H

Hairpin

► Stem-Loop

Hairpin Loop

► Stem-Loop

Hairpin Structure

Yan Zhang
Key Laboratory of Systems Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China

Synonyms

Stem-loop structure

Definition

Hairpin structure is a pattern that can occur in single-stranded DNA or, more commonly, in RNA. The structure is also known as a stem-loop structure. It occurs when two regions of the same strand, usually complementary in nucleotide sequence when read in opposite directions, base-pair to form a double helix that ends in an unpaired loop. The resulting structure is a key building block of many RNA secondary structures.

Characteristics

Formation and Stability

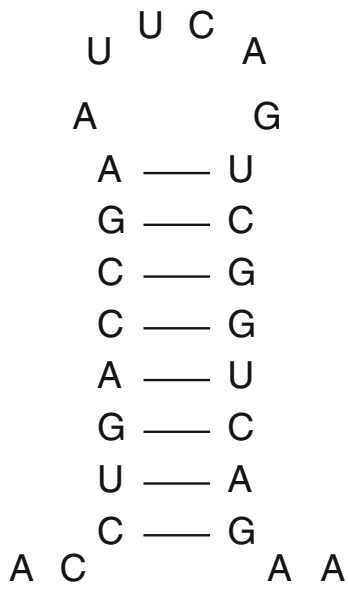
The formation of a hairpin structure is dependent on the stability of the resulting helix and loop regions. The first prerequisite is the presence of a sequence that can fold back on itself to form a paired double helix. The stability of this helix is determined by its length, the number of mismatches or bulges it contains (a small number are tolerable, especially in a long helix), and the base composition of the paired region. Pairings between guanine and cytosine have three hydrogen bonds and are more stable compared to adenine–uracil pairings, which have only two. In RNA, guanine–uracil pairings featuring two hydrogen bonds are as well common and favorable. Base-stacking interactions, which align the pi orbitals of the bases' aromatic rings in a favorable orientation, also promote helix formation. An example of a simple hairpin structure in RNA is shown in Fig. 1.

The stability of the loop also influences the formation of the hairpin structure. "Loops" that are less than three bases long are sterically impossible and do not form. Large loops with no secondary structure of their own (such as pseudoknot pairing) are also unstable. Optimal loop length tends to be about 4–8 bases long. One common loop with the sequence UUCG is known as the "tetraloop" and is particularly stable due to the base-stacking interactions of its component nucleotides.

Structural Contexts

Hairpin structures occur in pre-microRNA structures and most famously in transfer RNA, which contain three true stem-loops and one stem that meet in a cloverleaf pattern. The anticodon that recognizes a

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



Hairpin Structure, Fig. 1 Example of a hairpin structure in RNA

codon during the translation process is located on one of the unpaired loops in the tRNA. Two nested hairpin structures occur in RNA pseudoknots, where the loop of one structure forms part of the second stem.

Many ribozymes also feature hairpin structures. The self-cleaving hammerhead ribozyme contains three stem-loops that meet in a central unpaired region where the cleavage site lies. The hammerhead ribozyme's basic secondary structure is required for self-cleavage activity.

The mRNA hairpin structure forming at the ribosome binding site may control an initiation of translation.

Hairpin structures are also important in prokaryotic rho-independent transcription termination. The hairpin loop forms in an mRNA strand during transcription and causes the RNA polymerase to become dissociated from the DNA template strand. This process is known as rho-independent or intrinsic termination, and the sequences involved are called terminator sequences.

References

Svoboda P, Di Cara A (2006) Hairpin RNA: a secondary structure of primary importance. Cell Mol Life Sci 63(7–8):901–908

Krol J, Krzyzosiak WJ (2006) Structure analysis of microRNA precursors. Methods Mol Biol 342:19–32

Wadkins RM (2000) Targeting DNA secondary structures. Curr Med Chem 7(1):1–15

Half-Life

▶ Life Span, Turnover, Residence Time
▶ Lymphocyte Population Kinetics

Half-life Time

Jean Clairambault
Equipe-Projet BANG, INRIA Paris-Rocquencourt
BP 105, Le Chesnay, Paris, France

Definition

Half-life time ($t_{1/2}$) of a drug in a tissue is the time it takes to decrease by one half from its maximum concentration. If the decay law is exponential with exponent λ, i.e., obeys first-order kinetics with rate λ, i.e., $\frac{dX}{dt} = -\lambda X$, then $t_{1/2} = \frac{\ln 2}{\lambda}$. If the decay curve looks like a sum of two exponential curves, then in a phenomenological perspective it can be modeled as a solution curve to second-order kinetics or equivalently a chain of two first-order kinetics with exponents λ and μ ($\lambda > \mu$), and then pharmacologists define a $t_{1/2\alpha}$ and a $t_{1/2\beta}$ ($t_{1/2\alpha} < t_{1/2\beta}$) measurable from the decay curve, which are nothing but $\frac{\ln 2}{\lambda}$ and $\frac{\ln 2}{\mu}$, respectively. A semiphysiological interpretation can be found to such kinetics, consisting of "distribution" (i.e., fast diffusion) followed by slow elimination (renal, or by binding to plasma proteins). Sometimes are also mentioned three half-life times, $t_{1/2\alpha}$, $t_{1/2\beta}$, and $t_{1/2\gamma}$ when three consecutive episodes are clearly distinguishable in the decay curve.

HAM/TSP

▶ Human T-Lymphotropic Virus Type-I-associated Myelopathytropical Spastic Paraparesis

Hamartoma

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

Definition

Hartoma is a disorganized benign mass of tissue found in the particular site at which it develops and, therefore, considered a developmental malformation.

Cross-References

► Cancer Pathology

Hamilton's Rule

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

Definition

Due to Hamilton (1964), the rule concerns the fixation by natural selection of an altruistic behavior: an act with cost c for its actor and benefit b for its beneficiary will evolve if and only if $c < br$, where r is the ► relatedness of the beneficiary to the focal individual. C and b are measured in terms of ► fitness. Originally supposed to underpin the process of kin selection, it proved to be the general rule for the evolution of cooperation, given that many cases where cooperation emerges among non-kin behave according to such rule; the main reason is that relatedness as such measures a statistical association between individuals rather than a degree of kinship, even if the latter yields ipso facto an association.

Cross-References

► Explanation, Evolutionary

References

Frank SA (1998) The foundations of social evolution. Princeton University Press, Princeton
Hamilton WD (1964) The genetical evolution of social behaviour. J Theor Biol 7:1–52
Van Veelen M (2007) Hamilton's missing link. J Theor Biol 246(3):551–554

Haplotype Map

► HapMap

HapMap

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

Synonyms

Haplotype map; International HapMap project

Definition

The shortened term HapMap (from Haplotype Map) is generally used to refer to the International HapMap project that aims to find genes affecting health, disease, and responses to medications and environmental factors. The information produced by the project is stored in the HapMap database and made freely available to the public (Thorisson et al. 2005).

References

Thorisson GA, Smith AV, Krishnan L, Stein LD (2005) The International HapMap Project Web site. Genome Res 15:1591–1593

Hardy-Weinberg Law

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

Definition

In a panmictic infinite population of diploid sexual individuals with no selection, the frequencies p and (1−p) of two alleles A and a (recessive) at one locus would reach an equilibrium given by the Hardy-Weinberg formula: p^2 AA, 2p (1−p) Aa, $(1-p)^2$ aa. The infinity of the population is requested in order to avoid the effects of genetic drift.

Cross-References

▶ Explanation, Evolutionary

References

Gillespie J (2004) Population genetics – a concise guide. The John Hopkins University Press, Baltimore

Hazard Ratios

Sigrid Rouam
Université Paris-Sud, Villejuif, France

Definition

The hazard ratio is a measure commonly used in survival analysis to compare the risk of occurrence of an event of interest (e.g., death) in two groups (e.g., treatment group vs. control group) at a given time. For example, the hazards ratio can be used to describe the outcome of therapeutic trials where the question is to what extent treatment can delay the apparition of a disease.

The hazard ratio can be calculated using the hazard rate which are defined as

$$\lambda(t) = \lim_{h \to 0+} \frac{\Pr(t \leq T \leq t + h | T \geq t)}{h} \qquad (1)$$

The hazard rate specifies the instantaneous rate at which failures occur for items that are surviving at time t and gives the risk of failure per unit time.

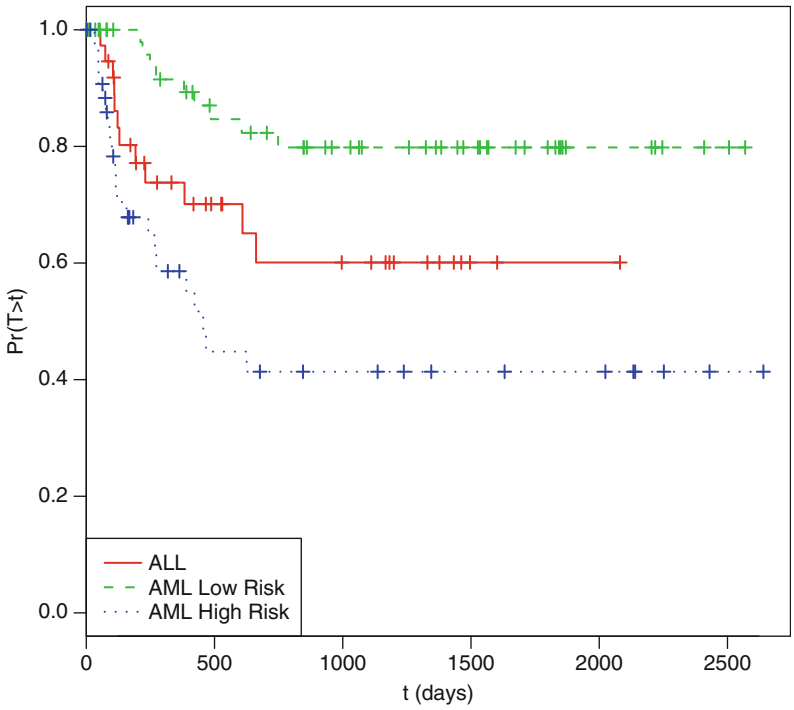
The hazard ratio is simply equal to $\mathrm{HR} = \frac{\lambda(t|\text{group 1})}{\lambda(t|\text{group 2})}$. If $\mathrm{HR} = 1$, the risk of occurrence of the event of interest is the same in the two groups of patients. If $\mathrm{HR} > 1$ ($\mathrm{HR} < 1$), the risk of occurrence of the event of interest is higher in group 1 (in group 2, respectively).

The hazard ratio can be calculated at different points in time. The hazard ratio differs from the relative risk in the sense that this latter is calculated over the whole period of follow-up and not at a given time t.

