISC).

This protein-RNA complex structure can bind, in a sequence-dependent manner, to the 3′ untranslated region (UTR) of messenger RNA targets (▶ MicroRNA Biogenesis, Regulation). Hybridizations with the 5′ UTR and ORF of targets have been observed but are typically nonfunctional in terms of target repression.

MicroRNA Target Regulation. In miRNA-induced target repression, the mature miRNA, previously loaded to the RISC complex, binds to a specific locus on the targeted messenger RNA (▶ Target Site). Success in the hybridization of miRNA-RISC with the target site depends on the level of sequence complementarity, where the so-called seed (▶ microRNA, Seed) region, nucleotides 2–8 nt of the miRNA, plays a major role in miRNA-target recognition. Successful hybridization of miRNA-RISC with the target side leads to post-transcriptional repression of the target gene (Fig. 2). The degree of sequence complementarity between the targeted mRNA and the microRNA can designate the mechanism by which the expression of a gene is hindered. Rarely observed in animals but common in plants are sites with extensive or perfect sequence complementarity, in which miRNAs direct Argonaute-catalyzed cleavage of the target mRNA (▶ Target Cleavage). With imperfect complementarity, a miRNA bound to a mRNA target site can induce deadenylation and subsequent destabilization (target deadenylation), or it can repress protein synthesis of the target by blocking translation initiation or elongation, or by causing early translation termination.

The identification of miRNA targets can be achieved through various in silico, in vitro, and in vivo techniques (target identification, microRNA). These range from predictions of computational algorithms to experimentally driven predictions using gene expression assays, sequencing, or mass spectrometry. Targets are typically validated by reporter gene analyses, immunoblotting, or immunoprecipitation. The aim of *target regulation studies* is to quantify changes in target mRNA and protein concentration upon miRNA overexpression or knockdown.



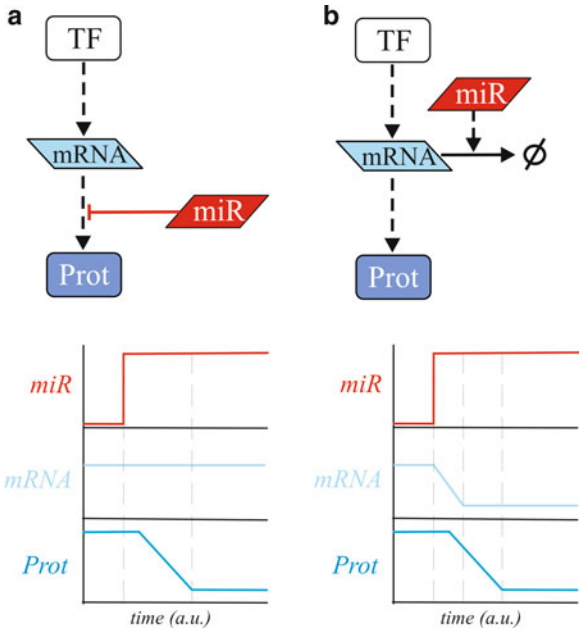
Computational microRNA Biology, Fig. 1 *Sketch of the miRNA biogenesis pathway*. The primary transcript of a miRNA gene (pri-miRNA) is cleaved in the nucleus by the endonuclease Drosha, yielding a 60–70-nt stem-loop intermediate (pre-miRNA). The pre-miRNA is translocated into the cytoplasm and there cleaved by Dicer. The resulting double-stranded RNA associates with Argonaute2 a member from the Argonaute protein family (Ago2) to form the miRNA-RISC loading complex (miRLC). One strand of the miRNA duplex guides the RNA-induced silencing complex (RISC) to specific mRNA target that is recognized by a certain sequence complementarity to the miRNA. This leads to either repression or Argonaute-induced degradation of the target, with subsequent reduction in the protein level. ▶ MicroRNA Biogenesis, Regulation

Networks Involving miRNA Regulation

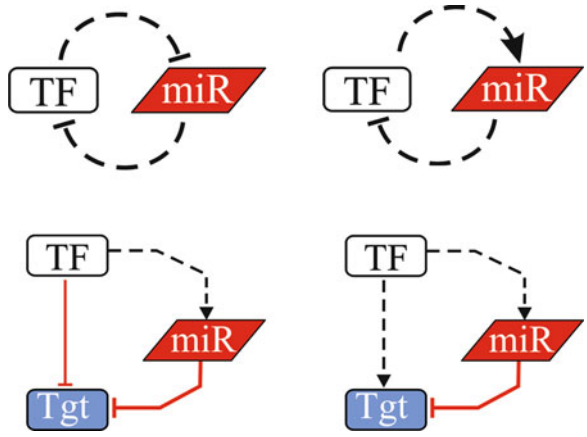
MicroRNAs are often embedded in complex gene and signaling networks composed of transcription factors, miRNAs, and signaling proteins. Interestingly, miRNA-regulated networks are often enriched in regulatory motifs like feedback and feed-forward loops. Moreover, they exhibit other complex motifs that are specific to miRNA regulation.

Feedback Loops. Often miRNAs regulate the expression of signaling proteins that are involved in feedback loop structures. In other cases, the expression of miRNAs can be regulated by their own targets or their interaction partners (▶ MicroRNA Regulation, Signaling Pathways). MiRNAs embedded in feedback loop systems can reinforce a given transcriptional program or be used as a backup option to ensure the activation of those programs.

MiRNAs are involved in *positive feedback loops*; in the simplest of these structures, the expression of a miRNA is inhibited by one of its target proteins (Fig. 3 top left). This motif can induce bistable expression of both the miRNA and its target, while under certain conditions this system transforms a transient activation signal into a sustained cellular response, driven by the miRNA target or downstream signal mediators. This confers robustness to those systems against fluctuations in gene expression and noise in the activation signal (Inui et al. 2010; ▶ MicroRNA Regulation, Feed-Forward Loops). MiRNAs also participate in *negative feedback loops*; in the simplest



Computational microRNA Biology, Fig. 2 *Sketch of the different modes of miRNA target repression.* A microRNA can regulate its target by (**a**) blocking translation initiation or causing early translation termination (*left, translation repression*) or (**b**) by inducing deadenylation and subsequent destabilization (*right, target deadenylation and destabilization*). Extensive or perfect sequence complementarity between microRNA and mRNA leads to Argonaute-catalyzed cleavage of the target mRNA, which is rare in animals but oftener in plants (not shown, ▶ target cleavage)

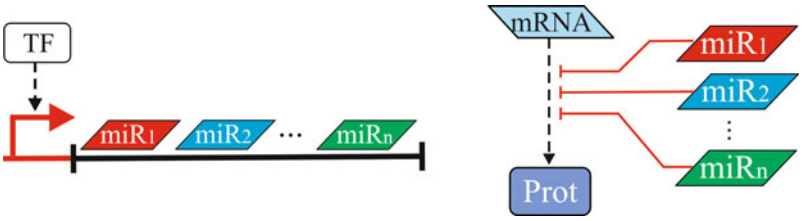


Computational microRNA Biology, Fig. 3 Basic regulatory motifs involving miRNA regulation

configuration, a target activates the expression of its repressing miRNA (Fig. 3 top right). Those systems confer homeostasis to network response, fine-tune gene expression, and can induce fast decay or signal termination (Tsang et al. 2007, ▶ MicroRNA Regulation, Feed-Forward Loops).

Feed-Forward Loops. MicroRNAs are also involved in many gene and signaling networks with feed-forward structures. In this case, the combination of transcriptional and miRNA-mediated post-transcriptional regulation plays a crucial role in controlling gene expression. Feed-forward loops containing miRNAs consist of three components: a transcription factor (TF), a miRNA directly regulated by the TF, and a target gene, which is regulated by both the TF and the miRNA (▶ MicroRNA Regulation, Feed-Forward Loops). There exist two types of miRNA-mediated feed-forward loops. In *coherent feed-forward loops*, a target gene is consistently regulated by direct TF interaction and indirect TF interaction realized through miRNA repression (Fig 3. bottom left). In these motifs, the transcription factor and the miRNA repress the expression of the target gene at the transcriptional and post-transcriptional level to ensure complete gene silencing. In *incoherent feed-forward loops*, a target is regulated by a TF and a miRNA, but with opposite signs (positively and negatively), and is therefore inconsistently regulated (Fig 3. bottom right). Incoherent miRNA-mediated feed-forward loops can display accelerated response as compared to the simple regulation motif, but they can also realize a noise buffering function, fine-tuning of the target gene, and can have the ability to convert step-like activation into transient peak activation (Osella et al. 2011).

MicroRNA Clusters. miRNA clusters are sets of two or more miRNAs that are (a) transcribed from physically adjacent miRNA genes, (b) transcribed in the same orientation, and (c) not separated by a transcriptional unit or a miRNA in the opposite orientation (▶ MiRNA Cluster). Typically, members of miRNA clusters are transcribed in synchrony upon an activation event (Fig. 4 left). MiRNAs integrated in a cluster sometimes target common but more often different target genes. One example is the miR-17-92 cluster, which is found on the human chromosome 13, and it is composed of six miRNAs. Among others, the transcription factor c-Myc binds to the upstream region of the miR-17-92 cluster to induce its expression. Interestingly, two miRNAs belonging to the cluster



Computational microRNA Biology, Fig. 4 Illustration of a miRNA cluster (*left*) and miRNA target hub regulation (*right*)

(miR-17-5p, miR-20a) post-transcriptionally regulate E2F1, which is itself a transcription factor that cross activates c-Myc. Thus, a complex regulatory network of miRNAs and TFs is formed, including negative and positive feedback loops (Aguda et al. 2008).

MicroRNA Target Hubs. MiRNA target hubs are genes targeted by many miRNAs (▶ Target Hub, Fig 4. right). These miRNAs might belong to the same miRNA cluster or are subject to totally different and independent transcriptional regulation. Under the assumption that miRNA target hubs are regulated by at least 15 miRNAs, Shalgi and coauthors (2007) found 470 of such genes in the human genome. It has been suggested that miRNA target hubs are integrated in larger regulatory networks, involving transcription factors, signaling proteins, and miRNAs, enriched with many of these regulatory motifs introduced above (feed-forward and feedback loops). This multiplicity of miRNA target hub regulation can lead to highly nonlinear features like miRNA-mediated cross talk, cooperation between different transcription factors, and generation of tissue-specific expression patterns (Lai et al. 2012).

Bioinformatic Tools and Algorithms for miRNA Investigation

Advancements in the understanding of miRNA biology have always been paved by bioinformatic tools and algorithms. Computational methods are being used for the prediction of miRNA genes and their targets, for the analysis of high-throughput expression experiments and functional assessment of miRNA targets. Furthermore, bioinformatic approaches are used for data integration and knowledge transfer.

MiRNA Gene Identification

Around the beginning of this century, it became clear that there exist more miRNAs than just *lin-4* and *let-7*. And it was found that many of them are conserved in different species. At that time, methods based on ▶ non-coding RNA prediction algorithms were developed that could help to uncover many more putative miRNA genes. In ▶ miRNA gene prediction, sequence-based approaches try to find genomic regions which may reveal stem-loop structures when being expressed. Such methods are often complemented by a conservation analysis of candidate sequences and their predicted structures. Additional approaches investigate structural properties, like stability and robustness, of the stem loop in order to reduce the space of candidates. Machine-learning approaches use known miRNA precursors as training set and exploit their features as criteria to classify new candidates. Transcriptiomics based approaches search in ▶ RNA-seq data for short transcripts that can be mapped to genomic loci that may form stem loops upon expression. Nowadays, the search for miRNA genes in newly sequenced genomes benefits from the large body of known miRNA genes, which can be used in a homology search in order to come up with an initial set of miRNA candidates (Mendes et al. 2009).

Target Prediction

The first step in the functional characterization of miRNAs and their association with biological processes is the identification of miRNA-regulated mRNA targets. Shortly after the first miRNA-target pairs have been identified experimentally, computational algorithms for the prediction of more targets have been developed (▶ MicroRNA Target Prediction). Soon, patterns in miRNA-target recognition emerged and were exploited to improve target prediction results. A central criterion for target candidates is the sequence complementarity between a potential ▶ target site and the mature miRNA, with special emphasis on the ▶ seed region. Other criteria that suggest the presence of functional miRNA-mRNA pairs are the evolutionary conservation of the target site sequence, the thermodynamic stability of the putative miRNA-mRNA duplex, and the multiplicity of miRNA binding sites in the 3′ UTR of the mRNA target. For many algorithms, target predictions were made accessible in public databases. Some examples

Computational microRNA Biology, Table 1 *MicroRNA Web resources*. It has to be noted that URLs might change for unforeseen reasons. Similarly, users should be aware of the date of the last update of the database to avoid the use of outdated information

Relevant microRNA Web resources		
Name	Description	URL
miRBase	Primary miRNA registry	www.mirbase.org
	Pre-miRNA and mature miRNA sequence and structure information	
microRNA.org	miRanda predicted miRNA targets	www.microrna.org
miRecords	Database of validated miRNA-target interactions	http://mirecords.biolead.org
miRGator v2.0	miRNA expression profiles	http://mirgator.kobic.re.kr
	Functional characterization of miRNAs	
miR2Disease	miRNA disease relationships	www.mir2disease.org
miRWalk	Target predictions from eight algorithms	http://mirwalk.uni-hd.de
	Validated targets	
	Predicted functional associations	

of single and composite Web-accessible resources of miRNA target predictions are listed in Table 1.

MiRNA Web Resources

Apart from databases for miRNA target predictions, there exists a large variety of other useful miRNA Web resources and Web services (► MicroRNA Web Resources). The data provided includes (a) primary data of miRNA genes, their sequences, and expression profiles; (b) processed data, like functional assignments and computational predictions; and (c) collections of literature-based knowledge. The most prominent example of a miRNA database is the ► miRBase database, the primary miRNA sequence repository. New miRNA sequences are registered and annotated in this database. Additionally, it is a resource for sequence and structure information of pre-miRNAs and mature miRNAs (Enright and Griffiths-Jones 2008). Other examples are databases of target predictions (e.g., miRWalk), collections of validated miRNA-target interactions (e.g., miRecords), databases of miRNA expression profiles (e.g., microRNA.org), and databases of functional miRNA characterization (e.g., miR2Disease). An overview of some relevant miRNA Web resources, a brief description, and their URL is provided in Table 1.

Mathematical Modeling of miRNA Regulation

In complex biochemical networks involving transcriptional regulation, miRNA-mediated repression, protein-protein interactions, and subcellular compartmentalization of species, the spatiotemporal organization of those networks can hardly be understood without mathematical modeling. There is no default modeling approach to be used in the investigation of microRNA regulation. The appropriate modeling framework to be used depends on several circumstances, for example, the extent of biomedical knowledge that exists, the quality and quantity of the experimental data available, and the nature of the system's properties to be analyzed.

Kinetic Modeling of miRNA Regulation. Kinetic models describe the evolution in time of the expression levels for the different biochemical species (mRNAs, microRNAs, TFs, signaling proteins) involved in regulatory networks. In this kind of models, reaction rates account for biochemical events like processing, degradation, and interaction of messenger RNAs, miRNAs, and proteins, including the miRNA-induced regulation of targeted genes. An exemplary model structure is presented in Table 2.

Values of the model parameters characterizing the reaction rates are assigned by using parameter values taken from relevant publications and/or by applying data fitting techniques over quantitative data. In the context of miRNA regulation, several studies have used kinetic models to investigate miRNA regulation, some of which have addressed general design principles of miRNA regulation. For example, Levine and coauthors (2007) developed a simple quantitative model for miRNA-mediated target repression and used it to investigate how distinctive

Computational microRNA Biology, Table 2 An exemplary model structure

$\frac{d}{dt}mRNA_i = k_{s\_e2} \cdot F_{act}(TF_{e2}) - k_{d\_me2} \cdot mRNA_i - \sum_i k_{a\_mRi} \cdot miR_i \cdot mRNA_i$ $\frac{d}{dt}miR_i = k_{s\_mRi} \cdot F_{act}(ProtX, TF_{mR_i}) - k_{d\_mRi} \cdot miR_i - k_{a\_mRi} \cdot miR_i \cdot mRNA_j$ $\frac{d}{dt}Tgt_i = k_{s\_e2} \cdot mRNA_i - k_{d\_e2} \cdot Tg\mathrm{Prot}_i - \sum_j Tg\mathrm{Prot}X_i \cdot ProtX_j + \sum_k F_k(ProtX)$	In these equations, $mRNA_i$ accounts for messenger RNAs, $ProtX_i$ for proteins and protein complexes, and TF_i for transcription factors. Reaction rates account for TF-mediated synthesis of messenger RNAs, degradation of messenger RNAs, miRNA-mediated regulation of target genes, protein synthesis and degradation, and protein-protein interactions. Rate equations are constructed using mass action kinetics, Hill equations, power-law terms, and other formalisms. In addition, parameters k_x account for the rate constants

responses in the mRNA and protein expression levels of the target gene are affected by parameter settings, which are thought as being target gene specific. Others have shown how to employ kinetic modeling to investigate the regulation of miRNA-regulated networks that are of biomedical interest. For example, Lee and collaborators (2010) developed a network model that characterized miR-204 as tumor suppressor miRNA and uncovered previously unknown connections between network topology and expression dynamics. Additionally, they validated 18 target genes related to tumor progression.

Logical Modeling of miRNA Regulatory Networks. When microRNAs are embedded in regulatory networks, the size limits the applicability of the kinetic modeling approach. Logical models represent a suitable alternative for larger networks. These can, for example, be regulatory networks involving dozens of transcription factors, miRNAs, and other types of functional molecules (e.g., repressors, coactivators, corepressors, ribonucleoproteins). The logical approach can be classified into Boolean logic and multivalued logic (▶ MicroRNA-embedding Regulation Networks, Logical Modeling). Boolean logic works on the basis of binary categories. For instance, a miRNA can only be treated as absent or present, while a transcription factor as inactive or active. Their time evolution is restricted to transitions between two generic states encoded by 0 and 1. Multivalued logic instead generalizes the Boolean approach and allows the use of multiple ordinal categories. In this way, the effective inhibitory level of a miRNA can in multivalued logic be described by several stages, for example, null, low, high, and maximum inhibition, while the expression of the target might be categorized into silenced, low, high, and maximal expression.

Graph Analysis of miRNA Networks. MicroRNA-mRNA networks can be described by using graphs that capture the processes by which mature miRNAs control the translation of target mRNAs. In these graphs, elements involved in the network are represented as nodes. Edges that connect two nodes account for a relationship (interaction) between them, for example, the repression of a mRNA by a given miRNA (▶ MicroRNA-mRNA Regulation Networks). Using this approach, miRNA-mRNA regulation networks have proven to display topological properties like fat-tail degree distribution, abundance of cybernetic motifs, and modular structures.

Cross-References

- ▶ MicroRNA Biogenesis, Regulation
- ▶ MicroRNA Clusters
- ▶ MicroRNA Families, Cancer Progression
- ▶ MicroRNA Gene Prediction
- ▶ MicroRNA Regulation, Feedback Loop
- ▶ MicroRNA Regulation, Feed-Forward Loops
- ▶ MicroRNA Regulation, Signaling Pathways
- ▶ MicroRNA Target Prediction

- ▶ MicroRNA Web Resources
- ▶ microRNA, Disease and Therapy
- ▶ MicroRNA, Seed
- ▶ MicroRNA-embedding Regulation Networks, Logical Modeling
- ▶ MicroRNA-mRNA Regulation Networks
- ▶ MiRBase
- ▶ Non-coding RNA, Prediction
- ▶ RNA-seq
- ▶ Target Cleavage
- ▶ Target Hub
- ▶ Target Identification, microRNA
- ▶ Target Site

References

Aguda BD et al (2008) MicroRNA regulation of a cancer network: consequences of the feedback loops involving miR-17-92, E2F, and Myc. Proc Natl Acad Sci USA 105(50):19678–19683

Ambros V (2004) The functions of animal microRNAs. Nature 431(7006):350–355

Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116:281–297

Carrington JC, Ambros V (2003) Role of microRNAs in plant and animal development. Science (New York, NY) 301(5631):336–338

Enright AJ, Griffiths-Jones S (2008) MiRBase: a database of microRNA sequences, targets and nomenclature. In: Appasani K (ed) MicroRNAs: from basic science to disease biology. Cambridge University Press, Cambridge, pp 157–171

Filipowicz W, Bhattacharyya SN, Sonenberg N (2008) Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? Nat Rev Genet 9:102–114

Garzon R, Calin GA, Croce CM (2009) MicroRNAs in cancer. Annu Rev Med 60:167–179

Hurst DR, Edmonds MD, Welch DR (2009) Metastamir: the field of metastasis-regulatory microRNA is spreading. Cancer Res 69(19):7495–7498

Inui M, Martello G, Piccolo S (2010) MicroRNA control of signal transduction. Nat Rev Mol Cell Biol 11:252–263

Iorio MV et al (2005) MicroRNA gene expression deregulation in human breast cancer. Cancer Res 65(16):7065–7070

Lai X, Schmitz U, Gupta S, Bhattacharya A, Kunz M, Wolkenhauer O, Vera J (2012) Computational analysis of target hub gene repression regulated by multiple and cooperative miRNAs. Nucleic Acids Res 40:8818–8834

Lee RC, Feinbaum RL, Ambros V (1993) The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell 75:843–854

Lee Y et al (2010) Network modeling identifies molecular functions targeted by miR-204 to suppress head and neck tumor metastasis. PLoS Comput Biol 6(4):p.e1000730

Levine E, Ben Jacob E, Levine H (2007) Target-specific and global effectors in gene regulation by MicroRNA. Biophys J 93(11):L52–L54

Mendes ND, Freitas AT, Sagot MF (2009) Current tools for the identification of miRNA genes and their targets. Nucleic Acids Res 37(8):2419–2433

Osella M et al (2011) The role of incoherent microRNA-mediated feedforward loops in noise buffering. PLoS Comput Biol 7(3):p.e1001101

Shalgi R et al (2007) Global and local architecture of the mammalian microRNA-transcription factor regulatory network. PLoS Comput Biol 3(7):p.e131

Takamizawa J et al (2004) Reduced expression of the let-7 microRNAs in human lung cancers in association with shortened postoperative survival. Cancer Res 64(11):3753–3756

Tsang J, Zhu J, van Oudenaarden A (2007) MicroRNA-mediated feedback and feedforward loops are recurrent network motifs in mammals. Mol Cell 26(5):753–767

Computational Modeling

▶ Modeling Formalisms, Lymphocyte Dynamics and Repertoires

Computational Modeling Languages

▶ Cell Cycle Modeling, Process Algebra

Computational Models of Mechanisms

▶ Mechanism, Dynamic

Computer Modeling

▶ Lymphocyte Dynamics and Repertoires, Modeling

Computer Simulations

▶ Partial Differntial Equations, Numerical Methods and Simulations

Computer-Based Patient Records (CPR)

▶ Electronic Health Records

Concavity Tree for Protein–Protein Interaction Analysis

Virginio Cantoni[1,2], Alessandro Gaggia[1], Riccardo Gatti[1] and Luca Lombardi[1]
[1]Department of Computer Engineering and Systems Science, University of Pavia, Pavia, Italy
[2]Computational Biology, KTH Royal Institute of Technology, Stockholm, Sweden

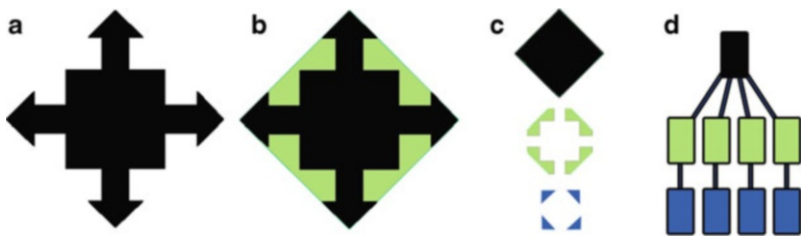
Definition

First introduced by Sklansky (1972), further researched by Bachelor (1979, 1980) and Borgefors and Sanniti di Baja (1992), a concavity tree is a data structure used for describing non-convex two-dimensional shapes. It is a rooted tree in which the root is a simple characterization of the whole object boundary, i.e., its ▶ convex hull. The next level describes the set of objects obtained by subtracting the object from the convex hull. All deeper levels are elaborated in the same way, iteratively. A node that represents a convex shape corresponds to a leaf in the tree, so it does not have any child.

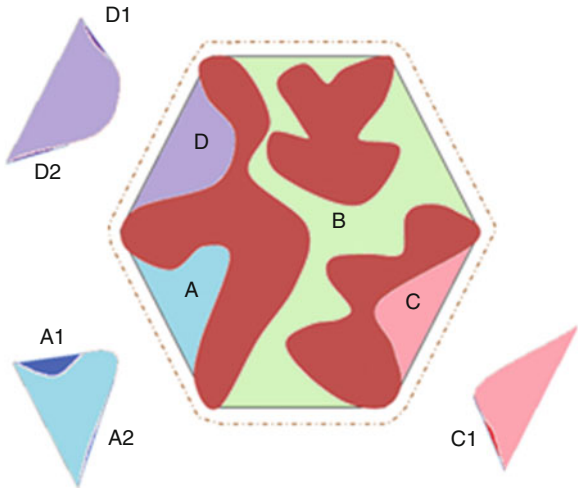
Figure 1 shows an example of a shape (a), its convex hull, concavities, and meta-concavities (b), and its corresponding concavity tree (c). The shape generates five concavities as reflected in level one of the tree. The four leaf nodes in level one correspond to the highlighted triangular concavities shown in (d), whereas the non-leaf node corresponds to the concavity shown in (e). Similarly, the nodes in levels two and three correspond to the meta-concavities highlighted in (f) and (g), respectively. Typically, each node in a concavity tree stores information pertinent to the part of the object the node is describing (a feature vector for example), in addition to tree metadata (like the level of the node; height, number of nodes, and number of leaves in the subtree rooted at the node, etc.).

Characteristics

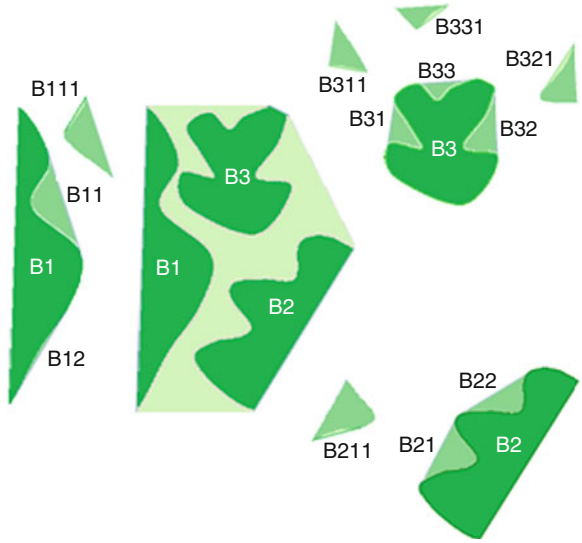
One of the most successful approaches for protein shape analysis and description is the structural one. A complex shape can be segmented into its component (e.g., a pockets set), and each pocket can be subsequently decomposed into simpler regions, and the complete description is given in terms of the



Concavity Tree for Protein–Protein Interaction Analysis, Fig. 1 An object (**a**), its convex hull with green concavities (**b**), the actual shape of each node (**c**) and the corresponding three levels concavity tree (**d**)



Concavity Tree for Protein–Protein Interaction Analysis, Fig. 2 A 2D representation example with a tunnel and three pockets in a section composed of three connected components (in *brown*). The closed curve in *black* corresponds to the first level convex hull, and the border in *brown dotted-dashed line* embodies the area under analysis. Part of the second level with three meta-concavities (A, B, C) is shown; in evidence also five third level termination-node components (A1, A2, C1, D1, and D2)



Concavity Tree for Protein–Protein Interaction Analysis, Fig. 3 Continuing the 2D representation example of Fig. 2, the details of the tunnel B component (*light green*) are shown. The second level is composed of the three meta-concavities (B1, B2, and B3). Each one has concavities at the third level: B1 has a termination-node concavity (B12) and a second component B11 which maintains a meta-concavity at the fourth level B111; B2 has a termination-node concavity (B22) and a second component B21 which maintains a meta concavity at the fourth level B211; finally, B3 has three node concavities (B31, B32, B33) each one maintaining a meta-concavity node-component B311, B321, and B331 respectively, at the fourth level

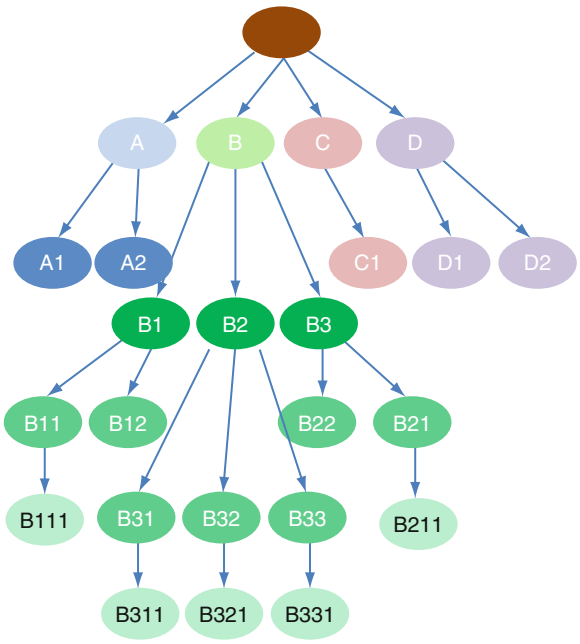
region's features and their spatial relationship. This process can be executed recursively. In this way, a sequence of approximations is built, and, at each stage, this structural hierarchical representation can be effectively applied for analyzing and comparing complex shapes. This approach is particularly fruitful in proteomics (Cantoni et al. 2010), in which the morphology plays a fundamental role, for example to study protein–protein interaction in which also a geometrical congruence for a normally extended part of the molecule surfaces is required.

A hierarchical refinement of the morphological analysis of extended regions is performed to describe further details. The Convex Hull of each region at every scale level is analyzed by the same process as that applied to the whole initial shape, going through concavity and meta-concavity. The process continues until all regions of the last level are convex, or until a level of detail sufficiently fine for the purpose is reached.

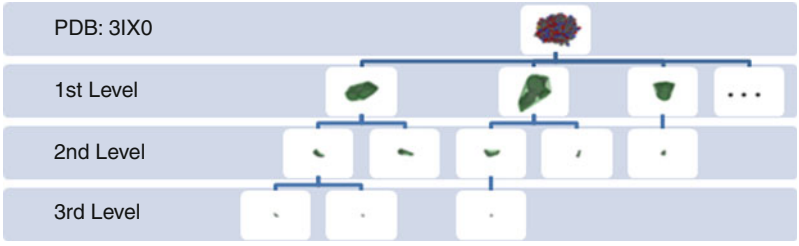
The final result is a hierarchical structure: the (meta) concavity tree. At each level, the concavities can be analyzed and described on the basis of a set of features evaluated at each node. Obviously, the features defined for concavities can also be computed for the meta-concavities of the same levels.

In Fig. 2 the concavities (three "pockets" and one "tunnel") and second-level meta-concavities of a 2D example are shown. Figure 3 shows concavities and meta-concavities of level two, three, and four for the tunnel of level one. The corresponding concavity tree is shown in Fig. 4. As an example of the proposed data description, Fig. 5 is

a partial representation of the concavity tree for the three main pockets of the Crystal Structure of uncomplexed HIV_1 Protease subtype A (PDB ID3IX0). Note that, each node can be described quantitatively through geometrical and biochemical parameters such as skewness, kurtosis, mouth aperture, surface to volume ratio, travel depth, etc.



Concavity Tree for Protein–Protein Interaction Analysis, Fig. 4 Representation of the concavity tree for the 2D section of Fig. 2 and Fig. 3. Note that each node contains the information of the feature vector previously presented



Concavity Tree for Protein–Protein Interaction Analysis, Fig. 5 The representation of the concavity tree for the Crystal Structure of uncomplexed HIV_1 Protease Subtype A (PDB ID 3IX0)

References

Batchelor BG (1979) Computers and digital techniques. J IEEE 2(4):157–168

Batchelor BG (1980) Shape descriptors for labeling concavity trees. J Cybernetics 10:233–237

Borgefors G, Sanniti di Baja G (1992) Methods for hierarchical analysis of concavities.In: Proceedings of the international conference on pattern recognition, vol 3. pp 171–175

Cantoni V, Gatti R, Lombardi L (2011) Geometrical constraints for ligand positioning. In: Proceedings of the first international conference on bioinformatics. pp 211–216

Sklansky J (1972) Measuring concavity on a rectangular mosaic. IEEE Trans Comput C-21:1355–1364

Concentration Control Coefficient

Emma Saavedra and Rafael Moreno-Sánchez
Department of Biochemistry, National Institute for Cardiology "Ignacio Chávez", Mexico City, Mexico

Definition

It is the degree of control that each enzyme exerts on the concentration of the pathway intermediaries. It is written as C^X_{ai} where X is any pathway metabolite and a is the activity of each pathway enzyme i.

Cross-References

▶ Metabolic Control Theory

Concentration Graph Model

▶ Graphical Gaussian Model

Concept Extraction Task

▶ Template Filling, Text Mining

Conditional Distribution

Lin Wang
School of Computer Science and Information Engineering, Tianjin University of Science and Technology, Tianjin, China

Synonyms

Conditional probability distribution; Conditional probability density function; Conditional probability mass function

Definition

Letting **X** and **Y** be two random variables over the same sample space **S**, the conditional probability distribution of **Y** given **X** is the probability distribution of **Y** when **X** is known to be a particular value. For discrete random variables, the conditional distribution of **Y** given the value **x** of **X** can be written as the following formula:

$$p_Y(y|X=x) = P(Y=y|X=x) = \frac{P(X=x \cap Y=y)}{P(X=x)}, \quad (1)$$

For continuous random variables, the conditional distribution of **Y** given the value **x** of **X** can be written as

$$f_Y(y|X=x) = \frac{f_{X,Y}(x,y)}{f_X(x)}, \quad (2)$$

where $f_{X,Y}(x,y)$ gives the joint density of X and Y, while $f_X(x)$ gives the density of X.