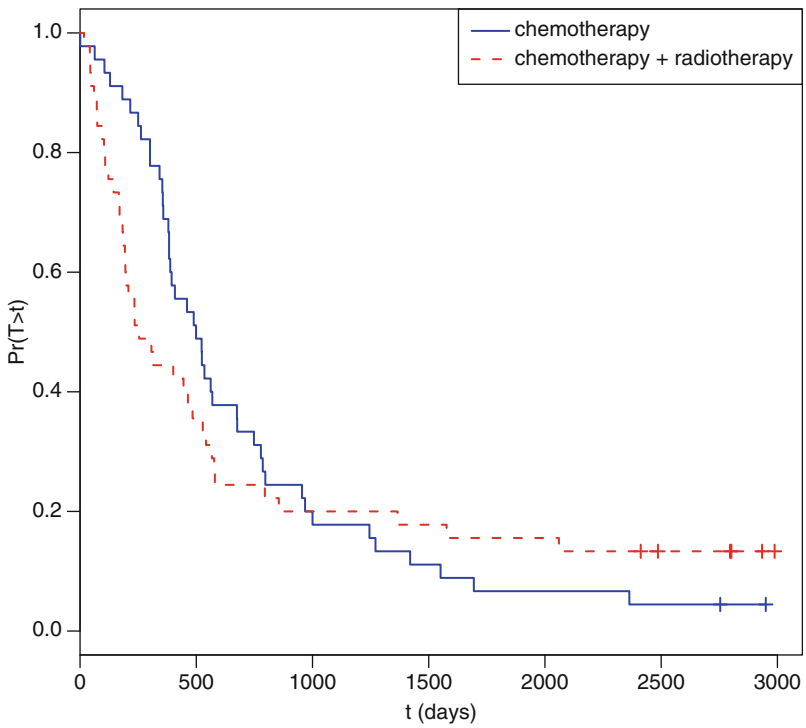
When the hazard ratio is constant over time, the hazards are said to be proportional. This is the assumption made in proportional hazards models.

An example of proportional hazards ratio is given in Fig. 1. It represents the disease-free survival curves of patients with acute leukemia classified into three risk categories (Copelan et al. 1991). These categories were defined according to the status of the patients at the time of transplantation as follows: acute lymphoblastic leukemia (ALL) (38 patients), acute myelocytic leukemia (AML) low risk (54 patients), and AML high risk (45 patients). The three curves are parallel to each other showing that the hazards are constant over time and thus proportional. Patients in group AML high risk have the greatest chance of failure. Patients in group AML low risk have the best prognosis.

Figure 2 displays an example of non-proportional hazards ratio. It represents the survival curve of gastric cancer patients receiving two different treatments: chemotherapy (45 patients) and chemotherapy plus radiotherapy (45 patients) (Stablein and Koutrouvelis 1985). Clearly, the hazards are not proportional. The figure shows that, at the beginning, patients receiving chemotherapy have a better prognosis than patients with chemotherapy plus radiotherapy. But after 2.7 years, the survival functions of the two groups intersect, and patients with chemotherapy plus radiotherapy have a better prognosis.



Hazard Ratios,
Fig. 1 Disease-free survival for the 137 bone marrow transplant patients



Hazard Ratios,
Fig. 2 Survival for the 90 gastric cancer patients

References

Copelan EA, Biggs JC, Thompson JM, Crilley P, Szer J, Klein JP, Kapoor N, Avalos BR, Cunningham I, Atkinson K, Downs K, Harmon GS, Daly MB, Brodsky I, Bulova SI, Tutschka PJ (1991) Treatment for acute myelocytic leukemia with allogeneic bone marrow transplantation following preparation with BuCy2. Blood 3(1):838–843

Stablein DM, Koutrouvelis IA (1985) A two-sample test sensitive to crossing hazards in uncensored and singly censored data. Biometrics 41(3):643652

Health Informatics

Guy Tsafnat
Centre for Health Informatics, Australian Institute for Health Innovation, University of New South Wales, Sydney, NSW, Australia

Definition

Health Informatics is a broad concept that covers all aspects of the use of information technology (IT) in the provision and implementation of health care (Coiera 2003). This includes the use of IT to:

- Facilitate communication at any level of a health care system, from the communication between clinicians and patients all the way to share clinical records from multiple electronic health record systems in a large-scale study
- Implement and maintain electronic health record management systems, ensuring appropriate privacy and access to the data
- Implement and maintain monitoring and reporting systems to ensure health service quality and patient safety
- Develop decision support systems for clinicians, clinical researchers, and translational researchers

Health Informaticians come from a range of background including social sciences, psychology, human-computer-interaction and ergonomics, communication science and linguistics, ethnography and information technology. As electronic health record management systems become more prevalent, health informatics becomes more central to health systems enabling better monitoring and quality control, improve decision making, and communication.

References

Coiera E (2003) Guide to health informatics. Hodder, London

Helper T Cell 1 Response

► Th1 Response

Helper T Cell 2 Response

► Th2 Response

Heterochromatin

Vani Brahmachari and Shruti Jain
Dr. B. R. Ambedkar Center for Biomedical Research, University of Delhi, Delhi, India

Synonyms

Closed chromatin; Compact chromatin

Definition

The heterochromatin is the compacted, tightly packed chromatin which is inaccessible to the transcription machinery and is intensely stained with DNA binding dyes. Two types of heterochromatin are present in cells: constitutive heterochromatin in the centromeres that largely remains transcriptionally silent in all cells and facultative heterochromatin which is transiently silenced. Besides, being highly compacted, heterochromatin is enriched in epigenetic modifications of ► histone and DNA associated with transcriptional inactivity.

Cross-References

► Epigenetics
► Histones

Heterochromatin and Euchromatin

Yota Murakami
Department of Chemistry, Hokkaido University, Sapporo, Japan

Synonyms

Silent chromatin and active chromatin

Definition

Heterochromatin and euchromatin are two major categories of chromatin higher order structure. Heterochromatin has condensed chromatin structure and is inactive for transcription, while euchromatin has loose chromatin structure and active for transcription. Heterochromatin is further divided into two subcategories: constitutive and facultative heterochromatin. Heterochromatin and euchromatin are defined by specific histone modifications. Since heterochromatin can spread into neighboring euchromatic region and repress gene expression, it is important to regulate boundaries between euchromatin and heterochromatin. Generally, the balance of euchromatic and heterochromatic histone-modifying enzymes determines the boundary. At particular region, sequence specific elements and their binding proteins define the boundary. In addition, the boundary elements determine the domain structure of genome that is important for wide range of regulation of transcription through association of nuclear structure or self-association.

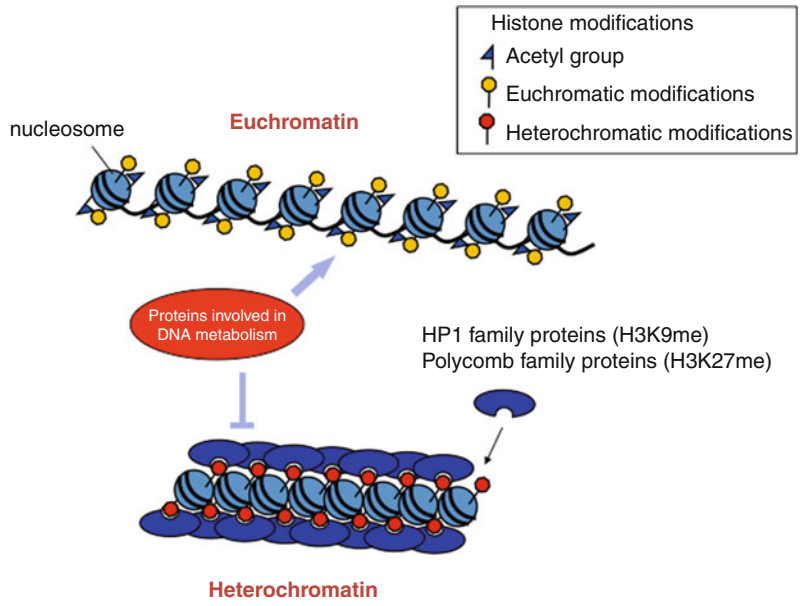
Characteristics

▶ Chromatin forms various higher order structure and the structure is classified into two categories; ▶ heterochromatin and ▶ euchromatin. Heterochromatin has condensed chromatin structure: ▶ nucleosomes are regularly positioned and packed tightly. In contrast, euchromatin has loose nucleosomes structure and the extent of the packaging of nucleosomes varies from region to region and dynamically regulated.

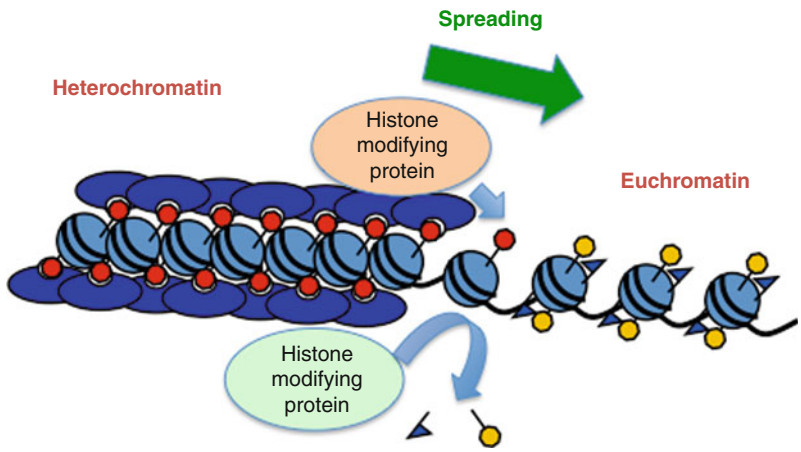
Heterochromatin is further divided into two categories, constitutive and facultative heterochromatin. Constitutive heterochromatin is observed in almost all eukaryotic cells and localized at repeated sequences and transposons and also localized at ▶ centromere and telomere, where repeated sequences are enriched. Constitutive heterochromatin is stably maintained and suppresses transcription and recombination of the embedded DNA sequences. Facultative heterochromatin is observed in higher eukaryote and represses transcription of protein coding genes during many cell generations. Formation of facultative heterochromatin is developmentally regulated.

The chromatin higher order structures are defined by covalent histone modifications (▶ Histone Post-translational Modification to Nucleosome Structural Change) (Fig. 1). Heterochromatin is generally hypo-acetylated. Constitutive heterochromatin is defined by methylation of lysine 9 of histone H3 and its binding proteins, ▶ heterochromatin protein 1 family proteins. Facultative heterochromatin is defined by methylation of lysine 27 of histone H3 and its binding partner, polycomb family proteins (▶ Polycomb Complexes). In contrast, euchromatin is generally hyper-acetylated and avoids methylation of histone H3 lysins 9 and 27 and has various states of histone modifications that tightly correlate with the transcription activities. Note that heterochromatin of budding yeast does not have the methylation mark of the histones and is defined by hypo-acetylated state and Sir proteins.

Heterochromatin is "inactive" chromatin, which prevent DNA metabolism such as transcription and recombination (Fig. 1). The basis of the inactiveness has been thought the tight packaging of the nucleosome array, which prevents access of enzymes promoting the DNA metabolism. However, recent study suggests that not only the condensed structure but also recruitment of "effector" proteins to heterochromatin by heterochromatin proteins are responsible for repression of the transcription (Grewal and Jia 2007). Intriguingly, an effector that activates transcription is also recruited to heterochromatin, suggesting that transcription in the heterochromatin could be actively regulated (Grewal and Jia 2007). One of the mechanisms that regulates the recruitment of the effectors are post-translational modifications of heterochromatin proteins, including phosphorylation, but the precise mechanism is not clear yet (Shimada and Murakami 2010). Hence, heterochromatin is a dynamic chromatin structure rather than an "inactive" static chromatin structure.