Cross-References

▶ Causal Relationship
▶ Conditional Independence

References

John R (1999) Mathematical statistics and data analysis. Duxbury, Belmont

Conditional Independence

Katsuhisa Horimoto
Computational Biology Research Center, National Institute of Advanced Industrial Science and Technology, Koto-ku, Tokyo, Japan

Synonyms

Bayes rule; Causal relationship; Conditional distribution; Correlation relationship

Definition

In probability theory, two events A and B are conditionally independent given a third event C, if the occurrence or nonoccurrence of A and the occurrence or nonoccurrence of B are independent events in their conditional probability distribution given C. In the standard notation of probability theory, A and B are conditionally independent given C if and only if

$$Pr(A \cap B|C) = Pr(A|C)Pr(B|C)$$

or equivalently,

$$Pr(A|B \cap C) = Pr(A|C)$$

In other words, A and B are conditionally independent if and only if, given knowledge of whether C occurs, knowledge of whether A occurs provides no information on the likelihood of B occurring, and knowledge of whether B occurs provides no information on the likelihood of A occurring. In statistics, two random variables X and Y are conditionally independent given a third random variable Z, if and only if they are independent in their conditional probability distribution given Z. This is generally written:

$$X \amalg Y|Z$$

and is read "X is independent of Y, given Z."

Cross-References

- ▶ Bayes Rule
- ▶ Causal Relationship

Conditional Independency

- ▶ Bayesian Network Model

Conditional Probability Density Function

- ▶ Conditional Distribution

Conditional Probability Distribution

- ▶ Conditional Distribution

Conditional Probability Mass Function

- ▶ Conditional Distribution

Confidence Intervals

Andreas Raue[1] and Jens Timmer[1,2,3]
[1]Institute for Physics, University of Freiburg, Freiburg, Germany
[2]BIOSS Centre for Biological Signalling Studies and Freiburg Institute for Advanced Studies (FRIAS), Freiburg, Germany
[3]Department of Clinical and Experimental Medicine, Linköping University, Linköping, Sweden

Definition

Often, parameter values are unknown and have to be estimated from experimental data. It is important not to rely on the mere estimated values and the predictions that correspond to these values. It is necessary to consider the uncertainties in the parameter estimation procedure: from measurement uncertainties to parameter uncertainties and possibly non-▶ identifiability, to uncertainties in the model predictions. In ▶ maximum likelihood estimation, uncertainties in the parameter estimates are usually described by confidence intervals. A confidence interval $[\sigma_i^-, \sigma_i^+]$ of a parameter estimate $\hat{\theta}_i$ to a confidence level α signifies that the true value θ_i^* is expected to be inside this interval with probability of α. For nonlinear models, confidence intervals can be defined by a threshold Δ_α in the likelihood. This threshold defines a confidence region

$$\left\{\theta|\chi^2(\theta) - \chi^2(\hat{\theta}) < \Delta_\alpha\right\} \quad \text{with} \quad \Delta_\alpha = Q\left(\chi^2_{df}, 1-\alpha\right) \tag{1}$$

whose borders represent likelihood-based confidence intervals (Meeker and Escobar 1995). $\chi^2(\theta)$ is the weighted sum of squared residuals and the threshold Δ_α is the $1-\alpha$ quantile of the χ^2_{df}– distribution. The choice of df yields confidence intervals that hold jointly for df number of parameters (Press et al. 1990); often $df = 1$ is desired. For high-dimensional models, the profile likelihood can be evaluated to derive likelihood-based confidence intervals; see in ▶ structural and practical identifiability analysis.

Cross-References

- ▶ Identifiability
- ▶ Maximum Likelihood Estimation
- ▶ Structural and Practical Identifiability Analysis

References

Meeker WQ, Escobar LA (1995) Teaching about approximate confidence regions based on maximum likelihood estimation. The Am Stat 49(l):48–53

Press WH, Teukolsky SLA, Flannery BNP, Vetterling WMT (1990) Numerical recipes: FORTRAN. Cambridge University Press, Cambridge

Configuration Type

- ▶ IMGT-ONTOLOGY, ConfigurationType

Confocal and Multiphoton Microscopy

Xiaohua Wu
Department of Pediatrics, Herman B Wells Center for Pediatric Research, Indiana University School of Medicine, Indianapolis, IN, USA

Definition

Confocal and multiphoton microscopy can achieve up to diffraction-limited resolution when imaging virtual tissue sections in intact tissue volumes, which is an aspect of these methods referred to as "tissue sectioning" or "optical sectioning." In particular, laser scanning confocal microscopy scans a focused laser beam inside the specimen and uses a pinhole to reject photons that arrive to the detector from out-of-focus areas.

Cross-References

- ▶ Spectroscopy and Spectromicroscopy

Conformational Epitope

- ▶ Discontinuous Epitope

Conjugation Reactions

- ▶ Phase II Enzymes

Connectionist Model

- ▶ Linear Additive Network Model

Connective Tissue

- ▶ Stroma

Connectivity Theorem

Emma Saavedra and Rafael Moreno-Sánchez
Department of Biochemistry, National Institute for Cardiology "Ignacio Chávez", Mexico City, Mexico

Definition

It links the kinetic properties of a particular enzyme (elasticity coefficients) with its ability to control the pathway flux (flux control coefficient). The sum of the products of the elasticity coefficient and the flux control coefficient of each of the pathway enzyme has to add up a value of zero.

Cross-References

- ▶ Metabolic Control Theory

Consensus Sequence

Jianhua Ruan
Department of Computer Science, University of Texas at San Antonio, San Antonio, TX, USA

Definition

In transcriptional regulation studies, a consensus sequence is a representation of a transcription factor-binding motif. It is obtained from the multiple alignment of a collection of binding sites recognized by the transcription factor, and depicts which nucleotides are most abundant in the alignment at each position. For example, consider the following DNA sequence, AccATGg. An upper-case letter means that the transcription factor has a very strong preference of the corresponding nucleotide at that position, while a lower-case letter represents the most abundant nucleotide at those positions. When multiple nucleotides are equally likely to appear, IUPAC nucleotide codes can also be used, for example, ATNNCY, where N stands for any base, and Y represents any pyrimidine.

Conservation Analysis

Sang Yup Lee
Department of Chemical and Biomolecular Engineering and Department of Bio and Brain Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea

Synonyms

Conserved mass analysis; Conserved moiety analysis; Stoichiometric mass balance analysis

Definition

Conserved moieties are molecular subgroups that are conserved throughout the course of a network evolution with respect to time due to stoichiometric constraints (Sauro and Ingalls 2004; Vallabhajosyula et al. 2006). Examples are the total amount of atomic elements (mass conservation) and the conservation of metabolites; the sum of ATP, ADP, and AMP remains constant with time if the synthesis of these nucleotides is not part of the network. A conservation analysis focuses on the identification of conserved moieties in networks and involves linear algebra methods (Sauro and Ingalls 2004; Vallabhajosyula et al. 2006).

References

Sauro HM, Ingalls B (2004) Conservation analysis in biochemical networks: computational issues for software writers. Biophys Chem 109(1):1–15
Vallabhajosyula RR, Chickarmane V, Sauro HM (2006) Conservation analysis of large biochemical networks. Bioinformatics 22(3):346–353

Conserved Mass Analysis

► Conservation Analysis

Conserved Moiety Analysis

► Conservation Analysis

Constant (C) Domain

Marie-Paule Lefranc
Laboratoire d'ImmunoGénétique Moléculaire, Institut de Génétique Humaine UPR 1142, Université Montpellier 2, Montpellier, France

Synonyms

C domain; C type domain; Constant domain

Constant (C) Domain, Table 1 C domain description per receptor type and chain type

C domain	Receptor type	C domain description per chain type	
C-DOMAIN	IG	CH1	Several CH[a] per IG-Heavy chain (ex: 4 for IG-Heavy-Mu)
		CH2	
		CH3	
		CH4	
		C-KAPPA[b]	
		C-LAMBDA[b]	
	TR	C-ALPHA	
		C-BETA	
		C-GAMMA	
		C-DELTA	
C-LIKE-DOMAIN	IgSF other than IG and TR	C-LIKE-DOMAIN	One or several domain(s) per chain

[a]CH: C-DOMAIN of IG-Heavy chain
[b]If the IG-Light chain is not specified, C-KAPPA and C-LAMBDA are described as CL (C-DOMAIN of IG-Light chain)

Definition

The Constant (C) domain is a type of structural unit (domain) that, with the ▶ variable (V) domain, characterizes a protein chain belonging to the ▶ immunoglobulin superfamily (IgSF) (Duprat et al. 2004; Lefranc et al. 2005). The C domain comprises the C-DOMAIN of the immunoglobulins (IG) or antibodies and T cell receptors (TR), and the C-LIKE-DOMAIN of the IgSF proteins other than IG and TR. The C domain description per receptor type and chain type, based on the ▶ IMGT-ONTOLOGY concepts (▶ Chain Type, ▶ Domain Type), is shown in Table 1. IMGT® labels are in capital letters.

A C domain (C-DOMAIN or C-LIKE-DOMAIN) is usually encoded by one exon of a gene. Although the C domain sequences may be very diverse, the structure is well conserved. The C domain (C-DOMAIN or C-LIKE-DOMAIN) is made of seven antiparallel beta strands (A, B, C, D, E, F, and G) linked by beta turns (AB, DE, and EF), a transversal strand CD, and loops (BC and FG), forming a sandwich of two sheets (four strands on one sheet and three strands on the other sheet).

According to the ▶ IMGT unique numbering (Lefranc et al. 2005), the four conserved amino acids of a C domain always have the same position, cysteine 23 (1st-CYS), tryptophan 41 (CONSERVED-TRP), conserved hydrophobic amino acid 89, and cysteine 104 (2nd-CYS). The amino acid at position 118, although not conserved, is highlighted in the ▶ IMGT Collier de Perles as it is one of the two anchors of the FG loop.

Analysis of C domain sequences can be performed by tools of IMGT®, the international ImMunoGeneTics information system® (http://www.imgt.org) (▶ IMGT® Information System):

- Nucleotide sequences of IG and TR C-DOMAIN by IMGT/V-QUEST
- Amino acid sequences of C-DOMAIN and C-LIKE-DOMAIN by IMGT/DomainGapAlign (Ehrenmann et al. 2010)

These tools align the user sequences with the closest C domains of the IMGT reference directory, create gaps according to the ▶ IMGT unique numbering for C domain (Lefranc et al. 2005), delimit the strands and loops, highlight differences with the closest reference (s), and generate the ▶ IMGT Collier de Perles.

Analysis of C domain three-dimensional (3D) structures and interactions is available in the 3D database (IMGT/3Dstructure-DB) of the ▶ IMGT® information system (Ehrenmann et al. 2010).

Allotypes in humans are antigenic determinants on C-DOMAIN of the IG (CH of IG-Heavy gamma and alpha chains, and C-KAPPA of IG-Light-Kappa chains) (Lefranc and Lefranc in press). They comprise:

- the Gm allotypes ('gamma markers') on CH of IG-Heavy-Gamma chains,

- the Am allotypes ('alpha markers') on CH of IG-Heavy-Alpha chains,
- the Km allotypes (or 'kappa markers') on C-KAPPA of IG-Light-Kappa chains (Lefranc and Lefranc in press) (IMGT Repertoire, http://www.imgt.org)

They have extensively been analysed in population genetics and molecular evolution of the IG. They presently regained a lot of attention in antibody engineering and antibody humanization, in the analysis of the immunogenicity of therapeutic antibodies. The corresponding amino acid changes are identified and described according to the IMGT unique numbering for C-DOMAIN (Lefranc and Lefranc in press).

Cross-References

- ▶ Chain Type
- ▶ Domain Type
- ▶ IMGT Collier de Perles
- ▶ IMGT Unique Numbering
- ▶ IMGT® Information System
- ▶ IMGT-ONTOLOGY
- ▶ Immunoglobulin Superfamily (IgSF)
- ▶ Variable (V) Domain

References

Duprat E, Kaas Q, Garelle V, Lefranc G, Lefranc M-P (2004) IMGT standardization for alleles and mutations of the V-LIKE-DOMAINs and C-LIKE-DOMAINs of the immunoglobulin superfamily. In: Pandalai SG (ed) Recent Research Developments in Human Genetics. Research Signpost, Trivandrum, vol 2, pp 111–136, chapter 6

Ehrenmann F, Kaas Q, Lefranc M-P (2010) IMGT/3Dstructure-DB and IMGT/DomainGapAlign: a database and a tool for immunoglobulins or antibodies, T cell receptors, MHC, IgSF and MhcSF. Nucleic Acids Res 38:D301–D317

Lefranc M-P, Pommié C, Kaas Q, Duprat E, Bosc N, Guiraudou D, Jean C, Ruiz M, Da Piedade I, Rouard M, Foulquier E, Thouvenin V, Lefranc G (2005) IMGT unique numbering for immunoglobulin and T cell receptor constant domains and Ig superfamily C-like domains. Dev Comp Immunol 29:185–203

Lefranc M-P, Lefranc G (in press) Human Gm, Km and Am allotypes and their molecular characterization: a remarkable demonstration of polymorphism. In: Tait B, Christiansen F (eds) Methods in Molecular Biology – Immunogenetics, chapter 34

Constant (C) Gene

Marie-Paule Lefranc
Laboratoire d'ImmunoGénétique Moléculaire, Institut de Génétique Humaine UPR 1142, Université Montpellier 2, Montpellier, France

C

Synonyms

C gene; Constant; Constant gene

Definition

The constant (C) gene, or "*constant*" is a ▶ leafconcept of the "▶ GeneType" concept of identification (generated from the ▶ IDENTIFICATION axiom) of ▶ IMGT-ONTOLOGY, the global reference in ▶ immunogenetics and ▶ immunoinformatics (Giudicelli and Lefranc 1999; Lefranc et al. 2004, 2005, 2008; Duroux et al. 2008), built by IMGT®, the international ImMunoGeneTics information system® (http://www.imgt.org) (▶ IMGT® Information System). "*Constant*" identifies a gene that codes the constant region of an immunoglobulin (IG) or antibody or of a T cell receptor (TR) chain (▶ Chain Type).

It is one of the four leafconcepts that are characteristics of the IG and TR loci (Lefranc and Lefranc 2001a, b), the other three being "variable" (V), "diversity" (D), and "joining" (J) (▶ Variable (V) Gene, ▶ Diversity (D) Gene, ▶ Joining (J) Gene).

An IG or TR constant (C) gene in contrast to V, D, and J genes does not rearrange itself and has a configuration identified as "undefined" (▶ Configuration Type). A C gene is characterized by an acceptor splice in 5′ (the splicing occurring with the 3′ donor splice of a J gene).

The C region encoded by a C gene comprises one or several domains (▶ Constant (C) Domain) depending on the chain type (▶ Chain Type).

Cross-References

- ▶ Chain Type
- ▶ Configuration Type

- ▶ Constant (C) Domain
- ▶ Diversity (D) Gene
- ▶ Gene Type
- ▶ IMGT® Information System
- ▶ IMGT-ONTOLOGY
- ▶ IMGT-ONTOLOGY, IDENTIFICATION Axiom
- ▶ IMGT-ONTOLOGY, Leafconcept
- ▶ Immunogenetics
- ▶ Immunoinformatics
- ▶ Joining (J) Gene
- ▶ Variable (V) Gene

References

Duroux P, Kaas Q, Brochet X, Lane J, Ginestoux C, Lefranc M-P, Giudicelli V (2008) IMGT-Kaleidoscope, the formal IMGT-ONTOLOGY paradigm. Biochimie 90:570–583

Giudicelli V, Lefranc M-P (1999) Ontology for immunogenetics: the IMGT-ONTOLOGY. Bioinformatics 12:1047–1054

Lefranc M-P, Lefranc G (2001a) The immunoglobulin FactsBook. Academic Press, London, pp 1–458

Lefranc M-P, Lefranc G (2001b) The T cell receptor FactsBook. Academic Press, London, pp 1–398

Lefranc M-P, Giudicelli V, Ginestoux C, Bosc N, Folch G, Guiraudou D, Jabado-Michaloud J, Magris S, Scaviner D, Thouvenin V, Combres K, Girod D, Jeanjean S, Protat C, Yousfi Monod M, Duprat E, Kaas Q, Pommié C, Chaume D, Lefranc G (2004) IMGT-ONTOLOGY for immunogenetics and immunoinformatics. In Silico Biol 4:17–29

Lefranc M-P, Clément O, Kaas Q, Duprat E, Chastellan P, Coelho I, Combres K, Ginestoux C, Giudicelli V, Chaume D, Lefranc G (2005) IMGT-Choreography for immunogenetics and immunoinformatics. In Silico Biol 5:45–60

Lefranc M-P, Giudicelli V, Regnier L, Duroux P (2008) IMGT, a system and an ontology that bridge biological and computational spheres in bioinformatics. Brief Bioinform 9:263–275

Constant Domain

- ▶ Constant (C) Domain

Constant Gene

- ▶ Constant (C) Gene

Constant

- ▶ Constant (C) Gene

Constraint

Jon Umerez and Matteo Mossio
IAS-Research Philosophy of Biology Group, Department of Logic and Philosophy of Science, University of the Basque Country (UPV/EHU), Donostia – San Sebastian, Spain

Synonyms

(Approx.) boundary condition; Control

Definition

Constraint refers to a reduction of the degrees of freedom of the elements of a system exerted by some collection of elements, or a limitation or bias on the variability or possibilities of change in the kind of such elements.

Although the term has several meanings in diverse scientific fields, the idea of a constraint is usually employed in relation to conceptualizations in terms of levels or ▶ hierarchies. Some general features of constraints such understood are the following: constraints do not interact with the elements they influence and their dynamics; they arise from dynamics at different levels of ▶ organization; constraint relations are always asymmetric, and may give rise to new phenomena.

In the tradition of the theories of emergent evolution of the early twentieth century and, even clearer, in the theory of levels of integration of the organicist tradition of the 1930s, higher levels are understood as arising from lower-level elements or processes, whose laws all obey, but concurrently exerting some specific influence on those very elements or processes (see Blitz 1992). Later on, in most approaches to hierarchy theory, the very concept of constraint is profusely used to account for the specificity of nontrivial inter-level

relation, where lower level and upper level act upon each other but in different ways (see, e.g., Allen and Starr 1982; Salthe 1985). Even the introduction of some specific senses of constraint in evolutionary biology (as developmental constraints and historical constraints) is due to the attempt to accommodate different explanations, stemming from diverse factors or forces, and is a potential alternative, if not rival, to adaptive selection, in an unavoidably multilevel explanatory construction.

Characteristics

The development of more specific characterizations of constraint (or alike) are attempts to spell out the interaction among levels of organization by deriving them from epistemologically legitimate concepts grounded in the physical sciences or in their explanatory armory.

The difficult issue of interlevel relation is rendered more concrete by Polanyi's (1968) account of organisms (and machines) as dual control systems, which relies on an indistinct concept of boundary conditions. Polanyi deliberately employs the machine example to introduce his idea of boundary conditions "harnessing the laws of nature" that govern matter and the forces acting on it. In a machine, the harnessing is exerted with a goal, which makes easier to describe the dual control: the design of the machine, by which the machine does what it is intended for, and the laws of physics and chemistry, that the components of the machine and the machine itself must obey.

In the organism "its structure serves as a boundary condition harnessing the physical-chemical processes by which its organs perform their functions" (Polanyi 1968, 1308). The machine analogy goes on to compare morphogenesis, as the process that develops that structure, with the shaping of the machine. Yet, as Polanyi emphasizes, the analogy ends here: Although both are systems under dual control (unlike inanimate systems which may exhibit boundary conditions without being subject to dual control), organisms are not artificially designed. Therefore, the harnessing principle in organisms has to be autonomous (▶ Autonomy), task that Polanyi attributes to the informational account of heredity.

Is precisely this the challenge that Pattee (see, e.g., 1968, 1972) undertakes in his research on the origin of life and the nature of biological function and heredity, using the concept of constraint as an explanatory tool derived from Mechanics. He develops in several papers the concept and the relevant distinction between holonomic (▶ Constraint, Holonomic) (structural-like) and non-holonomic (▶ Constraint, Non-holonomic) (▶ functional-like) kinds. Pattee claims that, in order to explain hereditary storage, transmission, and the action of genetic instructions, we should understand them in terms of non-holonomic constraints acting in the context of a very specific kind of interlevel relation (semantic closure).

Following his description we may recall that, in physics, initial conditions and laws of movement provide, in principle, an exhaustive description of the possible behavior, future and past, of a mechanical system. No other conditions are needed. The choice of the coordinates of the system that specifies its space of configuration or space of states defines all the possible degrees of freedom in the system or all the possible trajectories of movement of its elements. These coordinates establish the variables of the movement equations of the system. In this context, constraints are those additional equations that are introduced as auxiliary conditions in order to define the specific mechanical system subject to calculation (see, e.g., Sommerfeld 1952). In other areas of Physics, constraints may be expressed in a more general form as boundary conditions. In Chemistry, constraints refer to the steady state of elementary particles (chemical bonds). When referred to dissipative systems, the concept acquires an even more specific sense as it becomes dynamical (an unstable macroscopic pattern that remains as long as there is energy contribution). Finally, its presence is patent in any form of biological regulation (starting from a membrane) and more controversial in its contribution to the understanding of evolutionary paths. In social and artificial systems, it is clearly manifested in the form of rules. In sum, constraints refer to certain conditions or rules additional to the laws of dynamics (that are taken as basic), that rule/govern the behavior of the elements and that arise from their aggregation.

Whereas natural laws are, in principle, inexorable and incorporeal, constraints are, by necessity, accidental or arbitrary, and require some distinct physical materialization (as molecules, membranes, or surfaces). Constraints are alternative descriptions of part

of the system. Namely, constraints cannot be expressed in the same language than the microscopic description of matter. In fact, what a constraint does is to selectively ignore microscopic degrees of freedom in order to obtain a simplification in the prediction or explanation of movement. The concept of constraint means a selective loss of detail or a predetermined rule about what is going to be ignored.

Therefore, for Physics constraining forces are, unavoidably, linked to a new hierarchical level of description. When in Physics a constraint equation is added to the equations of movement, we are always dealing with two languages at the same time. The language of the equation of movement relates the detailed trajectory or the state of the system with dynamical time, whereas the language of the constraint does not deal at all with the same kind of system, but with another situation in which the dynamical detail has been purposely ignored. In other words, constriction forces are not detailed forces of the individual particles but forces of the collections of particles or, sometimes, forces of simple units averaged out in time. In any case, some form of statistical averaging process has substituted microscopic detail. In Physics, then, in order to describe a constraint, the detailed dynamical description has to be abandoned. A constraint requires, therefore, an alternative description.

Another relevant contribution to naturalize the concept of constraint in the biological domain has been provided by theoretical biologist Stuart Kauffman, who has recently proposed the idea of the "Work-Constraint cycle" (Kauffman 2000, 2003). The Work-Constraint cycle is supposed to capture what Kauffman takes as a central feature of all biological organisms, namely, the fact of acting "on their own behalf" (Kauffman 2003, 1089). Whereas this idea appears to be in accordance with common intuition, Kauffman's scientific challenge consists in giving a naturalized and consistent account of it. The concept of a Work-Constraint cycle plays precisely this role. The main idea is to link the idea of action to that of "work," the latter being defined, following Atkins, as "constrained release of energy into relatively few degrees of freedom" (Kauffman 2003, 1094). In this definition, the concepts of work and constraint are related: work is constrained release of energy. This connection gives a way to interpret the slogan acting "on their own behalf." A system acts on its own behalf if it is able to use its work to regenerate at least some of the constraints that make work possible. When this occurs, a Work-Constraint cycle is realized. In physical terms, it requires very specific conditions to occur. Actually, the cycle is inevitably a thermodynamic irreversible process, which dissipates energy and requires couplings between exergonic (spontaneous, energy releasing) reactions and endergonic (non-spontaneous, energy requiring) ones, such that exergonic processes are constrained in a specific way to produce work that may be used to generate endergonic processes, which in turn generate those constraints canalizing exergonic processes. In Kauffman's terms, "Work begets constraints beget work" (Kauffman 2000).

In evolutionary biology, the concept of constraint has mainly been introduced as a challenge by developmental approaches regarding the scope of selection and the extent to which ▶ adaptation remains the main explanatory factor (see ▶ Explanation, Developmental). Besides the specific issue of connecting developmental with evolutionary accounts, the concept of constraint plays an important role, for instance, in the criticisms to adaptationism, in the discussion regarding the relevance of stasis and other macroevolutionary patterns, or in the morphologically oriented proposals of structuralism or ▶ complexity theories (see Orzack and Sober 2001).

In an already classic paper that may be considered to be an attempt to build a "consensus" position on the subject, a developmental constraint is defined as "a bias on the production of variant phenotypes or a limitation on phenotypic variability caused by the structure, character, composition, or dynamics of the developmental system" (Maynard Smith et al. 1985, 266). The origin of this bias or limitation may be attributed to various sources (materiality, genetic dynamics, evolutionary pathways, complexity regime, etc.) but what is agreed upon is that they have an impact in evolution. A distinction is also drawn between universal constraints, deriving from general laws of physics or from invariant properties of some material or complex systems, and local constraints, confined to particular taxa.

Amundson (1994), while accepting that constraint "implies some sort of restriction on variety or on change," claims that the key rests on the answer to

the question: "What is being constrained?" Thus, he identifies two possible answers depending on whether the explanandum is adaptation or form. We would therefore have to distinguish between understanding the effect of developmental principles on evolution as constraints on adaptation, that is, as restrictions imposed by embryology on the adaptive optimality of adult organisms on the potential of adaptation, or constraints on form, in the sense that in the morphospace there are morphologies that cannot be achieved by the process of development.

This distinction is orthogonal to the previous one between local and universal kinds of constraints since both of them may operate, irrespectively, on the prospects of reaching more optimal adaptations and on the scope of generation of organic form.

Sometimes, other kinds of constraints are mentioned in the context of contending evolutionary explanations, even if they are not generally accepted as specific constraints. For instance, historical constraints are often brought up to contrast with developmental ones though maintaining the same feature of imposing restriction on adaptation, but they are due to more "evolutionary" common factors such as contingency or accident creating some kind of bias. Ecological constraints can also be sometimes evoked.

Regarding these uses and besides those distinctions about how to understand the concept, there is another point worth mentioning, since it allows connecting the concept of developmental constraint with the more general idea defined in the beginning of this entry. According to Schwenk and Wagner (2003), this collective definition highlights the separation between constraint and selection by distinguishing the "generation of variation from the operation of selection on that variation" (p. 54). In this sense, the force of selection would assume the role of the general law upon which some rules are imposed and, here too, some potential degrees of freedom are reduced due to what Polanyi was calling boundary conditions and what Pattee called constraints in their more general approaches.

Acknowledgments Funding for this work was provided by grants GV/EJ IT505-10 from the Government of the Basque Country, and FFI2011-25665 from the Ministerio de Economía y Competitivad (MEC) and FEDER funds from the E.C.

Cross-References

- ▶ Adaptation
- ▶ Autonomy
- ▶ Complexity
- ▶ Constraint, Holonomic
- ▶ Constraint, Non-holonomic
- ▶ Explanation, Developmental
- ▶ Explanation, Evolutionary
- ▶ Explanation, Functional
- ▶ Functional
- ▶ Hierarchy
- ▶ Interlevel Causation
- ▶ Organization

References

Allen TFH, Starr TB (1982) Hierarchy. Perspectives for ecological complexity. The University of Chicago Press, Chicago

Amundson R (1994) Two concepts of constraint: adaptationism and the challenge from developmental biology. Philos Sci 61:556–578

Blitz D (1992) Emergent evolution. Qualitative novelty and the levels of reality. Kluwer, Dordrecht

Kauffman S (2000) Investigations. Oxford University Press, Oxford

Kauffman S (2003) Molecular autonomous agents. Phil Trans R Soc A 361:1089–1099

Maynard Smith J, Burian R, Kauffman S, Alberch P, Campbell J, Goodwin B, Lande R, Raup D, Wolpert L (1985) Developmental constraints and evolution: a perspective from the mountain lake conference on development and evolution. Q Rev Biol 60(3):265–286

Orzack SH, Sober E (eds) (2001) Adaptationism and optimality. Cambridge University Press, Cambridge

Pattee HH (1968) The physical basis of coding and reliability in biological evolution. In: Waddington CH (ed) Towards a theoretical biology, vol 1, Prolegomena. Edinburgh University Press, Edinburgh, pp 67–93

Pattee HH (1972) Laws and constraints, symbols and languages. In: Waddington CH (ed) Towards a theoretical biology, vol 4, Essays. Edinburgh University Press, Edinburgh, pp 248–258

Polanyi M (1968) Life's irreducible structure. Science 160:1308–1312

Salthe SN (1985) Evolving hierarchical systems. Their structure and representation. Columbia University Press, New York

Schwenk K, Wagner GP (2003) Constraint. In: Hall BK, Olson WM (eds) Keywords in evolutionary developmental biology. Harvard University Press, Cambridge, MA, pp 52–61

Sommerfeld A (1952) Mechanics, 2nd edn. Academic, New York

Constraint, Holonomic

Jon Umerez and Matteo Mossio
IAS-Research Philosophy of Biology Group, Department of Logic and Philosophy of Science, University of the Basque Country (UPV/EHU), Donostia – San Sebastian, Spain

Definition

Holonomic constraints are auxiliary conditions that limit permanently the number of degrees of freedom of a system. They are, then, fixed and passive structures that, in a sense, are independent of time (at least of the specific time frame, dynamics, of the system in particular). They establish some fixed relations among some coordinates of the system that are mathematically integrable with the fundamental equations of movement. They are integrable since they introduce no new temporal dimension. They just reduce degrees of freedom once and for good. They are mainly found in most physical-chemical and artificial mechanistic systems. Constraints whose equations do not contain time are called scleronomic.

Cross-References

► Constraint

Constraint, Non-holonomic

Jon Umerez and Matteo Mossio
IAS-Research Philosophy of Biology Group, Department of Logic and Philosophy of Science, University of the Basque Country (UPV/EHU), Donostia – San Sebastian, Spain

Definition

Non-holonomic constraints are variable auxiliary conditions that limit in time the number of degrees of freedom of the system. They are, then, dynamical structures that establish time-dependent relations among degrees of freedom. These correlations among coordinates are nonintegrable with the equations of movement, since they introduce a different temporal scale. Classically, these constraints do reduce more degrees of freedom in the dynamical movement of the system than the number of degrees of freedom that were necessary to define in the static description that specifies the initial conditions and the states or configuration space of the system. They are mainly found in biological and social systems, where processes of selection and classification are common, opening up the state space and increasing the variety of behaviors. Constraints whose equations contain time are called rheonomic.

Cross-References

► Constraint

Constraint-based Modeling

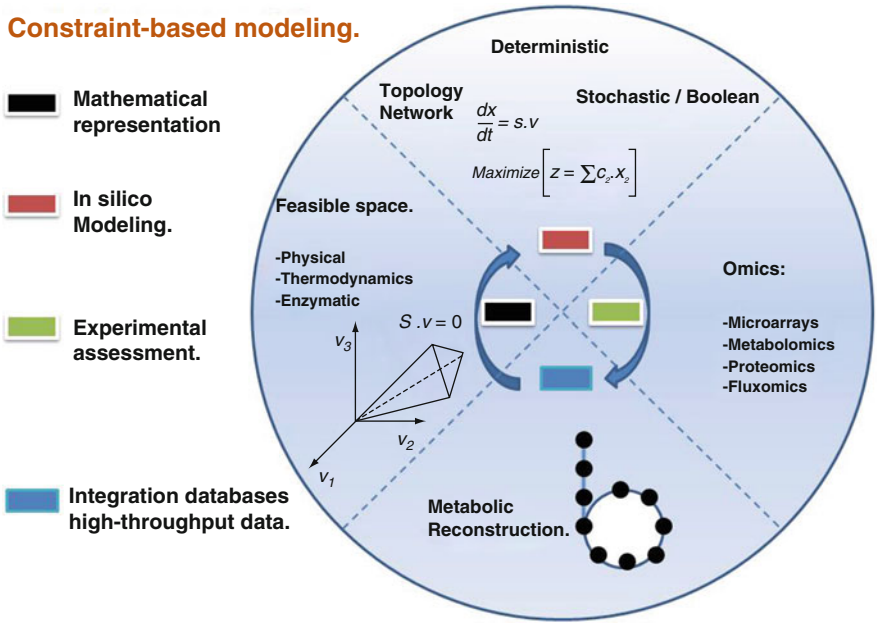
Osbaldo Resendis-Antonio
Center for Genomics Sciences-UNAM, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, Mexico

Synonyms

Genome-scale metabolic modeling

Definition

With the advent of high-throughput technology, there has been a growing need to develop computational frameworks that contribute to analyze these data for unveiling the biological principles underlying metabolic networks. Constraint-based modeling is a paradigm in systems biology that contributes to this latter aim by considering physical, enzymatic, and topological constraints underlying the phenotype in a metabolic network. In global terms, this method can be stratified into four steps: metabolic reconstruction of an organism, mathematical representation of the metabolic network, in silico analysis, and experimental



Constraint-based Modeling,
Fig. 1 *Constraint-based modeling*. Constraint-based modeling is a paradigm in systems biology which consists of four main steps. (1) genome scale metabolic reconstruction (region with *blue block*), (2) mathematical representation of the metabolic network (*black*), (3) in silico analysis (*red*), and (4) experimental assessment (*green*). A continuous feedback is needed to improve the in silico predictions, extend the metabolic reconstruction and design experiments for uncovering the organizing principles in metabolism

assessment of computational predictions. In the end, this approach supplies with a powerful scheme for: (1) integrating HT data, (2) exploring the metabolic phenotype in a metabolic reconstruction, and (3) designing experiments to assess biological hypothesis.