Heterochromatin and Euchromatin, Fig. 1 Heterochromatin and euchromatin

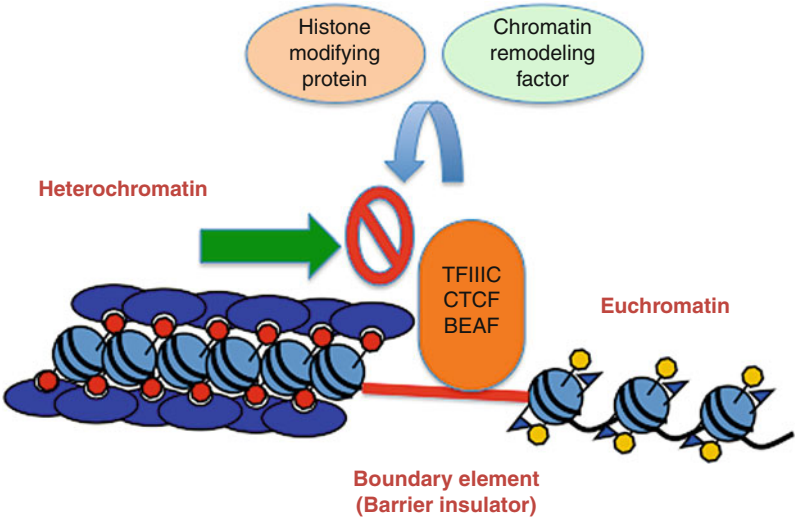


Heterochromatin and Euchromatin, Fig. 2 Spreading of Heterochromatin

Similarly, various combinations of histone modifications in euchromatin attract various kinds of proteins that regulate transcription, resulting in wide range of transcriptional states, from silent state to active state.

One characteristic of heterochromatin is "spreading"; heterochromatin can autonomously spread into neighboring euchromatic region, resulting in ▶ position effect variegation of neighboring genes. The exact mechanism of spreading is not clear yet, but a simple model is widely accepted that heterochromatin protein recruits histone-modifying enzymes via protein-protein interaction, which convert neighboring euchromatic nucleosome to heterochromatic state (Fig. 2). To avoid the uncontrolled silencing of the gene expression by spreading of heterochromatin, the "boundaries" between euchromatin and heterochromatin should be strictly regulated.

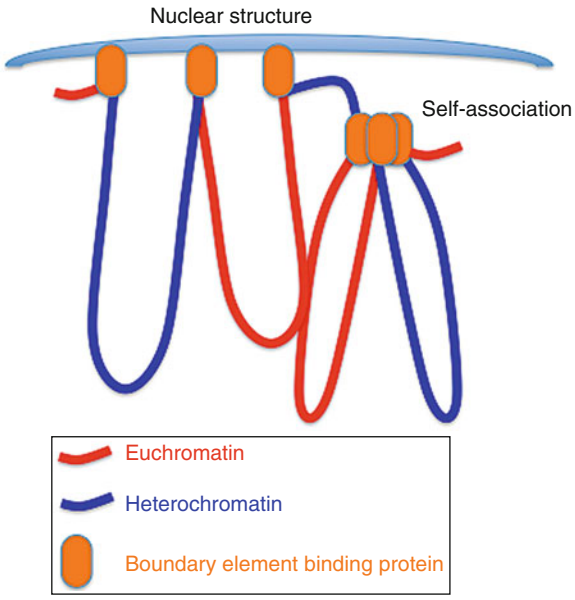
Heterochromatin and euchromatin are determined by histone modification. Therefore, primary determinant of the boundaries would be the competition between the heterochromatic modifying enzymes and the euchromatic ones. Indeed, the competitive determination mechanism is observed at the boundary



Heterochromatin and Euchromatin, Fig. 3 Boundary elements/ barrier insulators

between telomeric heterochromaitn and its neighboring euchormatin in budding yeast (Kimura et al. 2002).

Since competitive determination of the boundaries results in fluctuation of the boundaries, at certain region of the genome, the boundary should be tightly regulated to prevent the stochastic silencing of genes close to heterochromatin by spreading. Some DNA elements have been shown to have the barrier function against heterochromatin spreading in various organisms and are called boundary elements or barrier insulators (Fig. 3). The boundary elements exhibit the barrier function through their binding proteins. In budding yeast, tRNA genes, which are transcribed by RNA polymerase III (► RNA Polymerase), prevent the spreading of heterochromatin (Lunyak 2008; Sun et al. 2011). The proteins assembled onto the tRNA (Sun et al. 2011), including RNA polymerase III and its regulatory proteins are required for the barrier function. In fission yeast, tRNA genes also act as the boundary element (Scott et al. 2006). Interestingly, TFIIIC, a tRNA specific transcription factor, act as the boundary protein without tRNA transcription in fission yeast (Noma et al. 2006). In higher eukaryote, several DNA sequence specific DNA binding proteins, such as BEAF in *Drosophila* and CTCF in mammals, are shown to act as boundary proteins. These proteins are thought to recruit specific proteins including chromatin remodeling factors or histone-modifying enzymes, to establish the barrier against heterochromatin spreading (Fig. 3) (Lunyak 2008; Sun et al. 2011).



Heterochromatin and Euchromatin, Fig. 4 Chromatin domains and boundary elements

Search for the proteins that shows barrier activity when tethered close to heterochromatin identified ► nuclear pore proteins (Ishii et al. 2002). This suggests that anchoring of genomic region to nuclear structure including nuclear envelope is one of the determinants to establish the boundary. Supporting this idea, some of the boundary proteins associate with nuclear structure

(Vogelmann et al. 2011). In addition, tRNA genes as well as the boundary proteins including TFIIIC form cluster in nucleus (Noma et al. 2006). The association with nuclear structure or self-clustering of the boundary elements results in the formation of chromatin loop, and this chromatin loop defines silent heterochromatic domain and active euchromatic domain of the genome. Therefore, the boundary elements regulate not only the boundary between heterochromatin and euchromatin but also domain structure of chromatin in the nucleus, which might be important for regulation of whole genome organization (Vogelmann et al. 2011) Fig. 4.

Cross-References

- ▶ Centromere
- ▶ Chromatin
- ▶ Euchromatin
- ▶ Heterochromatin
- ▶ Heterochromatin Protein 1
- ▶ Histone Post-translational Modification to Nucleosome Structural Change
- ▶ Nuclear Pore
- ▶ Nucleosomes
- ▶ Polycomb Complexes
- ▶ Position Effect Variegation
- ▶ RNA Polymerase

References

Grewal SI, Jia S (2007) Heterochromatin revisited. Nat Rev Genet 8:35–46

Ishii K, Arib G, Lin C, Van Houwe G, Laemmli UK (2002) Chromatin boundaries in budding yeast: the nuclear pore connection. Cell 109:551–562

Kimura A, Umehara T, Horikoshi M (2002) Chromosomal gradient of histone acetylation established by Sas2p and Sir2p functions as a shield against gene silencing. Nat Genet 32:370–377

Lunyak VV (2008) Boundaries. Boundaries. . .Boundaries??? Curr Opin Cell Biol 20:281–287

Noma KI, Cam HP, Maraia R, Grewal SIS (2006) A role for TFIIIC transcription factor complex in genome organization. Cell 125:859–872

Scott K, Merrett S, Willard H (2006) A heterochromatin barrier partitions the fission yeast centromere into discrete chromatin domains. Curr Biol 16:119–129

Shimada A, Murakami Y (2010) Dynamic regulation of heterochromatin function via phosphorylation of HP1-family proteins. Epigenetics 5:30–33

Sun JQ, Hatanaka A, Oki M (2011) Boundaries of transcriptionally silent chromatin in *Saccharomyces cerevisiae*. Genes Genet Syst 86:73–81

Vogelmann J, Valeri A, Guillou E, Cuvier O, Nollmann M (2011) Roles of chromatin insulator proteins in higher-order chromatin organization and transcription regulation. Nucleus 2:358–369

Heterochromatin Protein 1

Yota Murakami
Department of Chemistry, Hokkaido University, Sapporo, Japan

Synonyms

HP1

Definition

Heterochromatin protein 1 (HP1) is widely conserved protein that localized at heterochromatin. HP1 is a main component of constitutive heterochromatin that localizes on repeated sequences including centromere and telomere. HP1 contains two distinct domains, chromo domain and chromo-shadow domain. Chromo domain of HP1 specifically recognizes di- or trimethylated lysine 9 of histone H3. Chromo-shadow domain promotes dimerization of HP1, which is thought to be responsible for the packaging of nucleosomes in heterochromatin. In addition, many proteins interact with HP1 through chromo-shadow domain (Grewal and Jia 2007). Those proteins play various roles in the regulation of heterochromatin function (Fanti and Pimpinelli 2008).

Cross-References

- ▶ Heterochromatin and Euchromatin

References

Fanti L, Pimpinelli S (2008) HP1: a functionally multifaceted protein. Curr Opin Genet Dev 18:169–174

Grewal SI, Jia S (2007) Heterochromatin revisited. Nat Rev Genet 8:35–46

Heuristic Optimization

Feng-Sheng Wang and Li-Hsunan Chen
Department of Chemical Engineering, National Chung Cheng University, Chiayi, Taiwan

Definition

Heuristic designates a computational procedure that determines an optimal solution by iteratively trying to improve a candidate solution with regard to a given measure of quality. Heuristics make few or no assumptions about the problem being optimized and can search large spaces of candidate solutions toward finding optimal or near-optimal solutions at a reasonable computational cost without being able to guarantee either feasibility or optimality, or even in many cases to state how close to optimality a particular feasible solution is. Heuristics implement some form of stochastic search optimization, such as ▶ evolution programming, evolution strategy, ▶ genetic algorithms, genetic programming, and differential evolution (Michalewicz 1996; Reeves 1995; Sharda et al. 2003; Zhilinskas and Žilinskas 2008). Other methods having a similar meaning as heuristic are derivative-free, direct search, and black-box optimization techniques.

Characteristics

Heuristic optimization algorithms are developed in all kinds of forms variant from simple "trial and error" to complicated algorithms as evolutionary algorithms. The methods are easy to understand and easy to implement and use. The mathematical formulation of the problem is flexible. Heuristic optimization can be applied to iteratively solve continuous/integer problem. They are usually applied when there is no known algorithm that guarantees for finding the optimal solution in efficient computational cost (i.e., time or memory space) or when a "near-optimal" solution is good enough in practical use.

The common advantages of heuristic optimization algorithms are as follows:

1. *Fast*: They can find a "near-optimal" solution in a short time.
2. *Small*: They can work in a relatively small memory space.

The common disadvantages of heuristic optimization algorithms are as follows:

1. *Not absolute optimal solution*: Heuristic algorithms cannot guarantee to find the optimal solution.
2. *Uncertainty*: The time required for finding a "near-optimal" solution can be large in an unlucky case.

Cross-References

- ▶ Evolution Programming
- ▶ Genetic Algorithms
- ▶ Particle Swarm Optimization (PSO)
- ▶ Simulated Annealing
- ▶ Tabu Search

References

Michalewicz Z (1996) Genetic algorithms + data structures = evolution programs. Springer, Berlin
Reeves CR (1995) Modern heuristic techniques for combinatorial problems. McGraw-Hill, Londres
Sharda R, Voß S, Woodruff DL, Fink A (2003) Optimization software class libraries. Springer, New York, pp 81–154
Zhilinskas A, Žilinskas A (2008) Stochastic global optimization. Springer, New York

Heuristic Search

Long Jason Lu[1] and Minlu Zhang[2]
[1]Division of Biomedical Informatics, Cincinnati Children's Hospital Research Foundation, Cincinnati, OH, USA
[2]Department of Computer Science, University of Cincinnati, Cincinnati, OH, USA

Definition

Heuristic search refers to a search strategy that attempts to optimize a problem by iteratively improving the solution based on a given heuristic function or a cost measure. A heuristic search method does not always guarantee to find an optimal or the best solution, but may instead find a good or acceptable solution within a reasonable amount of time and memory space. Several commonly used heuristic search methods include hill climbing methods, the best-first search,

the A* algorithm, simulated-annealing, and genetic algorithms (Russell and Norvig 2003). A classic example of applying heuristic search is the traveling salesman problem (Russell and Norvig 2003).

Cross-References

- ▶ Biological Applications of Network Modules
- ▶ Modules in Networks, Algorithms and Methods

References

Russell SJ, Norvig P (2003) Artificial intelligence: a modern approach. Prentice Hall, Upper Saddle River

Hfq

Tatsuhiko Someya[1], Nobukazu Nameki[2] and Gota Kawai[3]
[1]University of Tsukuba, Tsukuba, Japan
[2]Gunma University, Kiryu, Japan
[3]Department of Life and Environmental Sciences, Chiba Institute of Technology, Narashino, Chiba, Japan

Synonyms

Host factor 1

Definition

Escherichia coli Hfq is a 102 residues protein that is first identified as a host factor of the RNA phage Qβ (Chao and Vogel 2010). Hfq orthologous proteins are conserved among approximately half of all sequenced Gram-positive and Gram-negative bacteria (Brennan and Link 2007). Hfq protein forms homohexameric ring and binds to poly(A) and single-stranded AU-rich RNAs at two distinct binding surfaces (Brennan and Link 2007). In Gram-negative bacteria, Hfq promotes base-pairing between sRNA and its target mRNA by inducing the RNA structural changes acting as an RNA chaperone (Jousselin et al. 2009). The sRNA–mRNA interactions are known to regulate the expression of several genes (Jousselin et al. 2009).

References

Brennan RG, Link TM (2007) Hfq structure, function and ligand binding. Curr Opin Microbiol 10:125–133
Chao Y, Vogel J (2010) The role of Hfq in bacterial pathogens. Curr Opin Microbiol 13:24–33
Jousselin A, Metzinger L, Felden B (2009) On the facultative requirement of the bacterial RNA chaperone. Hfq Trends Microbiol 17:399–405

Hierarchic Mechanisms

- ▶ Mechanism, Multilevel

Hierarchical Agglomerative Clustering

Marie Lisandra Zepeda-Mendoza and Osbaldo Resendis-Antonio
Center for Genomics Sciences-UNAM, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, Mexico

Synonyms

Agglomerative hierarchical clustering; Agglomerative hierarchical data segmentation; Bottom-up hierarchical clustering

Definition

Cluster analysis consists to classify a set of objects (observations, individuals, cases) into subsets, called clusters, such that they have similar characteristics or properties. There are different ways to define the similarities among objects or variables through the use of different metrics. Some of them are as follows:

- The single-linkage clustering, or nearest neighbor clustering, takes into account the shortest distance of the distances between the elements of each cluster. This is one of the simplest methods.
- The complete linkage clustering, or farthest neighbor clustering, takes the longest distance between the elements of each cluster.

- The average linkage clustering takes the mean of the distances between the elements of each cluster. The merged clusters are the ones with the minimum mean distance.

There are a variety of clustering algorithms; one of them is the agglomerative hierarchical clustering. This clustering method helps us to represent graphically the results through a dendogram. The dendogram has a tree structure that consists of the root and the leaves; the root is the cluster that has all the observations, and the leaves are individual observations. The agglomerative hierarchical clustering starts with the individual observations and successively fuses the clusters that are closer together (the most similar ones).

Cross-References

▶ Modules, Identification Methods and Biological Function

References

Jain AK, Dubes RC (1998) Algorithms for clustering data. Prentice Hall, Englewood Cliffs

Jain AK, Murthy MN, Flynn PJ (1999) Data clustering: a review. ACM Comput Rev 31(3):264–323

Hierarchical Models

▶ Mixed and Multi-Level Models

Hierarchical Modularity

▶ Hierarchical Structure

Hierarchical Organization

▶ Hierarchical Structure

Hierarchical Structure

Marie Lisandra Zepeda-Mendoza and Osbaldo Resendis-Antonio
Center for Genomics Sciences-UNAM, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, Mexico

Synonyms

Hierarchical modularity; Hierarchical organization

Definition

A hierarchical structure in a network context is characterized by being a topological structure in which nodes are placed in different layers; and in these layers there can be small and highly connected modules. The nodes in one layer collect influences from each other and from the nodes of the superior layer, while also being able to influence nodes of the inferior layer.

The transcriptional regulatory network of *Escherichia coli* shows a clear example of this view of hierarchical structure. The nodes are organized in layers where every node of one layer can receive inputs from nodes in upper layers and can be linked to nodes in lower layers (Resendis-Antonio et al. 2005).

In metabolic networks, the high size-independent clustering coefficient (which is evidence for modularity) and the power law degree distribution (which supports the scale-free model that would rule out modular topology) pose an apparent contradiction that can be solved by proposing the existence of metabolic hierarchical modules (Ravasz et al. 2002). This model presents a $C(k) \sim k^{-1}$, and this property can be used as a quantitative indicator of hierarchy.

To better understand this model, imagine a starting point of a small cluster of four densely connected nodes. Next generate three replicas of the starting module and connect them to the central node of the first starting cluster, obtaining a large 16-node module made of 4 smaller modules. Then generate 3 replicas of this 16-node module and connect the peripheral nodes

to the central node of the first starting cluster. These steps of replication and connection can be repeated indefinitely (Ravasz et al. 2002).

Cross-References

- ▶ Hierarchy
- ▶ Modules, Identification Methods and Biological Function
- ▶ Module Network

References

Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabasi AL (2002) Hierarchical organization of modularity in metabolic networks. Science 297:1551

Resendis-Antonio O, Freyre-Gonzalez J, Menchaca-Mendez R, Gutierrez-Rios RM, Martinez-Antonio A, Avila-Sanchez C, Collado-Vides J (2005) Modular analysis of the transcriptional regulatory network of *E. coli*. Trends Genet 21:16

Hierarchy

Shihua Zhang
National Center for Mathematics and Interdisciplinary Sciences, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China

Synonyms

Hierarchical structure

Definition

Hierarchy is defined as an arrangement of items or simply an ordered set. In networks, hierarchy describes the hierarchical relationship among nodes (Ravasz et al. 2002).

References

Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabási AL (2002) Hierarchical organization of modularity in metabolic networks. Science 297(5586):1551–1555

High Throughput Computing

- ▶ Grid Computing, Parameter Estimation for Ordinary Differential Equations

Highly Structured System

- ▶ Complex System

High-Performance Computing

- ▶ Large-Scale and High-Performance Computing
- ▶ Multicore Computing

High-Performance Computing, Structural Biology

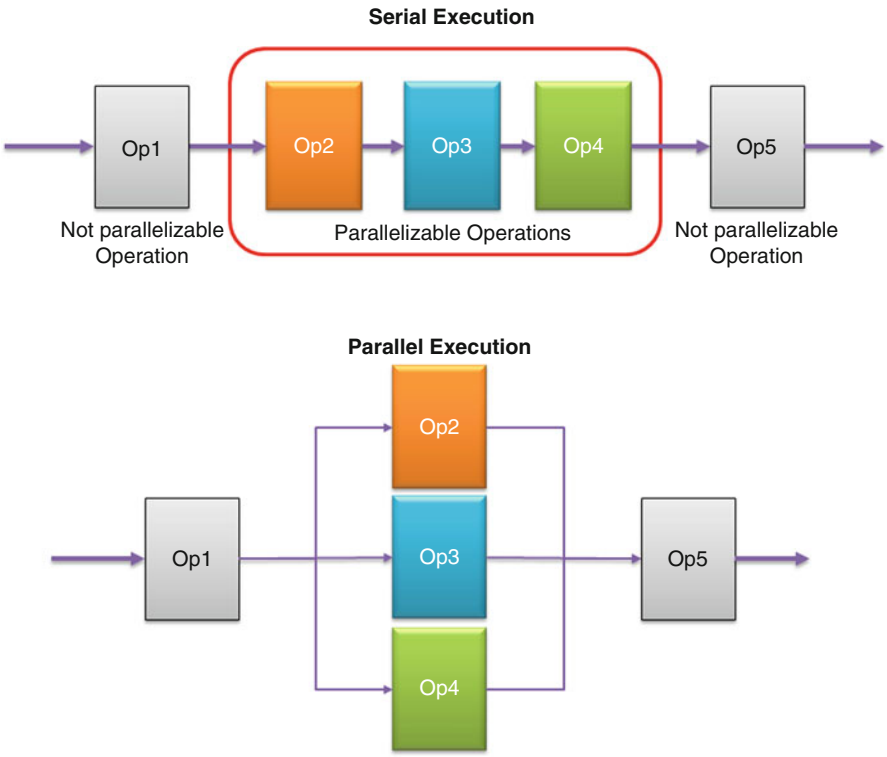
Piercarlo Dondi and Luca Lombardi
Department of Computer Engineering and Systems Science, University of Pavia, Pavia, Italy

Synonyms

Parallel computing

Definition

High Performance Computing (HPC) refers to technologies used for implementing systems able to execute time expensive elaborations and to manage a huge amount of data in a small amount of time. HPC solutions are commonly exploited in different scientific fields that require the solution of complex mathematical models, like climatology, physics, medicine, or biology. One of the most recent innovations, which presents a good compromise between hardware cost and performances, is the use of the ▶ GPU for parallel computation. This technology



High-Performance Computing, Structural Biology, Fig. 1 Serial versus parallel computation, the efficiency of parallelization depends on the relative portion of parallelizable operations with respect to the total (Eq. 3).

supplies promising results in simulation and modeling of biological systems and in real-time medical analysis.

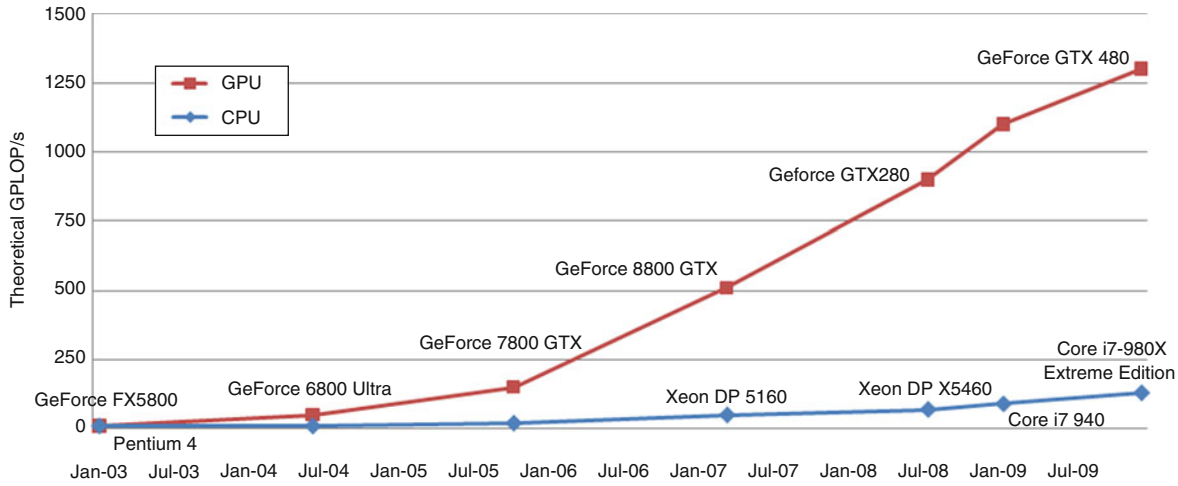
Characteristics

Parallel Computing

HPC (▶ Large-Scale and High-Performance Computing) comes from the need of more and more great computational power to elaborate complex mathematical models and to evaluate new scientific theories. Technical and physical constraints limit the maximum speed reachable by a single CPU, so all HPC solutions involve the use of parallel computing resources, in which multiple processing units cooperate to complete a task in a time as small as possible. Not all types of problems can be parallelized (e.g., a computation in which each step depends on the outcomes of the previous one) and also if the parallelization is possible, some conditions must be satisfied to work properly. It is crucial to redesign the program (the so-called porting of the application) in a parallel way: the processors must operate independently and have a workload equally distributed between them; the computational time has the priority, time needed for the communications; and the synchronization between processors and for data transfer has to be reduced to the minimum (Culler et al. 1998). Figure 1 shows an example of serial and parallel execution of a simple algorithm.