Characteristics

The advent of high-throughput technology has triggered the appearance of a significant number of databases that constitute important sources of data for understanding the mechanisms by which cells organize their biological activity at diverse biological levels. Given the high number of variables inherent to these databases, this enterprise requires computational methods that are capable of dealing with the heterogeneous and high dimensionality nature of high-throughput data in a coherent and systematic fashion. Constraint-based modeling serves as a model-concentric database that provides with a framework for describing and predicting metabolic phenotypes in microorganisms. Having selected a metabolic reconstruction in an organism, this computational approach involves the application of a series of constraints arising from the consideration of stoichiometric, thermodynamic, enzymatic capacities, regulatory and kinetic constraints when they are available. In general terms, this approach can be summarized by four steps: genome scale metabolic reconstruction, mathematical representation of the metabolic network, in silico analysis, and experimental assessment, see Fig. 1.

Metabolic Reconstruction

The current DNA sequence technologies and the bioinformatics tools available in the literature have facilitated the development of metabolic databases in a variety of organisms. In this contextual scheme, these databases can be used to survey how genes codified in the genome contribute to determine the innate metabolic capacities in microorganisms. These databases, such as KEGG (Kanehisa et al. 2008) and Biocyc (Caspi et al. 2010) to name a few, constitute a good base to construct a metabolic network by identifying its components and defining their interactions (Feist et al. 2009; Reed et al. 2006; Resendis-Antonio et al. 2007, 2010). Furthermore, in order to characterize the feasible metabolic capacities in the organism, this information usually is complemented with high-throughput data and a carefully genetic and metabolic literature review of the organism under study (Reed et al. 2006). During this curation process, it is important to verify

that all the reactions included in the reconstruction are such that they are charge and mass balanced, this issue is essential for ensuring proper in silico results and interpretations (Thiele and Palsson 2010).

Mathematical Representation of the Metabolic Network

The relation between the genotype and the phenotype is a central issue when studying microorganisms in systems and synthetic biology approaches (Palsson 2006). At a macroscopic level, the phenotype state in a cell is well characterized by the activity of metabolic pathways, which in turn, mirror the genetic and protein activity required for facing external environment. For this reason, the study of the metabolic capacities in a genome scale reconstruction becomes relevant for exploring the feasible phenotype space in an organism. To this end, the set of reactions that form the metabolic reconstruction are mathematically represented through the ▶ stoichiometric matrix, S. Thus, by assuming that metabolic concentrations are at steady state condition, the feasible metabolic space of flux distribution can be calculated by:

$$S \cdot v = 0 \tag{1}$$

where v represents a vector whose entries are the metabolic fluxes of the reactions included in the reconstruction. Hence, from a mathematical perspective, the problem associated to the feasible phenotype state is close related to solve the null space of the stoichiometric matrix (Palsson 2006).

In Silico Analysis

Physical, thermodynamical, and enzymatic constraints. Even though the feasible metabolic capacities in the metabolic network are quantitatively well defined through Eq. 1, the number of possibilities remains still huge and additional considerations are required to limit their spectrum. Thus, in order to limit even more this space, one imposes these main constraints when available:

- *Physical:* when we have information that certain reactions occur in well-defined compartmentalized organelle inside the cell, for instance, mitochondria or cytoplasm.
- *Thermodynamical:* when we have information related with the irreversibility or reversibility of certain reactions along the reconstruction. Other constraint belonging to this classification is related with the second law of thermodynamics, the latter being a fundamental law to regulate the direction of the biological process.
- *Enzymatic:* when one has information about the flux capacity that one enzyme can be able to carry out. For instance, uptake of oxygen is usually of 20 *nMol•gDW/Hr*.
- *Additional constraints* can be imposed to delimit the metabolic capacities, for instance, regulatory constraints when the transcriptional regulatory network is known in one organism (Covert et al. 2004).

Taking into account these constraints brings as a consequence a reduction of the feasible metabolic space, this point is an important step toward a more accurate modeling of genome scale metabolic reconstruction, see Fig. 1 (black region). A variety of approaches have been suggested to explore the phenotype capacities of a metabolic reconstruction and link their outputs to experimental and high-throughput data (Price et al. 2004). Among these approaches, ▶ flux balance analysis (FBA) is a paradigm in systems biology for exploring the metabolic phenotype in an organism and predicting its response under external perturbation (Orth et al. 2010). In this biased approach, the metabolic activity in a reconstruction is obtained, first, by defining a function that represents certain type of biological purpose – called objective function, Z, and then, identifying the metabolic flux distribution that maximizes it. Depending on the case of study, this objective function can represent a specific physiological state (as growth rate or bacterial nitrogen fixation) and it is calculated through linear optimization,i.e.,

$$\text{Maximize}\left[Z = \sum_{k=1} c_k \cdot X_k\right]$$

such that:

$$\sum S_{i,j} \cdot v_j = 0 \quad i = 1 \ldots .m$$

$$-\alpha_j \leq v_j \leq \beta_j \quad j = 1 \ldots n$$

where X_k is a metabolite participating in the objective function and c_k is a weight factor. Here m indicates the number of metabolites and n the number of

reactions along the reconstruction. Note that the latter two equations indicate that the solution is subject to a steady state condition and is constrained by thermodynamic and enzymatic limits. This approach has been successfully and extensively applied to explore the consequences that gene perturbations produce on phenotype in *archaea*, *eukarya*, and *bacteria*. The number of papers where FBA has contributed to understand the metabolic activity in microorganism has been rising recently; a representative set of applications in some organism can be reviewed in Resendis-Antonio et al. (2007, 2012), Chang et al. (2012) and Varum Mazumdar et al. (2009).

In parallel with this approach, a variety of methods have been proposed for extending the analysis of the metabolic phenotype inherent to a genome scale metabolic reconstruction (Price et al. 2004). The applicability of these methods ranges from sampling the entire phenotypes in the null space of the stoichiometric matrix to calculating the metabolic phenotype when knockout or gene deletion occurs in the metabolic reconstruction (Segre et al. 2009). In addition, there are some algorithms devoted to explore the relation between dynamic behavior of metabolites, topology of network, and biological functionality (Resendis-Antonio 2009; Jamshidi and Palsson 2008).

Experimental Assessment of In Silico Predictions

Constraint-based modeling lets us simulate real biological processes and explore the resulting phenotypes when changes in parameters occur due to genetic mutations or environmental perturbations. The structural organization of this approach is such that it lets us to move toward experimental assessment with high-throughput technology. Thus, transcriptome data, proteome, and metabolome profiles – and other *omics* – become valuable data for refining the model and assessing the biological hypothesis emerging from in silico analysis (Resendis-Antonio et al. 2012). This feedback, established between experiment design and computational evaluation, is a valuable and needed task for completing our understanding of how microorganisms organize and control their metabolic response at specific environments.

Conclusions

Constraint-based modeling of metabolic networks contributes to establish a systems biology platform capable of integrating high-throughput data and computational simulations. This integrative description at genome scale is useful for: (1) understanding the fundamental activity of metabolism, (2) predicting phenotype metabolic under internal or external perturbations, and (3) designing more informative experiments at various biological layers.

Cross-References

▶ Flux Balance Analysis
▶ Stoichiometric Matrix

References

Caspi R, Altman T, Dale JM, Dreher K, Fulcher CA, Gilham F, Kaipa P, Karthikeyan AS, Kothari A, Krummenacker M, Latendresse M, Mueller LA, Paley S, Popescu L, Pujar A, Shearer AG, Zhang P, Karp PD (2010) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res 38:D473–D479

Chang RL, Ghamsari L, Manichaikul A, Hom EFY, Balaji S, Fu W, Shen Y, Hao T, Palsson BO, Salehi-Ashtiani K, Papin JA (2012) Metabolic network reconstruction of Chlamydomonas offers insight into light-driven algal metabolisms. Mol Syst Biol 7:518

Covert MW, Knight EM, Reed JL, Herrgård MJ, Palsson BØ (2004) Integrating high-throughput and computational data elucidates bacterial networks. Nature 429(6987):92–96

Feist AM, Herrgard MJ, Thiele I, Reed JL, Palsson BO (2009) Reconstruction of biochemical networks in microbial organisms. Nat Rev Microbiol 7(2):129–143

Jamshidi N, Palsson BØ (2008) Formulating genome-scale kinetic models in the post-genome era. Mol Syst Biol 4:171

Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S, Tokimatsu T, Yamanishi Y (2008) KEGG for linking genomes to life and the environment. Nucleic Acids Res 36:D480–D484

Mazumdar V, Snitkin E, Amar S, Segre D (2009) Metabolic network model of a human oral pathogen. J Bacteriol 191(1):74–90

Orth JD, Thiele I, Palsson BØ (2010) What is flux balance analysis? Nat Biotechnol 28:245–248

Palsson BØ (2006) Systems biology: properties of reconstructed networks. Cambridge University Press, Cambridge/New York

Price ND, Reed JL, Palsson BO (2004) Genome-scale models of microbial cells: evaluating the consequences of constraints. Nat Rev Microbiol 2:886–897

Reed JL, Famili I, Thiele I, Palsson BO (2006) Towards multidimensional genome annotation. Nat Rev Genet 7(2):130–141

Resendis-Antonio O (2009) Filling kinetic gaps: dynamic modeling of metabolism where detailed kinetic information

is lacking. PLoS One 4(3):e4967. doi:10.1371/journal.pone.0004967
Resendis-Antonio O, Reed JL, Encarnacion S, Collado-Vides J, Palsson BØ (2007) Metabolic reconstruction and modeling of nitrogen fixation in Rhizobium etli. PLoS Comput Biol 3:10 e192
Resendis-Antonio O, Checa A, Encarnación S (2010) Modeling core metabolism in cancer cells: surveying the topology underlying the warburg effect. PLoS One 5(8):e12383. doi:10.1371/journal.pone.0012383
Resendis-Antonio O, Hernandez M, Salazar E, Contreras S, Martinez-Batalla G, Mora Y, Encarnacion S (2012) Systems biology of bacterial nitrogen fixation: high-throughput technology and its integrative description with constraint-based modeling. BMC Syst Biol 5:120
Segre D, Vitkup D, Church G (2009) Analysis of optimality in natural and perturbed metabolic networks. Proc Natl Acad Sci U S A 99(23):15112–15117
Thiele I, Palsson BO (2010) A protocol for generating a high-quality genome scale metabolic reconstruction. Nat Protoc 5(1):93–121

Continuous Model

Yong Wang
Academy of Mathematics and Systems Sciences, Chinese Academy of Sciences, Beijing, China

Synonyms

ODE

Definition

Generally speaking, continuous model is a mathematical model whose variables are real numbers, that is, variables take the continuous value (for example, a real number between 0 and 1). The related term is discrete model whose variables take the discrete value. Physicists often choose continuous mathematical models for problems ranging from the dynamical systems of classical physics to the operator equations and path integrals of quantum mechanics. These mathematical models use the real or complex number fields and we argue that the real-number model of computation should be used in the study of the computational complexity of continuous mathematical models.

In Systems Biology, ordinary differential equations (ODEs) are the most established continuous modeling representation. For example, one uses ODEs to model the concentration of each metabolite which is calculated by a single ODE encapsulating all the reactions where the metabolite is synthesized or consumed, with fluxes determining the transformations to and from other metabolites in the metabolic network.

The use of continuous models is the main technique of the quantitative sciences. The antonym of continuous model is logic model (Please refer to the definition of logic model).

Contractile Actomyosin Ring (CAR)

► Actomyosin Ring

Contractile Ring

► Actomyosin Ring

Control

► Constraint

Controlled Vocabulary

Jörg Hakenberg
Department of Computer Science and Department of Biomedical Informatics, Arizona State University, Tempe, AZ, USA

Definition

A controlled vocabulary (CV) is a set of preselected, predefined, and authorized terms pertaining to a specific domain.

Characteristics

Controlled vocabularies contain carefully selected terms to describe each anticipated concept in a certain domain. As opposed to natural language vocabularies, controlled vocabularies utilize fixed terms and phrases. CVs are employed to build subject headings, taxonomies, and thesauri. These are generally used for annotation and search of documents, clinical notes, studies, experiments, studied materials, etc. CVs alleviate both the processes of initial indexing as well as retrieval, when users are restricted to use the same vocabulary and thus ambiguities such as homonyms, acronyms, and synonyms are avoided.

Cross-References

- ▶ Entity Mention Normalization
- ▶ Information Retrieval
- ▶ Named Entity Recognition
- ▶ Text Mining

Conventional Gene

Marie-Paule Lefranc
Laboratoire d'ImmunoGénétique Moléculaire, Institut de Génétique Humaine UPR 1142, Université Montpellier 2, Montpellier, France

Definition

The conventional gene, or "*conventional*" belongs to the "▶ *GeneType*" concept of identification (generated from the ▶ IDENTIFICATION Axiom) of ▶ IMGT-ONTOLOGY, the global reference in ▶ immunogenetics and ▶ immunoinformatics (Giudicelli and Lefranc 1999; Lefranc et al. 2004, 2005, 2008; Duroux et al. 2008), built by IMGT®, the international ImMunoGeneTics information system® (http://www.imgt.org) (▶ IMGT® Information System). The conventional gene allows to identify any (coding or not coding) gene other than IG or TR genes. A conventional gene has by definition the characteristics of a classical gene (initiation codon and stop codon in the same gene unit).

The conventional gene includes two "GeneType" leafconcepts (▶ IMGT-ONTOLOGY, Leafconcept):

- "*Conventional-with-leader*" identifies any (coding or not coding) gene other than IG or TR genes with a leader L region (or signal peptide).
- "*Conventional-without-leader*" identifies any (coding or not coding) gene other than IG or TR genes with no leader L region (or signal peptide).

The other four "GeneType" leafconcepts are, in contrast, specific to the immunoglobulins (IG) or antibodies and T cell receptors (TR) (▶ Variable (V) Gene, ▶ Diversity (D) Gene, ▶ Joining (J) Gene, ▶ Constant (C) Gene), have special characteristics (▶ Recombination Signal (RS) for V, D, and J, initiation codon and stop codon in different genes V and C, respectively), and need DNA rearrangements during the biosynthesis of the IG and TR chains (Lefranc and Lefranc 2001a, b) (▶ Immunoglobulin Synthesis).

The configuration of a conventional gene is always identified as "undefined" (▶ Configuration Type).

Cross-References

- ▶ Configuration Type
- ▶ Constant (C) Gene
- ▶ Diversity (D) Gene
- ▶ Gene Type
- ▶ IMGT® Information System
- ▶ IMGT-ONTOLOGY
- ▶ IMGT-ONTOLOGY, IDENTIFICATION Axiom
- ▶ IMGT-ONTOLOGY, Leafconcept
- ▶ Immunogenetics
- ▶ Immunoglobulin Synthesis
- ▶ Immunoinformatics
- ▶ Joining (J) Gene
- ▶ Recombination Signal (RS)
- ▶ Variable (V) Gene

References

Duroux P, Kaas Q, Brochet X, Lane J, Ginestoux C, Lefranc M-P, Giudicelli V (2008) IMGT-Kaleidoscope, the Formal IMGT-ONTOLOGY paradigm. Biochimie 90:570–583

Giudicelli V, Lefranc M-P (1999) Ontology for immunogenetics: the IMGT-ONTOLOGY. Bioinformatics 12:1047–1054

Lefranc M-P, Lefranc G (2001a) The immunoglobulin FactsBook. Academic, London, pp 1–458
Lefranc M-P, Lefranc G (2001b) The T cell receptor FactsBook. Academic, London, pp 1–398
Lefranc M-P, Giudicelli V, Ginestoux C, Bosc N, Folch G, Guiraudo D, Jabado-Michaloud J, Magris S, Scaviner D, Thouvenin V, Combres K, Girod D, Jeanjean S, Protat C, Yousfi Monod M, Duprat E, Kaas Q, Pommié C, Chaume D, Lefranc G (2004) IMGT-ONTOLOGY for immunogenetics and immunoinformatics. In Silico Biol 4:17–29
Lefranc M-P, Clément O, Kaas Q, Duprat E, Chastellan P, Coelho I, Combres K, Ginestoux C, Giudicelli V, Chaume D, Lefranc G (2005) IMGT-Choreography for immunogenetics and immunoinformatics. In Silico Biol 5:45–60
Lefranc M-P, Giudicelli V, Regnier L, Duroux P (2008) IMGT, a system and an ontology that bridge biological and computational spheres in bioinformatics. Brief Bioinform 9:263–275

Convergent Evolution

William Bechtel
Department of Philosophy and Center for Chronobiology, University of California, San Diego, La Jolla, San Diego, CA, USA

Synonyms

Homoplasy

Definition

Convergent evolution is the process by which distantly related species independently evolve similar genotypic or phenotypic traits. Trait convergence is typically caused by the adaptation of organisms to similar environments by means of natural selection. However, traits can also converge as a result of design constraints imposed by biological systems and chance. Well-known examples representing the convergence of gross morphological traits include the visual systems of vertebrates, cephalopods, and arthropods and the wings of birds, bats, and insects. Traits that are similar due to convergent evolution are called "analogous," in contrast to "homologous" traits, which are similar due to common evolutionary or developmental descent.