Nearly all situations in which a parallel implementation is useful can be reduced to two cases: a general case in which different independent instructions can be executed on different data, or a single operation that must be executed on a large amount of data (e.g., meteorological analysis – each unit examines in the same way a small portion of a geographic area) (Hord 1998). In most cases, biological and medical simulations fall into the latter category.

The performance of a parallel algorithm are usually measured by two related indexes: the *speedup* (Eq. 1), that is, the ratio between the execution time with one processor and that with N processors, and the *efficiency* (Eq. 2), that is, the ratio between the speedup and the number of processors:



High-Performance Computing, Structural Biology, Fig. 2 Growing rate of computational power for CPU and GPU in terms of single precision floating point operations per second – based on data presented in Kirk and Hwu (2010)

$$S(N) = \frac{T(1)}{T(N)} \tag{1}$$

$$E(N) = \frac{S(N)}{N} \tag{2}$$

Their theoretical best values are, respectively, $S(N) = N$ and $E(N) = 1$, but actually these indexes depend on the fraction of the algorithm that is effectively parallelizable. Amdahl's law (Eq. 3) defines the maximum reachable speedup: P is the parallelizable part of the code, where smaller is P the farther the speedup from the optimal value. In the worst case an increase of the number of processors N generates a small (or null) enhancement of the performance (Culler et al. 1998):

$$S(N) \leq \frac{1}{(1-P) + \frac{P}{N}} \tag{3}$$

The available hardware for parallel computing is heterogeneous: multi-core processors, distributed computing, clusters, ▶ grid computing, FPGAs, and, more recently, GPUs.

GPGPU

The need of realistic and detailed 3D models and environments in very different fields, from science to entertainment, generated a growth rate of the computing power of GPUs incredibly high, higher than that of the CPU (Fig. 2) – actually a GPU is a multi-core device composed by hundreds of simple processor units.

Between 2001 and 2002, a crucial enhancement in the architecture of the graphic cards was achieved: the introduction of programmable modules. This evolution gave the capability of a more complex interaction with the GPU, and allowed to employ a series of realistic graphic effects unreachable with the precedent models based on fixed steps and instructions (Kirk and Hwu 2010). The high level of flexibility lets the programmers consider a new series of applications that are not directly involved the graphic, but could successfully exploit the great computational power of graphic hardware (e.g., matrices calculations or convolutions). The research field that studies the use of GPUs for parallel computation is called GPGPU (General-Purpose computing on GPU).

Even if first experiments performed promising results and a good speedup, they were still limited by architectural constraints and by the complexity of a low-level programming model. The interest of scientific community in this area (in optimal conditions a GPU can increase the performances hundreds of times with respect to a CPU) led the main developers of graphic cards (ATI and Nvidia) to the creation of new models completely programmable and enabled for

parallel computing. At the end of 2006, Nvidia introduced CUDA (Computer Unified Device Architecture) the first GPU architecture designed for allowing parallel computing (Kirk and Hwu 2010). Instead, for its hardware, ATI supplied Stream technology that adopts the standard OpenCL (Open Computing Languages) (Munshi et al. 2011) developed by Khronos Group (a no-profit consortium that includes up to now ATI and Nvidia). At the moment CUDA directly supports specific extensions for the most common programming languages (e.g., C/C++, Fortran, or Python) and it is the most diffused technology for GPGPU. Nvidia also provides a particular category of graphic card, the Tesla series, created expressly for HPC. With respect to a traditional GPU, a Tesla card renounces at the video output for high reliability, for a major quantity of memory, and for the capability of a continued and intensive use. In November 2010, three of the top five supercomputers in the world (including the first) employ Tesla cards.

The best performance in parallel computing with a GPU can be achieved using it as a sort of stream processor, which means that a GPU is most efficient if the same operation is applied on a large amount of independent data. Typical applications that can take advantages of graphic acceleration involve image elaboration (e.g., filtering or ▶ mathematical morphology operators), video processing, wheatear forecasting, geological modeling, fluid dynamics, medical imaging, molecular modeling, and biological structure simulations.

Application Examples

The GPU computing allows the researchers to use desktop computers for complex computations that up to now were a prerogative of large computer clusters. In consequence the integration of graphic cards in a traditional cluster can supply the computing power needed to cope with more complex classes of problems. The state of art and the practical impact of HPC with GPU in biology and medicine can be illustrated by the following examples.

NAMD (Nanoscale Molecular Dynamics) and VMD (Visual Molecular Dynamics), developed by the University of Illinois at Urbana-Champaign (UIUC), are two widely used tools for simulating and visualizing biomolecular processes and interactions. The UIUC researchers ported on TESLA the most computationally intense parts of their tools obtaining significant enhancements. The visualization of molecular orbitals (MOs) generally required up to hundreds of seconds on a CPU; with the CUDA implementation a high-quality rendering is achievable in less than a second (Stone et al. 2009). This time reduction allowed the creation of the first interactive animations of quantum chemistry simulation trajectories using real-time computation. Another example involves the analysis of the interaction of biological molecules and ions. The ion placement can be very computationally demanding: large structures, such as viruses, could need several days even using moderately sized clusters of computers. However the independency of data makes this problem ideal for the porting on GPU, for example, the Coulomb-based ion placement can reach a speedup of 100 times or more. The implementation developed by UIUC reduces the time to generate large ionized structures to a few minutes on a single desktop computer (Stone et al. 2007).

In modern medical imaging, one of the most important goals is the production of highly detailed images in a short period of time, in particular for human scanning, to be able to give more quick diagnosis.

Techniscan Medical Systems recently created a new system for ultrasound scanning, the Whole Breast Ultrasound (WBU™). However, even with a cluster of 16 computers equipped with the latest Xeon processors, the procedure takes too much time to examine numerous patients in a day. A new approach, that employs four Tesla GPUs, is able to run Techniscan's algorithm in less than 20 min, less than half the time taken by the cluster (Hardwick 2009). This speedup allows radiologists to perform a complete ultrasound scan and to see the results during a 30-min patient visit. Also by the economic point of view the hardware cost of a GPU solution is cheaper than adopting a traditional cluster.

Cross-References

- ▶ GPU
- ▶ Grid Computing, Parallelization Techniques
- ▶ Grid Computing, Parameter Estimation for Ordinary Differential Equations
- ▶ Mathematical Morphology Operators

References

Culler D, Singh JP, Gupta A (1998) Parallel computer architecture: a hardware/software approach. Morgan Kaufmann, San Francisco

Hardwick J (2009) Embedded Tesla-using CUDA and Tesla in a medical device. In: GTC09, GPU technology conference 2009, 30 Sept–2 Oct, San Josè

Hord RM (1998) Understanding parallel supercomputing. Piscataway, IEEE

Kirk D, Hwu W (2010) Programming massively parallel processors: a hands-on approach. Morgan Kaufmann, Burlington

Stone JE, Phillips JC, Freddolino PL, Hardy DJ, Trabuco LG, Schulten K (2007) Accelerating molecular modeling applications with graphics processors. J Comput Chem 28:2618–2640

Stone JE, Saam J, Hardy DJ, Vandivort KL, Hwu WW, Schulten K (2009) High performance computation and interactive display of molecular orbitals on GPUs and multi-core CPUs. In: 2nd workshop on general-purpose processing on graphics processing units, ACM international conference proceeding series, vol 383. ACM, New York, pp 9–18

Munshi A, Gaster B, Mattson TG, Fung J, Ginsburg D (2011) OpenCL programming guide. Addison-Wesley Professional, 1st edition, July 2011, Boston

High-Throughput Computing, Asynchronous Communication

Ettore Mosca, Ivan Merelli and Luciano Milanesi
Institute for Biomedical Technologies – CNR (Consiglio Nazionale delle Ricerche), Segrate, Milan, Italy

Definition

Asynchronous communication is a third component mediated communication in which sender and receiver are not concurrently engaged in communication. The main feature of asynchronous communication is the transmission of data without the use of an external signal for coordination, which results in a non-blocking policy of the message exchange.

References

Eiben AE, Smith JE (2003) Introduction to evolutionary computing, (Natural computing series). Springer, Berlin

Hill Equation

Pramod R. Somvanshi and Kareenhalli V. Venkatesh
Department of Chemical Engineering, Indian Institute of Technology Bombay, Powai, Mumbai, Maharashtra, India

Synonyms

Hill function; Hill kinetics

Definition

Several molecular interactions in cellular systems exhibit sigmoidal response curve to variations in the input concentrations. Such a response curve is typically represented by Hill equation as given below.

$$Y = \left(\frac{I^{n_H}}{K_{0.5}^{n_H} + I^{n_H}} \right) \quad (1)$$

where Y is the output response and I is the input concentration. Hill equation involves two parameters, Hill Coefficient (n_H) and half-saturation constant $(K_{0.5})$. While Hill coefficient characterizes the sensitivity of the response, the half-saturation constant quantifies the threshold concentration required for 50% output response. Hill equation is typically used to quantify cooperativity, where the initial binding of an effecter molecule (ligand, activator) to the receptor enhances the binding of the forthcoming effecter molecules (Goldbeter and Dupont 1990). This is observed in allosteric regulation of enzymes and of ligands binding to their respective macromolecules. The equation describes saturation of the receptor molecules as a function of the effector concentration (Klipp et al. 2005).

Characteristics

In 1910, Archibald Vivian Hill was the first to introduce this equation to describe equilibrium relationship between oxygen tension and the saturation of hemoglobin with oxygen. He observed a sigmoidal binding curve of oxygen with hemoglobin, which revealed the cooperative kinetics of O_2 binding to hemoglobin.

To mathematically represent this phenomenon, he introduced a cooperativity coefficient, which was termed as Hill coefficient (Hill 1910, 1913).

Hill Equation is a rate law, which is used to model biological interactions that demonstrate sigmoidal response. The equation is used to capture the biomolecular interaction that exhibit cooperativity among two binding molecules. The bound subunit has the cooperative effect on the binding of the next subunit by increasing its affinity toward the binding region, which in turn increases the rate of reactions as compared to the single receptor-ligand reaction (Weiss 1997; Klipp et al. 2005). Hill equation is the general formalism used to study emergent properties, such as ▶ ultrasensitivity, ▶ amplification, and ▶ bistability, in biological networks. Further, Hill equation is also used extensively in many PK-PD models to describe the nonlinear drug-dose response relationship (Goutelle et al. 2008).

Receptor-Ligand Binding Kinetics

Hill equation represents the physicochemical equilibrium between the reacting species and can be derived based on the ▶ law of mass action. We consider a reaction of n number of ligands binding to a receptor which is given by

$$R + nL \leftrightarrow RL_n \tag{2}$$

Appling the law of mass action, the equilibrium dissociation constant K_d is given by

$$K_d = \frac{[R][L]^n}{[RL_n]} \tag{3}$$

where *[L], [R]* and *[RL_n]* are the molecular concentrations of the ligand, receptor, and ligand-receptor complex. At equilibrium condition, the total receptor concentration is given as,

$$R_T = [R] + [RL_n] \tag{4}$$

where, *[R_T]* is the total receptor concentration. The fractional saturation of the receptor (Y) is given by the ratio of the number of bound receptors to the total number of receptors,

$$Y = \frac{[RL_n]}{[RL_n] + [R]} \tag{5}$$

Substituting [RL_n] from Eqs. 3 in 5, we obtain

$$Y = \frac{[L]^n}{[L]^n + K_d} \tag{6}$$

This is the Hill equation for the receptor-ligand binding kinetics. In biomolecular reactions, for any input function (I), the above equation can be written in terms of (I) by replacing (L), which gives the generalized form of Hill equation as given in Eq. 1.

The rates of the reactions with allosteric regulation (that exhibit cooperativity) can be represented using Hill equation, as given below,

$$v = V_{max}Y = \frac{V_{max}I^{n_H}}{K_{0.5}^{n_H} + I^{n_H}} \tag{7}$$

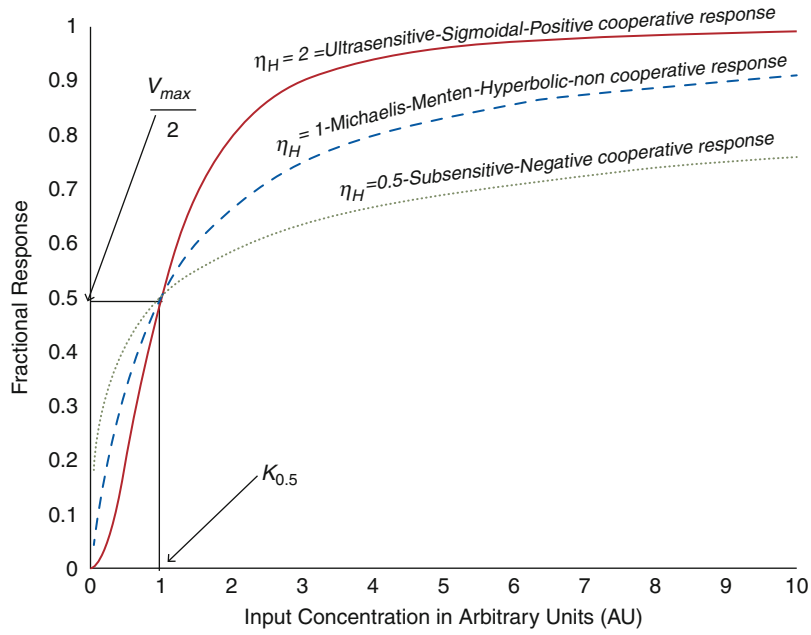
where v and V_{max} represent the net rate and maximum rate of reaction, respectively. The plot of the fractional saturation Y versus I yield a typical input-output response curve. The nature of the curve varies based upon the Hill coefficient (Fig. 1). The Hill coefficient of one gives a typical Michaelis-Menten hyperbolic response while Hill coefficient greater than one yields a sigmoidal curve with the inflection point at $K_{0.5}$. The analogy of the receptor-ligand interaction kinetics yielding Hill equation, can be applied to several biomolecular reactions such as binding of transcription factor to the promoter in a gene regulatory network, enzymatic reactions in metabolic network, and phosphorylation cycles in signaling pathways. This equation can also represent the kinetics of repression of a biomolecular reaction which is given by

$$Y = \left(\frac{K_{0.5}^{n_H}}{K_{0.5}^{n_H} + R^{n_H}}\right) \tag{8}$$

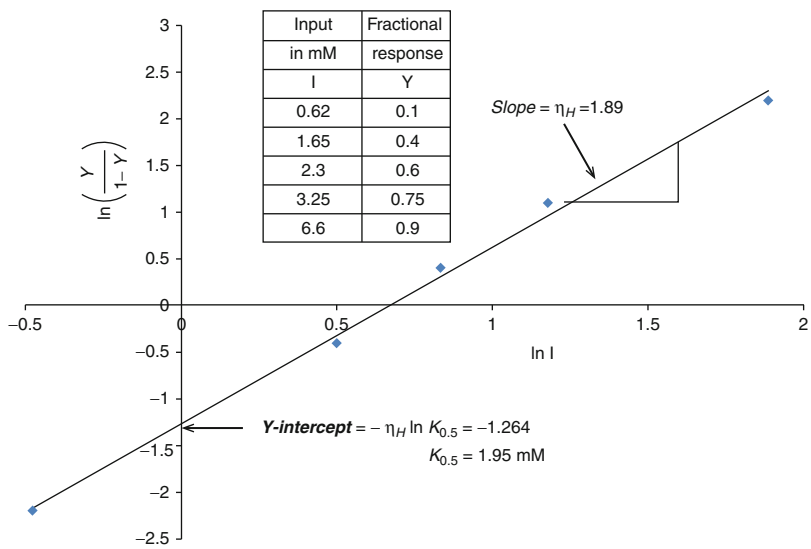
where Y is the fractional rate of expression, R is the concentration of repressor molecule, $K_{0.5}$ is half-saturation constant, and n_H is the Hill coefficient (Alon 2007). In this case, it can be noted that the output response Y tends to zero as the repressor concentration increases to infinity (see Eq. 8).

Parameter Estimation for Hill Equation

Hill coefficient (n_H) and the half-saturation constant $(K_{0.5})$ are the two parameters used in the Hill equation. These parameters can be obtained by linearizing the



Hill Equation,
Fig. 1 Typical input-output response obtained using Hill equation for different values of Hill coefficient



Input in mM	Fractional response
I	Y
0.62	0.1
1.65	0.4
2.3	0.6
3.25	0.75
6.6	0.9

Hill Equation,
Fig. 2 Graphical evaluation of parameters of Hill equation

Hill equation (Eq. 1). The linearized form of Hill equation is given by:

$$\ln\left(\frac{Y}{1-Y}\right) = n_H \ln I - n_H \ln K_{0.5} \qquad (9)$$

By plotting the LHS of the equation against *ln I* one can obtain the Hill coefficient (n_H) as the slope of the curve and the intercept of Y-axis can be used to estimate the half-saturation constant ($K_{0.5}$) (Covel 1970). Figure 2 shows an example to demonstrate the graphical evaluation of these parameters.

Hill Coefficeint and Cooperativity

Hill coefficient provides the measure of cooperativity that can be quantified based on the steepness of the binding curve saturation (Goldbeter and Dupont 1990). The measure of the steepness of the curve is captured

by Hill coefficient, which depicts the variation in the response curves from hyperbolic to sigmoidal. The Hill coefficient is computed based on the fold change in input stimuli required to take a response from 10% activation to 90% activation.

$$n_H = \frac{\log(81)}{\log\left(\frac{I_{90}}{I_{10}}\right)} \quad (10)$$

A typical Michaelis-Menten hyperbolic response has $n_H = 1$, which requires 81-fold change in the input stimulus to bring 90% of maximum response. A fractional Hill coefficient, that is $n_H < 1$, indicates negative cooperativity where binding of one ligand decreases the affinity for the binding of the other. The response generated through negative cooperativity is termed as subsensitivity where more than 81-fold change in the input stimulus is required to obtain 90% of maximum activation (see Fig. 1). The response indicates positive cooperativity when $n_H > 1$ leading to a sigmoidal response, which is also termed as ▶ ultrasensitivity. In such a case, less than 81-fold change in input stimulus is required to obtain 90% of the maximum activation (see Fig. 1) (Vinod and Venkatesh 2008; Koshland 1987).

Cross-References

- ▶ Amplification
- ▶ Bistability
- ▶ Law of Mass Action
- ▶ Pathway Modeling, Metabolic
- ▶ Ultrasensitivity

References

Alon U (2007) An introduction to systems biology-design principles of biological networks. Taylor & Francis Group/Chapman & Hall/CRC, Boca Raton

Covel ML (1970) Analysis of Hill interaction coefficients of the kwon and brown equation. J Biol Chem 245(23):6335–6336

Goldbeter A, Dupont G (1990) Allosteric regulation, cooperativity, and biochemical oscillations. Biophys Chem 37(1–3):341–353

Goutelle S, Maurin M, Rougier F, Barbaut X, Bourguignon L, Ducher M, Maire P (2008) The Hill equation: a review of its capabilities in pharmacological modeling. Fundam Clin Pharmacol 22:633–648

Hill AV (1910) The possible effects of the aggregation of the molecules of hemoglobin on its dissociation curves. J Physiol 40:4–7

Hill AV (1913) The combinations of haemoglobin with oxygen and carbon monoxide. Biochem J 7:471–480

Klipp E, Herwig R, Kowald A, Wierling C, Lehrach H (2005) Systems biology in practice concepts, implementation and application. WILEY-VCH, Weinheim

Koshland DE (1987) Switches, thresholds and ultrasensitivity. Trends Biochem Sci 12(6):225–229

Vinod PK, Venkatesh KV (2008) Quantification of signaling networks. J Indian Inst Sci 88:1

Weiss JN (1997) The Hill equation revisited: uses and misuses. FASEB J 11:835–841

Hill Function

▶ Hill Equation

Hill Kinetics

▶ Hill Equation

Histone Chaperones

Toshiya Senda[1] and Naruhiko Adachi[2]
[1]Biomedicinal Information Research Center (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

Definition

Histone chaperones are factors that interact with histones and mediate nucleosome assembly, disassembly, or both in an ATP-independent manner. They play critical roles in various nuclear events. The first identified histone chaperone, nucleoplasmin, was isolated in 1978 by Laskey (Wolffe 1998). Since then, many histone chaperones have been isolated (Eitoku et al. 2008). Histone chaperones can be categorized by the preference of histone binding, namely, those with a preference for histones H3–H4 and those with a preference for histones H2A–H2B. Histone chaperones are involved in histone storage, histone transfer, histone exchange (between old and newly synthesized histones, and between canonical and variant histones), and nucleosome structural change, which are required for most DNA-mediated

reactions in eukaryotes. Several lines of evidence have shown that histone chaperones play critical roles in transcription, replication, and DNA repair (Eitoku et al. 2008).

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

Wolffe AP (1998) Chromatin, structure function, 3rd edn. Academic, San Diego

Histone Covalent Modification

▶ Post-translational Modifications, Histone

Histone Modification

▶ Post-translational Modifications, Histone

Histone Post-translational Modification to Nucleosome Structural Change

Naruhiko Adachi[1] and Toshiya Senda[2]
[1]Structure-guided Drug Development Project, JBIC Research Institute, Japan Biological Informatics Consortium, Tokyo, Japan
[2]Biomedicinal Information Research Centre (BIRC), National Institute of Advanced Industrial Science and Technology (AIST), Tokyo, Japan

Characteristics

Although histone post-translational modification (PTM) and factors involved in nucleosome structural change have been studied as separate research fields, the findings of these two fields are being integrated into a comprehensive picture as the molecular mechanisms linking histone PTMs to nucleosome structural change continue to be discovered.

In 1964, Murray reported methylation of histones in the cell. This was probably the first report of histone modification. Soon after that, the Allfrey group revealed that histones are subjected to acetylation and methylation after the completion of the histone polypeptide chain, and that there is a correlation between histone acetylation and transcriptional activation (Allfrey et al. 1964). Although this discovery suggested the connection between histone post-translational modifications (PTMs) and nuclear events, the molecular mechanism underlying the observed correlation was at that time unknown due to lack of a sufficient knowledge of the chromatin template and biological signaling in the nucleus. It took more than 40 years of effort for researchers to begin to understand the molecular mechanisms connecting histone PTMs and transcription activation; for these breakthroughs to take place, substantial advances were required in three research fields, the structure of the chromatin template, factors involved in nucleosome structural change, and nuclear signaling with histone PTMs.