Cross-References

► Mechanism, Conserved

Convex Hull

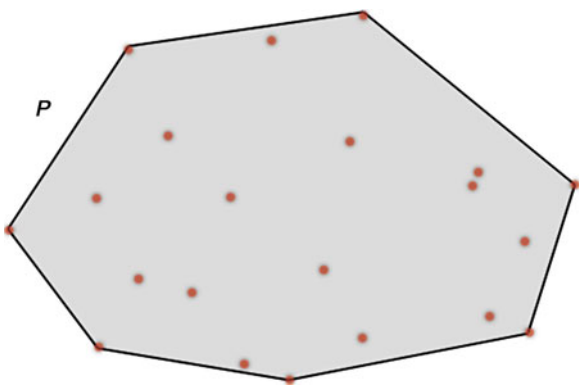
Virginio Cantoni[1,2], Alessandro Gaggia[1] and Luca Lombardi[1]
[1]Department of Computer Engineering and Systems Science, University of Pavia, Pavia, Italy
[2]Computational Biology, KTH Royal Institute of Technology, Stockholm, Sweden

Definition

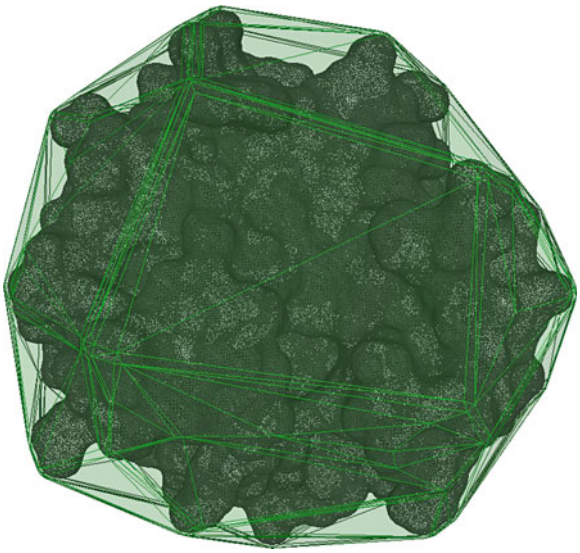
For a set of points S the Convex Hull $H(S)$, in 2D space, is the smallest convex polygon P that encloses S, smallest in the sense that there is no other polygon P' such that $S \subseteq P' \subset P$ (Fig. 1).

3D Convex HULL

The previous definition can be easily extended to the 3D space considering P as the smallest polyhedron that encloses the 3D set of points S (Fig. 2).



Convex Hull, Fig. 1 A 2D convex hull



Convex Hull, Fig. 2 A 3D convex hull

Cross-References

▶ Extended Gaussian Image for Pocket-Ligand Matching

Convex Optimization

Frank Allgöwer, Jan Hasenauer and Steffen Waldherr
Institute for Systems Theory and Automatic Control, University of Stuttgart, Stuttgart, Germany

Definition

Convex optimization is the task of solving a convex optimization problem. The goal of an optimization problem is to find a value for the optimization variable, lying within some given constraints, which minimizes an objective function.

Mathematically, an optimization problem over the optimization variable x in an n-dimensional vector space over the real numbers is written as

$$\begin{aligned} &\underset{X\in\mathbb{R}^n}{\text{minimize}} && f_0(x) \\ &\text{subject to} && f_i(x) \leqslant b_i, \quad i = 1, \ldots, m. \end{aligned} \tag{1}$$

The objective function is given by f_0, and the functions f_i are inequality constraints. If there does not exist a variable x which satisfies the constraints, the optimal value is ∞ by definition.

An optimization problem is called a convex optimization problem if the objective function f_0 and all of the inequality constraints f_i, $i = 1,\ldots, m$ are convex, i.e., satisfy the condition $f_i((1-\alpha)x + \alpha y) \leq (1-\alpha)f_i(x) + \alpha f_i(y)$ for all $\alpha \in [0,1]$ and $x,y \in \mathbb{R}^n$ (Boyd and Vandenberghe 2004). In convex optimization problems, any local minimum is guaranteed to be a global minimum. Convex optimization problems can be solved very efficiently by interior point methods.

An important concept in convex optimization is duality. The Lagrangian function associated with the optimization problem (Eq. 1) is defined as

$$L(x, \lambda) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x). \tag{2}$$

The dual optimization problem for (Eq. 1) is defined as

$$\max_{\lambda\in\mathbb{R}^m_+} \inf_{x\in\mathbb{R}^n} L(x, \lambda). \tag{3}$$

For certain classes of convex optimization problems, for example, linear programs or ▶ semidefinite programs, the dual problem is of the same class as the primal problem. Thus, solvers for these types of problems can also be applied to solve the dual problem.

The dual problem (Eq. 3) has the property that its optimal objective function value is always less than or equal to the optimal value of the primal problem (Eq. 1). Also, evaluation of the objective function in the dual problem for any feasible λ directly gives a lower bound on the optimal value of the primal problem and can be used to estimate how far the objective function value for any feasible x is away from the optimal value. For convex optimization problems, the optimal objective function values in the primal and dual problem are even equal in many cases. This property of duality can be used to certify nonexistence of solutions to ▶ feasibility problems where an algorithm to solve the dual problem is available.

Cross-References

- ▶ Convex Programming
- ▶ Model Falsification
- ▶ Semidefinite Programming

References

Boyd S, Vandenberghe L (2004) Convex optimization. Cambridge University Press, Cambridge

Convex Programming

Lin Wanglin
School of Computer Science and Information Engineering, Tianjin University of Science and Technology, Tianjin, China

Synonyms

Convex optimization

Definition

A following nonlinear programming is called a convex programming if $f(x),\ g_i(x),\ i = 1, \cdots, m,$ are convex functions in R^n:

$$\begin{aligned} &\text{Minimize} && f(x) \\ &\text{subject to} && g_i(x) \le 0, \quad i = 1, \cdots, m, \\ & && Ax = b, \end{aligned} \tag{1}$$

where A is an $p \times n$ matrix.

References

Zhang XS (2000) Neural networks in optimization. Kluwer, Dordrecht

Core Promoter Elements

Tetsuro Kokubo
Department of Supramolecular Biology, Graduate School of Nanobioscience, Yokohama City University, Yokohama, Kanagawa, Japan

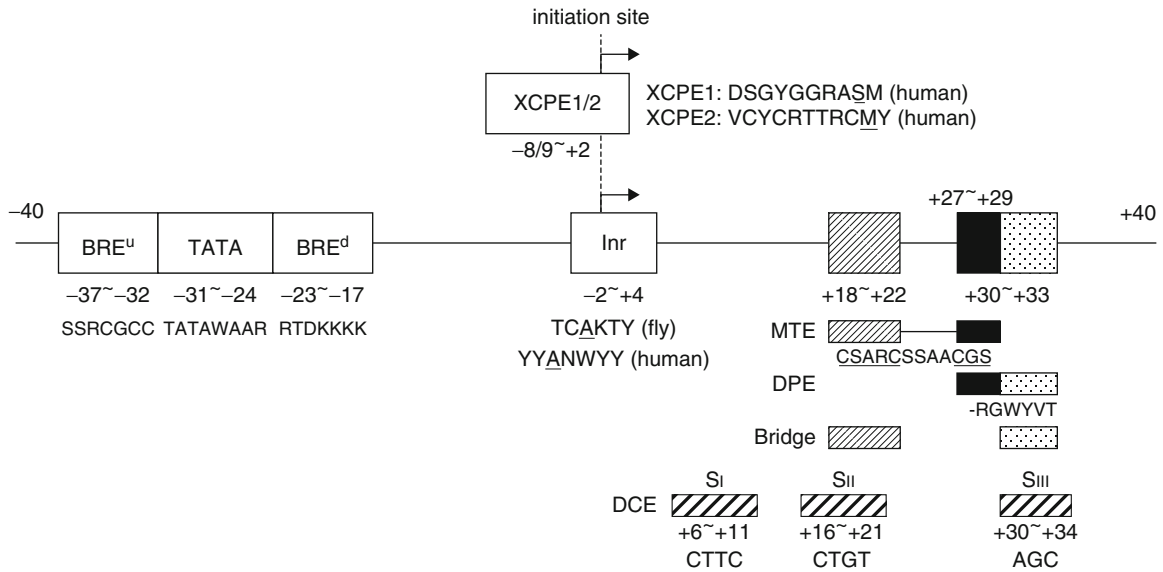
Synonyms

Cis-Acting sequences in the core promoter

Definition

The core promoter elements are *cis*-acting functional sequences embedded within the core promoters. Several representatives in the RNAPII system are summarized in Fig. 1 (Juven-Gershon and Kadonaga 2010; Smale and Kadonaga 2003). The first-discovered and best-characterized is the TATA box, which serves as a recognition site for TBP. This element is widely conserved from yeast to human, but is found in only a small fraction (approximately 10–20%) of all RNAPII promoters. Importantly, TBP is essential for transcription, regardless of whether the TATA box is present or absent in the core promoter. Therefore, TBP must play some indispensable but as-yet unidentified role(s) in transcription, other than simple binding to the TATA box. In some human promoters, the TATA box is flanked by BRE^u and/or BRE^d, both of which are recognition sites for TFIIB. These elements function positively or negatively depending on the architecture of promoters and/or transcription factors present in the system.

The DNA region encompassing the transcription initiation site(s) often contains the initiator (Inr) element (Fig. 1). Inr consensus sequences appear to be different in human (YYANWYY) and *Drosophila* (TCAKTY) (underlined A indicates initiation site/+1). In addition, even in human, there are multiple Inr consensus sequences ranging from a very relaxed motif (Y*R*) to a very strict motif (sINR; GSCGCCATYTTG) (Dikstein 2011), depending on the estimation method. Therefore, it is likely that there are a variety of core promoter elements



Core Promoter Elements, Fig. 1 S(G/C), K(G/T), R(A/G), M(A/C), W(A/T), Y(T/C), V(G/A/C), D(G/A/T), BRE (TFIIB recognition element), BRE^{u} (upstream BRE), BRE^{d} (downstream BRE), TATA (TATA box), Inr (Initiator), MTE (motif ten element), DPE (downstream promoter element), DCE (downstream core element), XCPE1/2 (X gene core promoter element 1/2)

encompassing the initiation site(s), with only a fraction belonging to the original Inr. Consistent with this, some human genes contain XCPE1 or XCPE2, which were originally identified in the core promoter of hepatitis B virus X gene. These elements are not only structurally but also functionally different from Inr.

In higher eukaryotes, some core promoter elements are located downstream of transcription initiation site(s) (Fig. 1). Although MTE and DPE were originally identified as separate elements, it was recently proposed that they comprised two distinct functional subregions, one of which is included in common in these two elements. The novel core promoter element that contains the 5′-subregion of MTE and the 3′-subregion of DPE is designated as "Bridge." DCE is another downstream core promoter element that also comprises three subregions (SI, SII, and SIII). DCE is not related to MTE, DPE, and Bridge, because the distance from the transcription initiation site(s) is highly restricted for the latter three elements but not for the former element. Conversely, the TATA box occurs frequently with the former element but not with the latter elements.

Nearly all of the core promoter elements are recognition sites for GTFs. As previously discussed, TATA box and $BRE^{u/d}$ are recognized by TBP and TFIIB, respectively. Importantly, TFIID (*t*ranscription *f*actor *IID*), a large protein complex comprising TBP and 14 TAFs (*T*BP-*a*ssociated *f*actors), is involved in the recognition of Inr, MTE, DPE, Bridge, and DCE. Specifically, Inr, MTE/DPE/Bridge and DCE may be recognized by TAF1/TAF2, TAF6/TAF9, and TAF1, respectively. However, it is still unclear which factor recognizes XCPE1/2. Native promoters contain only one or a few of these core promoter elements, probably because a weaker level of basal transcription is more advantageous for regulation by activators. In fact, an artificial core promoter containing TATA box, Inr, MTE, and DPE is highly active even in the absence of activators.

Core promoter elements such as Inr that overlap with the initiation site(s) could function as a platform for PIC assembly. However, these elements may also contain as-yet uncharacterized but crucial information that enables RNAPII to initiate transcription productively. Genetic studies indicate

that such information may be decoded by TFIIB, TFIIF, and RNAPII, since these three transcription factors are required for accurate initiation site selection.

Cross-References

► Transcription in Eukaryote

References

Dikstein R (2011) The unexpected traits associated with core promoter elements. Transcription 2(5):201–206
Juven-Gershon T, Kadonaga JT (2010) Regulation of gene expression via the core promoter and the basal transcriptional machinery. Dev Biol 339(2):225–229
Smale ST, Kadonaga JT (2003) The RNA polymerase II core promoter. Annu Rev Biochem 72:449–479

Cores

José Román Bilbao Castro
Supercomputing-Algorithms group, University of Alméria, Alméria, Spain

Definition

Core is the name currently given to a single processor within a multiprocessor chip. So, if we read "This processor is a quad core one," it really means "this chip contains four cores of processors." Cores are encapsulated together to form a single chip and some additional elements are added to improve communications among them.

Cross-References

► Multicore computing

Corneal Pocket Assay

Marsha A. Moses
Department of Surgery/Harvard Medical School, Vascular Biology Program/Children's Hospital Boston, Boston, MA, USA

Definition

This assay, originally developed for use in the rabbit but now routinely conducted using rodent cornea, requires creation of a micropocket in the normally avascular corneal immediately above the limbus into which test substances that represent putative angiogenesis stimulators or inhibitors (proteins, cells, or tissues) can be placed to determine their effect on blood vessels that arise from the limbus (Gimbrone et al. 1974). Alternatively, blood vessels stimulated by growth factors implanted in the pocket can be inhibited via systemic administration of anti-angiogenic agents (Moses et al. 1999).

Cross-References

► Neovascularization

References

Gimbrone MA Jr, Cotran RS, Leapman SB, Folkman J (1974) Tumor growth and neovascularization: an experimental model using the rabbit cornea. J Natl Cancer Inst 52:413–427
Moses MA, Wiederschain D, Wu I, Fernandez CA, Ghazizadeh V, Lane WS, Flynn E, Sytkowski A, Tao T, Langer R (1999) Troponin I is present in human cartilage and inhibits angiogenesis. Proc Natl Acad Sci USA 96:2645–2650

Corpora for Biomolecular Event Extraction

► Text Corpora, Molecular Event Extraction

Corpora for Event Extraction

▶ Text Corpora, Molecular Event Extraction

Corpus Luteum

Marsha A. Moses
Department of Surgery/Harvard Medical School, Vascular Biology Program/Children's Hospital Boston, Boston, MA, USA

Synonyms

Yellow body

Definition

The corpus luteum is a temporary endocrine structure formed during the mammalian menstrual cycle. In the early phase of the human menstrual cycle, 20–25 primary follicles containing oocytes begin to develop into secondary follicles, among which only one will mature into a Graafian follicle containing a mature egg. During ovulation, the egg is released upon rupture of the Graafian follicle, in which the follicular cells then undergo reprogramming to form the corpus luteum (Rizzo 2006). During this process, the follicular cells exit cell cycle and undergo terminal differentiation into estrogen- and progesterone-producing luteal cells. If fertilization and implantation do not occur, the corpus luteum soon degenerates to form the corpus albicans (white body), and a new menstrual cycle is initiated (Stocco et al. 2007). Intense angiogenesis occurs during the formation of the corpus luteum, which is regulated by key angiogenic factors such as vascular endothelial growth factor (VEGF) and matrix metalloproteinase-2 (MMP-2) (Stocco et al. 2007; Zhang et al. 2005).

Cross-References

▶ Neovascularization

References

Rizzo D (2006) Fundamentals of anatomy & physiology. Delmar Cengage Learning, Clifton Park
Stocco C, Telleria C, Gibori G (2007) The molecular control of corpus luteum formation, function, and regression. Endocr Rev 28(1):117–149
Zhang B, Yan L, Tsang PC, Moses MA (2005) Matrix metalloproteinase-2 (MMP-2) expression and regulation by tumor necrosis factor alpha (TNFalpha) in the bovine corpus luteum. Mol Reprod Dev 70(2):122–132

Correlation

▶ Correlation Coefficient

Correlation Coefficient

Fuyan Hu
Institute of Systems Biology, Shanghai University, Shanghai, China

Synonyms

Coefficient of correlation; Correlation; Correlational statistics; Correlativity

Definition

A correlation coefficient is a statistic index which measures the degree to which two variables are related. Generally, it varies from −1 (perfect negative correlation) through 0 (no correlation) to +1 (perfect positive correlation). If a negative correlation is detected, whenever one variable has a high (low) value, the other will have a low (high) value. If it is a positive correlation, whenever one variable has a high (low) value, so does the other.

The correlation coefficient of x_1 and x_2 is given by the formula:

$$\frac{\mathrm{cov}(x_1, x_2)}{\sigma_1 \sigma_2} = \frac{E[(x_1 - \mu_1)(x_2 - \mu_2)]}{\sigma_1 \sigma_2} \quad (1)$$

where μ_1 is the arithmetic mean of x_1, μ_2 is the arithmetic mean of x_2, $\mathrm{cov}(x_1, x_2)$ is the covariance of x_1 and x_2, σ_1 is the standard deviation of x_1, and σ_2 is the standard deviation of x_2.