In 1974, Kornberg discovered the nucleosome structure (Kornberg and Lorch 1999). The nucleosome core particle was considered to be a complex of the histone octamer, which comprises two copies of each canonical histone (H2A, H2B, H3, and H4), and DNA. As was easily recognized from the tertiary structure of the nucleosome, the nucleosome structure inhibits enzyme reactions with DNA such as transcription, replication, and DNA repair due to tight interactions between DNA and histone proteins. Therefore, the nucleosome structure must be disassembled to initiate these reactions. On the other hand, the nucleosome structure is necessary to stably maintain genetic information in the nucleus. To resolve the functional conflicts of the nucleosome, regulation of nucleosome assembly and disassembly is required.

Two types of factors involved in the nucleosome structural change, histone chaperones and ATP-dependent nucleosome-remodeling factors, have so far been identified. Nucleoplasmin, which is categorized as a histone chaperone, was the first factor found to have a nucleosome structural change activity (Wolffe 1998; Eitoku et al. 2008). Although many

histone chaperones have been found since then, their biological significance had remained elusive until the functional identification of the histone chaperone CIA/Asf1 (Eitoku et al. 2008). Genetic, biological, and biochemical studies showed that CIA/Asf1 is involved in various nuclear events such as transcription, replication, and DNA repair through nucleosome structural change.

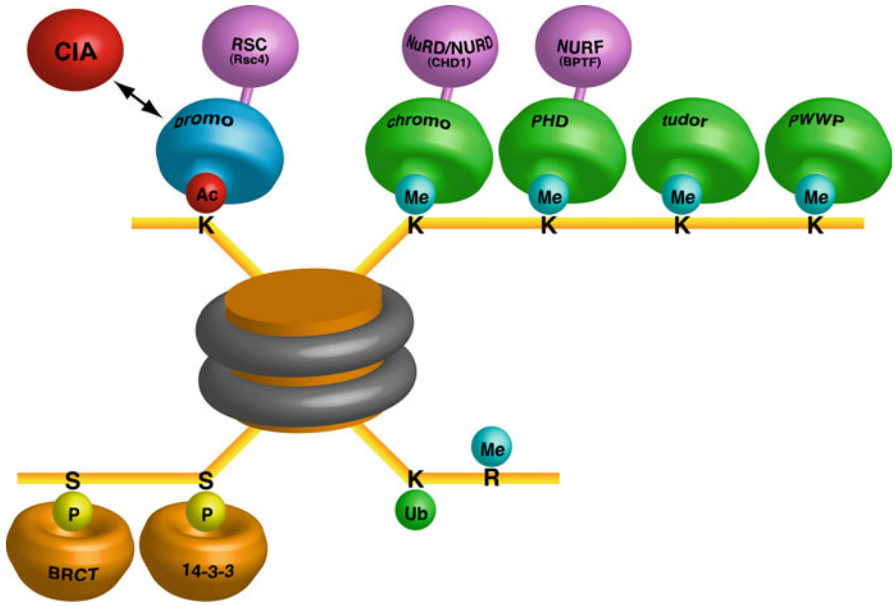
The first-identified ATP-dependent nucleosome-remodeling factor, the SWI/SNF complex, was isolated as a complex that contained the gene product of *swi2*/*snf2*. The *swi2*/*snf2* gene was initially identified from mutant yeast strains that showed the SWI and SNF phenotypes. Since these phenotypes were not observed in mutant yeast strains harboring destabilized chromatin structures, the *swi2*/*snf2* gene was considered to be functionally relevant to the chromatin structure. A purified protein complex including the *swi2*/*snf2* gene product showed ATP-dependent nucleosome-remodeling activity (Turner 2001; Elgin and Workman 2001). After this finding, several ATP-dependent nucleosome-remodeling factors such as NURF, RSC, ACF, CHRAC, NuRD/NURD, and INO80 were isolated. Biochemical and biological studies have shown that ATP-dependent nucleosome-remodeling factors adjust the position of nucleosomes and are involved in nuclear events such as transcription and DNA replication (Eberharter and Becker 2004; Elgin and Workman 2001).

The last critical advance is attributed to functional and mechanistic studies of histone PTMs. The discovery of a histone acetyltransferase and the subsequent identification of several histone modification enzymes accelerated studies of histone PTMs and their functions (Allis et al. 2006). Since some transcriptional coactivators contain a histone acetyltransferase domain/subunit, the connection between histone PTMs and nucleosome structural change that is required for transcription activation came to be recognized. It was proposed that histone PTMs themselves affect the structure of chromatin through changing the electrostatic and/or physical properties of nucleosomes. However, this hypothesis seemed to be insufficient to explain all nucleosome structural changes required for transcriptional activation. The situation was advanced by the structural and functional analysis of bromodomains. These studies revealed that bromodomains recognize acetylated lysines on histone proteins. Following this finding, several histone PTM recognition domains have been identified, as summarized in Fig. 1 (Allis et al. 2006). These findings established the notion that histone PTMs are recognized by their corresponding recognition domains, which are likely to function as adaptor domains linking histone PTMs and chromatin factors, such as ATP-dependent nucleosome-remodeling factors and histone chaperones.

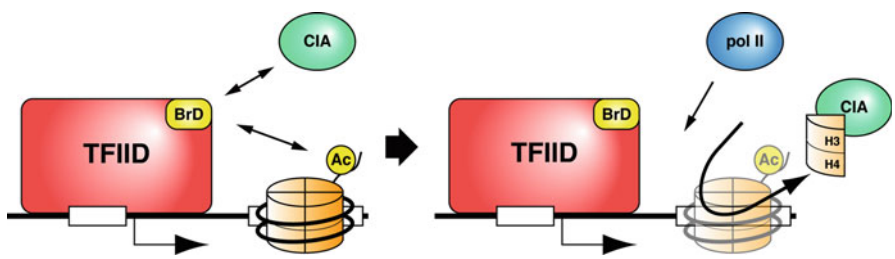
Indeed, ATP-dependent nucleosome-remodeling factors contain histone PTM recognition domains. For example, the Rsc4 subunit of RSC contains tandem bromodomains that physically interact with acetylated Lys14 of histone H3 (H3-K14) on the histone tail region, suggesting a relationship between histone acetylation and the activity of RSC. Genetic and biological analyses have suggested that RSC is involved in site-specific nucleosome remodeling in transcription in response to histone PTMs including acetylated H3-K14. The connection between acetylation of histone proteins and nucleosome structural change seems to be mediated by the tandem bromodomains in the Rsc4 subunit.

In contrast with ATP-dependent nucleosome-remodeling factors, no histone chaperones have histone PTM recognition domains in the molecule. The Horikoshi group, however, discovered physical and genetic interactions between histone chaperone CIA/Asf1 and double bromodomains in the general transcription initiation factor TFIID (Chimura et al. 2002), leading to the idea that histone acetylation could regulate the function of the histone chaperone CIA/Asf1. A molecular model that connects histone PTMs and transcription activation through nucleosome structural change by histone chaperone CIA/Asf1 was proposed on the basis of two crystal structures, CIA/Asf1–H3–H4 and CIA/Asf1–double-bromodomain complexes. Structure-based biochemical, genetic and molecular biological analyses have suggested that acetylated histones recruit histone chaperone CIA/Asf1 via the double bromodomain in the TFIID complex to a promoter region, resulting in histone eviction around the promoter site (Fig. 2) (Akai et al. 2010).

The correlation between histone PTMs and transcription activation (through nucleosome structural change) was discovered about 45 years ago. In the past 45 years, structural details of chromatin, the signaling system with histone PTMs, and the structural and functional relationship of histone chaperones and ATP-dependent chromatin-remodeling factors have been revealed. These findings have begun to be integrated such that the



Histone Post-translational Modification to Nucleosome Structural Change, Fig. 1 Histone PTM recognition domains. Histone acetylation (*red*), methylation (*cyan*), and phosphorylation (*yellow*) recognition domains are shown in *blue*, *green*, and *orange*, respectively. The names of the domains are labeled. Histone chaperones interacting with these histone PTM recognition domains are shown in *red*. Usually ATP-dependent nucleosome-remodeling factors contain a histone PTM recognition domain in the complex. (The relationships are shown by *thick purple lines*.) ATP-dependent nucleosome-remodeling factors are shown in *purple* with labels. (Names of the subunit containing a histone PTM recognition domain are given in parentheses.) No recognition domains have been reported for histone ubiquitination (*green*) and arginine methylation (*cyan*)



Histone Post-translational Modification to Nucleosome Structural Change, Fig. 2 A schematic representation of the hi-MOST model (linking the biological signaling from *hi*stone *mo*difications to *st*ructural change of the nucleosome). Biochemical, genetic, and structural analyses indicate that the double bromodomain of TFIID recruits CIA to specific promoter regions. Recruited CIA evicts histones and promotes RNA polymerase II entry. *BrD* bromodomain, *Ac* acetylation

connection between histone PTMs and the nucleosome structural change can be reasonably explained at the molecular level. However, our understanding of these molecular processes is still limited. For example, the details of the molecular mechanism for nucleosome disassembly by histone chaperones – such as H2A–H2B and DNA dissociation from the nucleosome – remain elusive. In addition, much is unknown about the signaling mechanism with histone PTMs. Although the histone code hypothesis was proposed to explain the signaling mechanism with histone PTMs (Allis et al. 2006), the hypothesis is too simple to explain the results of several

mutational analyses of the histone tail region. In order to understand the whole picture of histone PTM signaling, a macroscopic theory that considers the whole network of histone PTM signaling is required. A combination of experimental and theoretical approaches should lead to a comprehensive understanding of molecular processes between histone modification and various nuclear events including nucleosome structural change.

Cross-References

- ▶ Histones
- ▶ Nucleosome Structure
- ▶ Nucleosomes
- ▶ Post-translational Modifications

References

Akai Y, Adachi N, Hayashi Y, Eitoku M, Sano N, Natsume R, Kudo N, Tanokura M, Senda T, Horikoshi M (2010) Structure of the histone chaperone CIA/ASF1-double bromodomain complex linking histone modifications and site-specific histone eviction. Proc Natl Acad Sci USA 107:8153–8158

Allfrey VG, Faulkner R, Mirsky AE (1964) Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. Proc Natl Acad Sci USA 51:786–794

Allis CD, Jenuwein T, Reinberg D, Caparros M-L (2006) Epigenetics, 1st edn. Cold Spring Harbor Laboratory Press, New York

Chimura T, Kuzuhara T, Horikoshi M (2002) Identification and characterization of CIA/ASF1 as an interactor of bromodomains associated with TFIID. Proc Natl Acad Sci USA 99:9334–9339

Eberharter A, Becker PB (2004) ATP-dependent nucleosome remodeling: factors and functions. J Cell Sci 117: 3707–3711

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

Elgin SCR, Workman JL (2001) Chromatin structure and gene expression (Frontiers in molecular biology), 2nd edn. Oxford University Press, Oxford

Kornberg RD, Lorch Y (1999) Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. Cell 98:285–294

Turner BM (2001) Chromatin and gene regulation: molecular mechanisms in epigenetics, 1st edn. Wiley-Blackwell, Hoboken

Wolffe AP (1998) Chromatin, structure function, 3rd edn. Academic, San Diego

Histones

Vani Brahmachari and Shruti Jain
Dr. B. R. Ambedkar Center for Biomedical Research, University of Delhi, Delhi, India

Synonyms

Chromatin proteins; Nucleosome

Definition

Histones are highly basic proteins with molecular weights ranging from 11 to 23 kDa and form the basic unit of ▶ nucleosomes. There are four core histones named H2A, H2B, H3, and H4; two subunits of each interact to form the histone octamer. Histone H1 is called the linker histone as it binds to the DNA between two nucleosomes. The histone-DNA complex brings about the first level of compaction of the DNA in the nucleus. The H1, H2A, and H2B are rich in lysine amino acid, whereas H3 and H4 are rich in arginine. The amino-terminal tails of these proteins extend beyond the nucleosome and, hence, are accessible for covalent modification, while the octamer is bound to DNA and participates in the interaction with various other gene regulatory proteins.