Cross-References

▶ Correlation Relationship

References

Rodgers JL, Nicewander WA. Thirteen ways to look at the correlation coefficient. Am Stat. 1988;42(1):59–66.

Correlation Network

▶ Relevance Networks

Correlation Relationship

Katsuhisa Horimoto
Computational Biology Research Center, National Institute of Advanced Industrial Science and Technology, Koto-ku, Tokyo, Japan

Synonyms

Bayes rule; Causal relationship; Correlation coefficient

Definition

A correlation is the degree of the relationship between two variables. A positive correlation is a relationship that the amounts of the two variables simultaneously increase. In a negative correlation, as the amount of one variable increases, the amount of another variable decreases. Note that there is no evidence that changes in one variable cause changes in the other variable. In statistics, correlation frequently means the degree to which two or more quantities are linearly associated. In particular, the degree of correlation between the values on the two axes in a two-dimensional plot is quantified by the correlation coefficient. Also, correlating values of one variable with corresponding values at different times is called autocorrelation.

Cross-References

▶ Causal Relationship
▶ Conditional Independence

Correlational Statistics

▶ Correlation Coefficient

Correlativity

▶ Correlation Coefficient

CoSBiLab

▶ BlenX

Cost Function

▶ Objective Function

Covariance Graph

▶ Relevance Networks

Covariance Selection Model

▶ Graphical Gaussian Model

CpG Island

Jianhua Ruan
Department of Computer Science, University of Texas at San Antonio, San Antonio, TX, USA

Definition

In genetics, CpG islands refer to genomic regions that contain a high frequency of CG dinucleotides, where the "p" simply indicates that C and G are on the same strand connected by a phosphodiester bond. CpG islands are typically located near the transcription start site of housekeeping genes or other genes that are highly expressed.

CPM

► Cellular Potts Model

Criterion Function

► Objective Function

Cross Tissue Disease Network

Sanjeev Kumar[1] and Shipra Agrawal[2]
[1]BioCOS Life Sciences Private Limited, Bangalore, Karnataka, India
[2]BioCOS Life Sciences Pvt. Limited, Institute of Bioinformatics and Applied Biotechnology, Bangalore, Karnataka, India

Synonyms

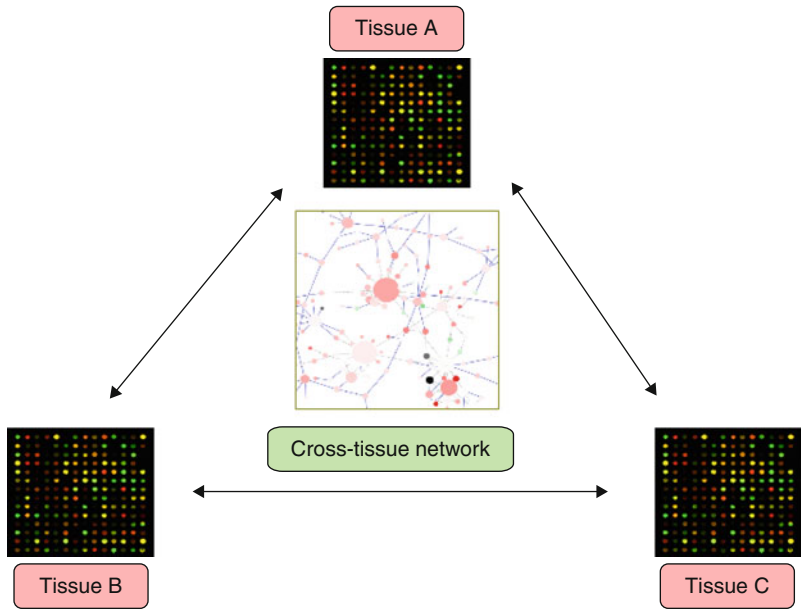
Cross tissue signature network; Tissue-to-tissue network

Definition

A cross view of the different molecular states, interaction of physiological functions and pathways across multiple tissues of a diseased system could be appropriately and comparably presented through ► cross tissue disease networks. CTDNs are constructed from specific gene/protein expression data across important tissues of diseased organisms (Fig. 1).

The application of CTDN could be properly defined from the following examples:

- It could be used to study the inter-tissue relationships and identify genes, which are important in different tissues and can also act as information relays in the control of peripheral tissues in obese mice. The subnetworks, which are specific to the tissue-to-tissue interactions, are enriched in the genes having corresponding disease-relevant biological functions (Drake 2010).
- CTDNs derived from gene co-expression data have been shown to determine the pleiotropic effects of single disease gene in human. Such genes differ in their expression-based tissue-specific interactions (Schadt 2009).
- The gene enrichment and genetic association analyses of CTDN modules facilitate identification of important modules relating to macrophage function and metabolic traits in human obese patients. This module has been shown to have high correlation with obesity-specific clinical parameters. Such networks can also describe causal relationship of genes with disease-associated phenotypes (Schadt 2009).
- Tissue-to-tissue networks enable the identification of disease-specific genes that respond to the physiological changes in one tissue, which is actually triggered in some other tissues (Sieberts and Schadt 2007).
- CTDN could also identify genes related to communication between tissues. These networks provide a first step toward understanding of the complex diseases by laying out the hierarchy of interacting molecular networks. These interacting networks in turn define various physiological states in the mammalian systems (Drake 2010).



Cross Tissue Disease Network, Fig. 1 A tissue-to-tissue network is generated from each possible tissue pairs by identifying the significantly correlated gene co-expression traits

References

Drake TA (2010) Genes and pathways contributing to obesity: a systems biology view. Prog Mol Biol Transl Sci 94:9–38

Schadt EE (2009) Molecular networks as sensors and drivers of common human diseases. Nature 461:218–223

Sieberts SK, Schadt EE (2007) Moving toward a system genetics view of disease. Mamm Genome 18:389–401

Crossover Design

► Repeated Measures

Cross Tissue Signature Network

► Cross Tissue Disease Network

Cross-Validation

Joo Chuan Tong
Data Mining Department, Institute for Infocomm Research, Singapore, Singapore
Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, Singapore

Definition

Training data set will be randomly partitioned into *i*-number of subsets with equal size. Then one subset will be evaluated by a model trained by the data from the *I-1* subsets, this process is repeated for *i* times. A good model trained on data points that are representative for the population should consistently produce high accuracy across the *i*-number of validation. This process is called cross-validation.

Cross-References

► TAP Translocator, In Silico Prediction

CRSP

► Mediator

Cumulative Causation

► Positive Feedback

Curation

Sabina Leonelli
ESRC Centre for Genomics in Society, University of Exeter, Exeter, Devon, UK

Definition

Ensemble of activities involved in developing and maintaining a database (Howe and Rhee 2008; Stein 2008; Leonelli 2010).

Cross-References

► Bio-Ontologies
► CellML Model Curation

References

Howe D, Rhee SY (2008) The future of biocuration. Nature 455:47–50
Leonelli S (2010) Packaging data for re-use: databases in model organism biology. In: Howlett P, Morgan MS (eds) How well do facts travel? The dissemination of reliable knowledge. Cambridge University Press, Cambridge, MA
Stein LD (2008) Towards a cyberinfrastructure for the biological sciences: progress, visions and challenges. Nat Rev Genet 9(9):678–688

Cyclin Detection

► Quantifying Lymphocyte Division, Methods

Cyclin-dependent Kinase 1 (CDK1)/ Anaphase-Promoting Complex (APC) Oscillator

► Cell Cycle of Early Frog Embryos

Cycling Egg Extract

► Xenopus laevis Egg Extract

Cyclins and Cyclin-dependent Kinases

Marcos Malumbres
Spanish National Cancer Research Centre (CNIO), Madrid, Spain

Synonyms

Cell cycle-regulated kinase complexes

Definition

Cyclins are proteins initially identified by their differential protein levels during different phases of the ► cell cycle. This characteristic depends on specific amino acid sequences that are recognized by the protein degradation machinery. Cyclins are known to function as the regulatory subunit of heterodimeric protein kinase complexes that also contain a catalytic subunit with serine/threonine kinase activity, known as cyclin-dependent kinase (Cdk).

Characteristics

Cyclins were first discovered as oscillating proteins that accumulate and suddenly disappear during cell division in sea urchin eggs (Jackson 2008). These proteins are usually not expressed in quiescent cells but they are synthesized in dividing cells to regulate entry into the cell cycle and progression throughout the different phases of the cell cycle. However, what makes these proteins special is the tight control of their levels by protein degradation. Cyclins contain specific amino acid sequences, such as the destruction (D)-box, that are recognized by the ubiquitin-mediated protein degradation machinery. Specific ubiquitin ligases recognize these residues and tag these proteins through the addition of ubiquitin residues. These ubiquitinated cyclins are then quickly degraded by the proteasome.

Cyclins contain a common amino acid stretch, known as the cyclin box, responsible for binding and activation of their partner kinases, the Cdks. Monomeric Cdks are inactive as kinases and their activation requires the interaction with cyclins. The binding between cyclins and Cdks results in crucial changes in the structural conformation of the kinase domain of Cdks (Morgan 2007). In addition, cyclins are thought to provide additional recognition sites for specific substrates. Thus, cyclins determine the substrate specificity and timing of activation of Cdks. In addition, their interaction with Cdks modulates other crucial aspects of Cdk function such as its control by upstream regulatory enzymes or by subcellular localization (Malumbres and Barbacid 2005).

Mammalian Cyclins

Cyclins are defined by the presence of one or two cyclin boxes. About 35 cyclins exist in the human genome (Fig. 1). Many of these proteins display little sequence similarity outside the cyclin box and their functional homology is therefore unclear. The N-terminal region usually contains regulatory and targeting domains, including the D-box, that are specific for each cyclin class.

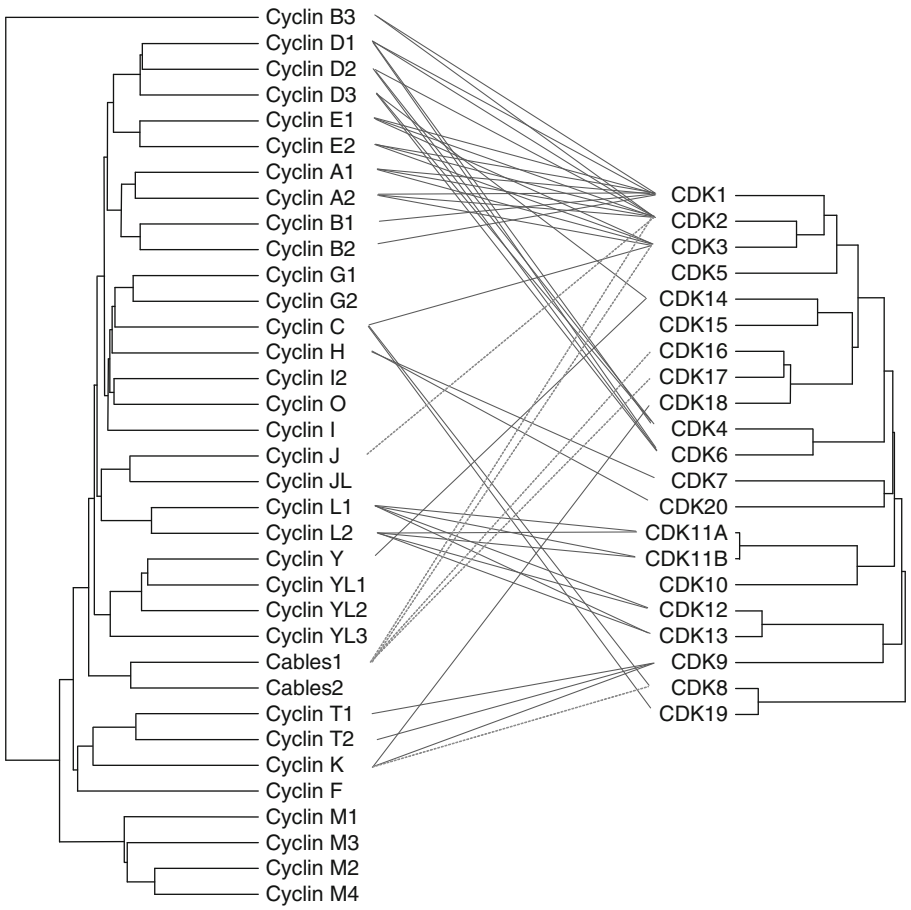
Major cell cycle cyclins are grouped into a cluster that contains A-, B-, D-, and E-type cyclins. Cyclins C, H, K, L, and T are mostly involved in the control of transcription and RNA processing. G-type cyclins have been proposed to play a role in the response to DNA Damage (▶ Cell Cycle Arrest After DNA Damage) whereas M-cyclins are divergent cyclins involved in ion transport. Cables1,2 are cyclin-like proteins that are both partners and substrates of Cdks and may be involved in apoptosis. Several other cyclins including Cyclin F, G1,2, I, I2, JL, M1-4, and O do not have cognate Cdk partners. Despite their name, it is also unclear whether the protein levels of many of these cyclins are tightly regulated by protein synthesis and degradation.

Mammalian Cdks

The Cdk denomination was originally applied to kinases whose function was clearly demonstrated to be dependent on a cyclin. This denomination is now applied to 20 different kinases (CDK1-20; Fig. 1) although their control by cyclins is not well characterized in some cases (Malumbres et al. 2009). All Cdks display high similarity in their kinase domain and contain a specific stretch of amino acids, usually known as the PSTAIRE domain (or α1 helix), involved in the binding to cyclins (Morgan 2007). Similarly to cyclins, this family of proteins includes members with a clear implication in the cell division cycle. Thus, CDK1, 2, 3, 4, and 6 play distinct functions in the progression throughout the cell cycle upon activation by specific members of the A-, B-, D-, and E-cyclins. On the other hand, other Cdks, such as CDK7-9 and CDK11 display critical roles in ▶ transcription and RNA processing, usually in combination with C-, H-, K-, L-, and T-cyclins (Malumbres and Barbacid 2005).

Regulation and Function of Cdks in the Control of the Cell Cycle

Only a few cyclin-Cdk complexes are known to be directly involved in the regulation of the cell cycle. CDK1, the mammalian ortholog of the yeast CDC2/CDC28 kinase (▶ Cell Cycle, Budding Yeast; ▶ Cell Cycle, Fission Yeast), is a major regulator of G2 and ▶ mitosis (▶ Mitotic Kinases) in combination with A- and B-type cyclins. CDK2 (and perhaps CDK3) are activated by E- and A-type cyclins and may participate in cell cycle control during ▶ DNA replication and G2 (▶ Cell Cycle Transitions, G2/M) and in DNA repair. CDK4 and CDK6, on the other hand, are specifically activated by D-type cyclins and control the entry into the cell cycle and G1 progression in



Cyclins and Cyclin-dependent Kinases, Fig. 1 Representative interactions between mammalian cyclins and Cdks. CDK5 may interact with some cyclins (e.g., D-type cyclins) although this binding does not result in kinase activation. Dotted lines indicate interactions whose biological significance is not clear. Evolution distances in cyclins and Cdks are represented at different scale

response to mitogenic signaling (Malumbres and Barbacid 2005). It is not clear whether most of the other Cdks play direct roles in the regulation of the cell cycle. CDK5 for instance is activated by cyclin-like proteins that are almost uniquely expressed in brain cells. As indicated above, CDK7-9 and CDK11 are involved in transcription and RNA processing and the information available for most other Cdks is too scarce to assign a clear role in the control of the cell cycle.

The activation of cyclin-Cdk complexes is tightly controlled at different levels including phosphorylation and dephosphorylation events, as well as folding and subcellular localization. Cdks are regulated by both activating and inactivating phosphorylations. In mammals, CDK7 is the catalytic subunit of the complex known as Cdk-activating protein (CAK). CAK phosphorylates several Cdks at a critical threonine residue at the T-loop of the kinase domain, and this phosphorylation is required to unblock the active-site cleft. On the other hand, two additional kinases, WEE1 and MYT1, are able to phosphorylate two residues located in the roof of the kinase ATP-binding site and these phosphorylations inhibit Cdk kinase activity. Dephosphorylation of these sites is mediated by CDC25(A-C) phosphatases. Both WEE1 and CDC25 enzymes are critical regulators of Cdk activation in response to multiple signals such as those

involved in the DNA damage response (Morgan 2007; Jackson and Bartek 2009) (▶ Cell Cycle Arrest After DNA Damage). Cdks are also regulated by direct binding to several ▶ Cdk inhibitors of the Ink4 ($p16^{Ink4a}$, $p15^{Ink4b}$, $p18^{Ink4c}$ and $p19^{Ink4d}$) and Cip/Kip ($p21^{Cip1}$, $p27^{Kip1}$ and $p57^{Kip2}$) families. Ink4 proteins bind CDK4 and CDK6 thus preventing the interaction of the monomeric Cdk with their activating cyclins. Cip/Kip inhibitors, on the other hand, generally bind and inhibit cyclin-Cdk complexes generating inactive trimeric complexes (Sherr and Roberts 1999)

The typical phosphorylation sequence recognized by Cdks is [S/T]*PX[K/R], where [S/T]* indicates the phosphorylated serine or threonine residue, P is proline, X represents any amino acid, and [K/R] represents a basic amino acid lysine or arginine two positions downstream of the phosphorylated residue. Cdks exert their effects in the cell cycle by phosphorylating a large number of proteins in the cell, thus regulating many aspects of cellular architecture and metabolism. In response to mitogenic signals, cyclin D-CDK4/6 complexes phosphorylate and inhibit the retinoblastoma protein (pRB), a transcriptional ▶ repressor that defines the ▶ restriction point in mammals by preventing the synthesis of cell cycle proteins. pRb exerts this function by inhibiting E2F transcription factors and recruiting chromatin remodeling complexes thus repressing transcription at specific sites. pRB may be also phosphorylated by many other Cdks such as CDK1, CDK2, CDK3, CDK9. Upon pRB inactivation, many proteins required for the synthesis of DNA and mitosis are synthesized, and cells become committed to ▶ DNA replication and progression throughout the following phases of the cell cycle. CDK2 and CDK1, in complex with E-, A-, and B-type cyclins, are responsible for the phosphorylation of a significant number of proteins. CDK1 (▶ Mitotic Kinases), the major Cdk activity during G2 and M, phosphorylates more than 70 proteins involved in DNA condensation, formation of the spindle, Golgi remodeling, nuclear envelop breakdown, etc. (Malumbres and Barbacid 2005). Mitotic exit (▶ Cell Cycle Transitions, Mitotic Exit) requires the proteasome-dependent degradation of A- and B-type cyclins, thus switching off Cdk activity and allowing DNA decondensation, spindle disassembly, reformation of the nuclear envelop, etc. Normal progression throughout the cell cycle is monitored by several ▶ cell cycle checkpoints which sense possible abnormalities and generally result in the inhibition of Cdk activity and cell cycle arrest.

In general, Cdks are considered as the major engines that drive entry and progression throughout the different phases of the eukaryotic cell cycle. Not surprisingly, their activity is deregulated in human cancer (▶ Cell Cycle, Cancer Cell Cycle and Oncogene Addiction) and inhibiting Cdk activity is now considered as an attractive therapeutic approach against tumor cell proliferation (Malumbres and Barbacid 2009).

Cross-References

- ▶ Cdk Inhibitors
- ▶ Cell Cycle
- ▶ Cell Cycle Arrest After DNA Damage
- ▶ Cell Cycle Checkpoints
- ▶ Cell Cycle Transition, Principles of Restriction Point
- ▶ Cell Cycle Transitions, G2/M
- ▶ Cell Cycle Transitions, Mitotic Exit
- ▶ Cell Cycle, Budding Yeast
- ▶ Cell Cycle, Fission Yeast
- ▶ DNA Replication
- ▶ Mitosis
- ▶ Mitotic Kinases
- ▶ Repressor
- ▶ Transcription

References

Jackson PK (2008) The hunt for cyclin. Cell 134:199–202
Jackson SP, Bartek J (2009) The DNA-damage response in human biology and disease. Nature 461:1071–1078
Malumbres M, Barbacid M (2005) Mammalian cyclin-dependent kinases. Trends Biochem Sci 30:630–641
Malumbres M, Barbacid M (2009) Cell cycle, Cdks and cancer: a changing paradigm. Nat Rev Cancer 9:153–166
Malumbres M, Harlow E, Hunt T, Hunter T, Lahti JM, Manning G, Morgan DO, Tsai LH, Wolgemuth DJ (2009) Cyclin-dependent kinases: a family portrait. Nat Cell Biol 11:1275–1276
Morgan DO (2007) The cell cycle: principles of control. New Science Press, London
Sherr CJ, Roberts JM (1999) CDK inhibitors: positive and negative regulators of G1-phase progression. Genes Dev 13:1501–1512

Cyclosome

- ▶ Anaphase-Promoting Complex (APC/C)
- ▶ Anaphase-Promoting Complex Inhibitors

Cytogenetics

Jingky Lozano-Kühne
Department of Public Health, University of Oxford, Oxford, UK

Definition

Cytogenetics is a branch of genetics that deals with the study of the structure and function of the chromosomes (Prabhu Britto and Ravindran 2007).

References

Prabhu Britto A, Ravindran DG (2007) A review of cytogenetics and its automation. J Med Sci 7:1–18

Cytokines

Rohan Dhiman
Center for Pulmonary and Infectious Disease Control, University of Texas Health Science Center at Tyler, Tyler, TX, USA

Definition

These are low molecular weight molecules secreted by various immune cells which modulate the immune system by acting as intercellular mediators (Gilman et al. 2001)

References

Gilman A, Goodman LS, Hardman JG, Limbird LE (2001) Goodman & Gikman's the pharmacological basis of therapeutics. Mc Graw-Hill, New York

Cytokinesis

Sebastian Mana-Capelli and Dannel McCollum
Department of Molecular Genetics and Microbiology, University of Massachusetts, Worcester, MA, USA

Definition

Cytokinesis is the process by which one cell divides into two daughter cells with equivalent genetic and cytoplasmic content. Cytokinesis initiates after the chromosomes have segregated and culminates by dividing the cell perpendicular to the mitotic spindle after the completion of telophase, thus bringing closure to the ▶ cell cycle. The process of cell division requires extensive rearrangements of the cytoskeleton and vesicle trafficking apparatus. Much of the spatial and temporal regulation of cytokinesis is controlled by ▶ small GTPases and regulatory kinases that coordinate this process with the nuclear cycle.

Characteristics

Cytokinesis can be universally divided into three major events: determination of the cleavage plane, assembly (and constriction) of the cytokinesis apparatus, and cell abscission. However, different organisms have developed remarkably variable approaches to cytokinesis. Specification of the division plane is the least conserved of the three major events. Once the site is determined, metazoans and fungi divide through the use of a contractile ▶ actomyosin ring. In contrast, plants assemble membranes and cell wall in a structure called the phragmoplast, which assembles in the cell middle and then expands outward toward the cell cortex. The final joining of cell membranes to bring about completion of cytokinesis is called abscission (Guertin et al. 2002). In this chapter, we will survey the mechanisms by which different cell types undergo cytokinesis, with special emphasis on the model systems that have been extensively studied. Both commonalities and differences (e.g., presence of a cell wall, or the use of a contractile ring) are discussed below.

Determination of the Cleavage Plane

Animal Cells and Some Lower Eukaryotes

In animal cells and some lower eukaryotes (but not yeast), the selection of the division plane is determined by the position and orientation of the mitotic spindle at anaphase. Although the cell division plane typically bisects the mitotic spindle, it has long been argued whether the signal to position the furrow originates from the central spindle or overlapping astral microtubules from opposite poles. More recent evidence suggests that both types of microtubules can induce furrow formation, with the emphasis depending on the cell type. For example in large cells such as early embryos, the central spindle is a long distance from the cortex, and in these cells, the astral microtubules seem to have a more important role in positioning the cell division site. There has also been a debate over the role of astral microtubules, which could act either by promoting constriction in the cell middle, or inhibiting cortical tension at the cell poles (Eggert et al. 2006). At the molecular level, there is evidence that spindle microtubules control the equatorial cortex contractility by regulating the activity of the small GTPase RhoA, the master driver of actomyosin ring formation and constriction. The RhoA GAP MGCRacGap and the GEF ECT2 both localize to the spindle midzone in anaphase, where they form a complex that is regulated by several mitotic kinases, including Cdk1, Aurora B, and Plk1. Unfortunately, it is not clear yet how the signal is transduced from the spindle to the cortex (Piekny et al. 2005).

Budding Yeast

Unlike animal cells, the plane of division in *Saccharomyces cerevisiae* is determined before ▶ mitosis. Since this type of yeast divides through budding, the plane of cytokinesis is the narrow connection between the mother and daughter cell, known as the bud neck. The site selection for bud growth is determined in G1/S by the localization of landmark proteins. Interestingly, as in mammalian cells, the landmark proteins recruit a small GTPase (in this case RSR1), which starts a GTPase cascade resulting in growth of the bud and assembly of the septin ring, which acts as a scaffold for actomyosin ring formation (Oliferenko et al. 2009).

Fission Yeast

The division plane in *Schizosaccharomyces pombe* is determined by the position of the nucleus in late G2/early mitosis (Oliferenko et al. 2009). Fission yeast cells are cylindrical in shape, and the nucleus is roughly positioned at the center of the cell by interphase microtubules, leading to the generation of two daughter cells of approximately the same size. Mid1, a paralog to metazoan anillins, plays a central role in the determination of the cleavage plane. Upon entry into mitosis, the bulk of Mid1 translocates out of the nucleus in a Polo-kinase-dependent manner and localizes to protein assemblies called nodes at the equator of the cell. Mid1 then recruits other components of the actomyosin ring to the medial plane of the cell, stimulating equatorial interphase nodes to mature into cytokinesis nodes that will eventually form the actomyosin ring (Pollard and Wu 2010).

Plant Cells

Plant cells do not divide through actomyosin ring constriction but by the continuous deposition of membrane and cell-wall components at a medial plane between the recently divided nuclei. Shortly after mitotic entry, microtubules, microtubule-associated proteins, and actin form a cortical structure, called ▶ pre-prophase band, at the future division site. The pre-prophase band marks the cortex by localization of landmark proteins before it is disassembled in pro-metaphase. Although a number of proteins have been implicated in pre-prophase band formation and function, the precise mechanisms governing their action are still unclear (Müller et al. 2009).

Assembly and Action of the Cytokinesis Apparatus

Animal Cells and Some Lower Eukaryotes

In mammalian cells, actomyosin ring assembly and constriction is regulated by RhoA (see above). Active RhoA localizes to the cortical side of the division plane, where it interacts with formin, promoting assembly of actin filaments. RhoA also interacts with the multidomain protein anillin, which binds actin, myosin, and possibly anchors the actomyosin ring to the cell membrane. Later in anaphase, RhoA activates Rock kinase, which phosphorylates the regulatory chains of myosin, inducing bipolar myosin filament assembly and motor activity. A large number of other

components contribute to actomyosin ring assembly and dynamics, including septins and various actin-binding proteins, such as the Arp2/3 complex and cofilin. Actomyosin ring constriction was initially explained through the "purse string" model whereby bipolar myosin filaments walk along actin filaments to bring about constriction of the ring in a manner similar to muscle constriction. Although there is considerable evidence for this model, other models are also supported by experimental data, including global contraction of the cortex and weakening of resistive forces in the cleavage furrow (Eggert et al. 2006). Alternatively, cell division could come about by global contraction of the cortex and astral microtubule stimulated relaxation at the cell poles distal from the cleavage furrow. The amoeba *Dictyostelium discoideum* usually undergoes cytokinesis in a similar fashion to metazoan cells. However, cells defective in myosin II still form equatorial furrows through a passive mechanism that requires adhesion to a substrate and involves cytoskeleton actin dynamics and cortical tension. This could be explained by models showing that contraction can be generated in absence of motor proteins through a ratchet diffusion based mechanism involving filament cross-linking and bundling proteins (Sun et al. 2010).

Budding Yeast

Cytokinesis in budding yeast occurs at the bud neck, a small connection of about 1 um in diameter between the mother cell and the bud. Septins arrive at the bud neck by late G1 and sequential localization of evolutionary conserved components of the actomyosin ring result in a functional contractile ring by late anaphase. Since the metaphase spindle resides in the mother cell, cytokinesis is halted until one ▶ spindle pole body enters the bud at the end of anaphase. Multiple mechanisms appear to block the mitotic exit network (MEN) from triggering initiation of actomyosin ring constriction until the daughter spindle pole body has passed through the bud neck and into the daughter cell (Balasubramanian et al. 2004). Ingression of the actomyosin ring is closely followed by deposition of a septum. It is thought that because of the high turgor pressure in yeast cells, the actomyosin ring cannot cause constriction on its own, but instead acts to guide the primary septum, which provides most of the force required for constriction.

Fission Yeast

As previously described, actomyosin ring assembly in *S. pombe* is driven by protein complexes called cytokinesis nodes. Actin filaments form from nodes and are cross-linked by myosin, whose contractile activity is thought to bring about condensation of the nodes into a discrete actomyosin ring. Interestingly, the Mid1-dependent pathway only appears to operate prior to anaphase. At anaphase onset, a signaling pathway called the SIN takes over from Mid1 to promote actomyosin ring maturation and stability. SIN mutants form rings in early mitosis through the Mid1-dependent pathway, but the rings fall apart in anaphase. In contrast, cells lacking Mid1 fail to assemble rings in early mitosis, but assemble misplaced rings in late mitosis. In addition, activation of the SIN pathway in interphase can trigger actomyosin ring assembly and constriction independent of Mid1. Together, these results suggest that fission yeast uses two pathways to promote ring assembly; the Mid1 pathway, which positions the ring in early mitosis, and the SIN-dependent pathway, which promotes ring assembly in anaphase (Pollard and Wu 2010).

Plants Cells

During anaphase, most plant cells use the remnants of the mitotic spindle to interdigitate microtubules and actin at the plane previously marked by the pre-prophase band, forming the phragmoplast. Molecular motors then use microtubules to drive Golgi-derived vesicles containing cell-wall components and membrane to the phragmoplast, where the new cell wall is synthesized. The phragmoplast then expands toward the region of the cortex marked by the pre-prophase band (Müller et al. 2009).

Cell Abscission

Abscission is the process by which the membrane connections between daughter cells are severed. Although it is clear that vesicle fusion is important for this step, cell abscission as a whole is still poorly understood, and the various model systems will be dealt with together.

In animal cells, furrow ingression leads to the formation of an intercellular bridge connecting both daughter cells. Inside the bridge, stabilized overlapping microtubules remaining from the central spindle recruit γ-tubulin and other proteins, organizing

into an electron dense structure called the ▶ midbody (Steigemann and Gerlich 2009). Actomyosin ring constriction proceeds perpendicularly to the midbody until both structures interact. At this point, two events must happen at the cellular bridge before the two daughter cells complete cytokinesis: microtubule disassembly/severing and membrane fusion. Microtubules at one side of the midbody are usually severed through a mechanism that is not fully understood but involves the microtubule severing protein spastin. Spastin is recruited to the midbody by the endosomal sorting complex, or ESCRT, which is essential for cell abscission. The thin connection between daughter cells is then sealed by vesicle fusion. Recycling endosomes and exocytic vesicles accumulate at the cellular bridge through interactions with t-SNARE proteins and target-site components of the exocyst complex, recruited at specialized membrane domains at both sides of the midbody. However, the exact nature of the vesicle fusion events at the midbody is not known. In addition, there is evidence that ESCRT proteins can promote cell abscission by self-assembly into spiral-shaped fibers that induce membrane curvature (Steigemann and Gerlich 2009).

Although fission and budding yeast both require a mechanism for final partitioning of the cellular membranes upon contractile ring constriction, evidence for a discrete abscission step has only been observed in budding yeast. However, the mechanisms governing abscission are unclear. Once abscission partitions the cytoplasm of the two daughter cells, cell separation is achieved through degradation of the division septum. This process is tightly regulated to avoid cell lysis and involves glucanases (fission yeast) and chitinases (budding yeast) (Balasubramanian et al. 2004).

In higher plants, vesicle fusion at the division plane results in an outward growth of the cell plate toward the cell cortex. The plane of cell plate growth is oriented by the expanding phragmoplast, which is guided toward the spatial cues left by the preprophase band. How membrane fusion is coordinated to occur all around the cell cortex at once remains a mystery.

Cross-References

- ▶ Actomyosin Ring
- ▶ Cell Cycle
- ▶ Midbody
- ▶ Mitosis
- ▶ Pre-Prophase Band
- ▶ Small GTPases
- ▶ Spindle Pole Body

References

Balasubramanian MK, Bi E, Glotzer M (2004) Comparative analysis of cytokinesis in budding yeast, fission yeast and animal cells. Curr Biol 14:806–818

Eggert US, Mitchinson TJ, Field CM (2006) Animal cytokinesis: from parts list to mechanisms. Annu Rev Biochem 75: 543–566

Guertin DA, Trautmann S, McCollum D (2002) Cytokinesis in eukaryotes. Microbiol Mol Biol Rev 66:155–178

Müller S, Wright AJ, Smith LG (2009) Division plane control in plants: new players in the band. Trends Cell Biol 4:180–188

Oliferenko S, Chew TG, Balasubramanian MK (2009) Positioning cytokinesis. Genes Dev 6:660–674

Piekny A, Werner M, Glotzer M (2005) Cytokinesis: welcome to the Rho zone. Trends Cell Biol 12:651–658

Pollard TD, Wu JQ (2010) Understanding cytokinesis: lessons from fission yeast. Nat Rev Mol Cell Biol 11:149–155

Steigemann P, Gerlich DW (2009) Cytokinetic abscission: cellular dynamics at the midbody. Trends Cell Biol 11:606–616

Sun SX, Walcott S, Wolgemuth CW (2010) Cytoskeletal crosslinking and bundling in motor-independent contraction. Curr Biol 15:649–654

Cytometry, Cytofluorimetry, Microfluorimetry, Analytical Cytology

- ▶ Cell Cycle Analysis, Flow Cytometry

Cytostatic Factor (CSF) Egg Extract

- ▶ Xenopus laevis Egg Extract