Cross-References

- ▶ Epigenetics
- ▶ Nucleosomes

Histopathology

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

Definition

Histopathology is the study of microscopic structures of diseased tissues (Crissman et al. 2004).

References

Crissman JW, Goodman DG, Hildebrandt PK, Maronpot RR, Prater DA, Riley JH, Seaman WJ, Thake DC (2004) Best practices guideline: toxicologic histopathology. Toxicol Pathol 32:126–131

Holism

Alvaro Moreno
Department of Logic and Philosophy of Science,
University of the Basque Country,
San Sebastian, Spain

Definition

Holism is the idea that whole entities are fundamental components of reality and have an existence, which is irreducible to the sum of their parts. Thus, all the properties of a given system cannot be determined or explained in terms of the properties of its component parts alone. Instead, the system as a whole is what determines in an important way how the parts behave. Holism in biology is the theory that emphasizes that living entities can only be understood as wholes, because they show global emergent properties that cannot be attributed to specific or well-distinguishable parts. Systems are said to be *holistic* when the linear aggregation of their parts cannot explain the functioning of the system as a whole: "the whole is more than the sum of the parts" as the main slogan of the sciences of complexity runs.

There are different types of holism, depending on which kind of reductionism they are opposed to (Gatherer 2010). For example, some authors defend holism because they are epistemological antireductionists (e.g., that some phenomena are so complex that their behavior cannot be deduced from the knowledge of their fundamental properties), while others go beyond the epistemological argument and adopt ontological antireductionism (e.g., that there exist, in certain complex systems, true emergent properties, often considered endowed with specific, "downward" causal powers) (Andersen et al. 1996).

History

Holism in biological sciences has its roots in German eighteenth-century biologists and philosophers, who emphasized the impossibility to study living beings according to the mechanistic-analytic tradition (nowadays sometimes called reductionism), which purports to understand systems by dividing them into their smallest possible or discernible elements and describing their elemental properties alone. In the anti-mechanistic view, known as "organicism" (the term "organicism" refers likely to the properties of the whole organism, whereas "holism" is a more encompassing concept, thus applicable to any level of organization in biology), living processes are studied in relation to the integrated organized whole, (namely, the entire organism), rather than to any of its parts. Both Goethe and Kant (1790), for example, considered that, contrary to what happens in man-made machines, in a natural organism, every part could neither be explained nor even exist, isolated from the whole. Each part must be an organ producing the other parts and being produced by them.

The debate between the mechanistic-reductionistic and the organicistic-antireductionistic views traversed the history of nineteenth- and twentieth-century biology, taking many different forms, as, for instance, the dispute between mechanists and vitalists in the early decades of the past century. Yet, the modern history of holism in biology begins with the work of researchers like Rachevsky (1938), Bertalanffy (1952), and Elsasser (1966), who argued for the necessity to adopt an integrated approach to deal with living systems (and other complex systems as well). British emergentists of the early twentieth century, like Broad (1925), Morgan (1923), and Alexander (1920), are also important references in the modern history of holism in biology and philosophy of biology. (Actually, it was in this context that Smuts (1926) proposed the first modern version of the concept of holism). The common idea of these authors is that as the organization of systems becomes more and more complex, it gets structured in levels, and new properties and causal interactions appear, in addition to those of the more fundamental levels. Their emergent properties are systemic features of biological systems which could not be predicted, despite a thorough knowledge of the features of their parts and of the laws governing them.

The experimental success of the research program of molecular biology during the second half of the twentieth century undermined anti-reductionist arguments and caused holistic objections to fade away. At the end of the last century, however, the reductionist research program faced a dead end (Morange 2003). As knowledge on biological systems became more detailed and

fine-grained, researchers progressively realized that components were acting in strongly holistic ways. For example, at the end of the last century, research in the structure of the genome faced increasing evidences that genetic components acted in a complex web of interactions. Thus, as E. F. Keller has pointed out (2007), research in biology changed from a program inscribed in DNA analysis to a new distributed (namely, more holistic) program in which DNA, RNA, and protein components operate alternatively as instructions and data.

To be sure, this turn has not only been a consequence of the internal development of biology but also of the recent development of new scientific tools alternative to the traditional analytic-reductionistic methods. The combination of increasingly powerful computers along with new modeling techniques (such as cellular automata, genetic algorithms, Boolean networks, chaos, and dynamical systems theory) has allowed a blossom of holism in modern science. All these deep innovations are affecting many scientific disciplines, giving rise to what is currently called the new "sciences of complexity." This impact is especially important in biology, where new approaches like artificial life, synthetic biology, and systems biology constitute its main expression. Interestingly, it is precisely in the field of systems biology that the idea of holism and the debate between reductionist and emergentist views has resurfaced with renewed force (Cornish-Bowden et al. 2005; Conti et al. 2007; Boogerd et al. 2007). (It must be admitted, however, that the blossoming of holism in systems biology is rather a reject of reductionism than a return to the holism of the 1930s, as emphasized by Gatherer (2010)).

Characteristics

Holistic phenomena occur even in relatively "simple" physical or chemical systems: an ensemble of interacting units producing a global property, or pattern of behavior, that cannot be ascribed to any of them but only to the whole. However, as Keller (2007) has pointed out, in these systems, the emergent, holistic pattern lacks any form of functional differentiation. This is not the case in biological systems, in which the presence of holistic properties goes together with functional diversification. Compared to nonliving holistic systems (like ordinary dissipative structures), biological systems involve a much richer internal structure: They are made of parts with different functionalities, acting in a selective and harmonized way, coordinating themselves at different timescales, interacting hierarchically in local networks, which form, in turn, global networks, meta-networks, etc.

This complex organization shows that, in fact, holism and mechanistic de-composition can be combined for the purposes of biological explanation, as Bechtel has recently pointed out (Bechtel 2010). In biological systems, holistic-emergent processes (which are continuously taking place) produce both dissipative patterns and more complex structures, which, in turn, are bound to become selective functional constraints acting on the dynamic processes that underlie those holistic processes. Moreover, these functionally diverse constraints may give rise (once a certain degree of variety is reached) to new self-organizing holistic processes, which, in turn, may be functionally reorganized. In this way, an increase in organizational complexity can take the paradoxical form of an apparent "simplification" of the underlying complicatedness, giving rise to levels of organization in which a mechanistic de-compositional strategy might be locally applicable. Interestingly, these functional constraints can be described as localized mechanisms (and therefore, to a certain degree, they are amenable to functional de-composition) because they act as distinguishable parts (or collections of parts) related to particular tasks performed in the system (for example, catalytic regulation). These two types of processes – the holism of the global network of processes and the local control devices/actions – are anyhow complementary: Both are required to produce and maintain the high level of complexity that characterizes biological systems. Thus, de-composition and re-composition strategies turn out to be complementary.

In sum, holism is a pervasive phenomenon in the world of complex systems. Yet, the high degree of complexity of biological systems relies on a specific form of organization that combines holism and locally differentiated functions and mechanisms. How to understand this complex entanglement between emergent holistic processes and functionally localized structures is probably the most difficult challenge of future research in systems biology.

Cross-References

- ▶ Complex System
- ▶ Complexity
- ▶ Emergence

- ▶ Interlevel Causation
- ▶ Reduction
- ▶ Self-Organization

References

Alexander S (1920) Space, time, and deity. Macmillan, London
Andersen PB, Emmeche C, Finnemann NE, Christiansen PV (1996) Downward causation: minds, Bodies and Matter. Aarhus University Press, Aarhus
Bechtel W (2010) The cell, locus or object of inquiry? Stud Hist Philos of Biol Biomed Sci. doi:10.1016/j.shpsc.2010.07.006
Boogerd F, Bruggeman F, Hofmeyr J-H, Westerhoff HV (2007) Systems biology. Philosophical foundations. Elsevier, Amsterdam
Broad CD (1925) The mind and its place in nature. Routledge & Kegan Paul, London
Conti F, Valerio MC, Zbilut JP, Giuliani A (2007) Will systems biology offer new holistic paradigms to life sciences? Syst synth biol 1:161–165
Cornish-Bowden A, Cardenas ML (2005) Systems biology may work when we learn to understand the parts in terms of the whole. Biochem Soc Trans 33:516–519
Elsasser W (1966) Atom and organism. Princeton University Press, Princeton
Gatherer D (2010) So what do we really mean when we say that systems biology is holistic? BMC Syst Biol. doi:10.1186/1752-0509-4-22
Kant I (1790/1952) Critique of judgment. Oxford University Press, Oxford
Keller EF (2007) The disappearance of function from 'self-organizing systems'. In: Boogerd F, Bruggeman F, Hofmeyr J-H, Westerhoff HV (eds) Systems biology. Philosophical foundations. Elsevier, Ámsterdam, pp 303–317
Morgan CL (1923) Emergent evolution. Williams and Norgate, London
Rachevsky N (1938) Mathematical biophysics: physico-mathematical foundations of biology. Dover Publications, New York
Smuts JC (1926) Holism and evolution. Macmillan, London
von Bertalanffy L (1952) Problems of life: an evaluation of modern biological thought. Watts, London

Holm's Method

Winston Haynes
Seattle Children's Research Institute, Seatlle, WA, USA

Synonyms

Holm's procedure; Holm-Bonferroni method; Sequentially rejective Bonferroni test

Definition

Designed for multiple hypothesis testing, the Holm's method iteratively accepts and rejects hypotheses. The Holm's method is a close relative to the ▶ Bonferroni correction with slightly different threshold levels.

Let α be the determined significance threshold for rejecting the null hypotheses and k be the number of hypotheses. Begin by ordering the k hypotheses by their respective p-values. Select the lowest p-value and compare it to α/k. If the p-value is lower, then reject that hypothesis and perform the same selection with the remaining $k-1$ hypotheses and a threshold of $\alpha/(k-1)$. Repeat the process until the selected p-value is not smaller, at which time all remaining hypotheses should be accepted.

By progressively adapting the threshold values, the Holm's method gains power over the ▶ Bonferroni correction. Whereas in the Bonferroni correction all values are thresholded relative to a/k, the Holm's method utilizes α/k, $\alpha/k-1,\ldots,\alpha$. Therefore, the probability of rejecting a hypothesis with the Bonferroni method is less than or equal to the same probability for the Holm's method.

References

Holm S (1979) A simple sequentially rejective multiple test procedure. Scand J Stat 6(2):65–70

Holm's Procedure

- ▶ Holm's Method

Holm-Bonferroni Method

- ▶ Holm's Method

Holobiont

- ▶ Microbiome

Holoenzyme Recruitment Pathway

▶ PIC Assembly Pathways

Homeostasis

▶ Life Span, Turnover, Residence Time
▶ Lymphocyte Population Kinetics

Homeostatic Proliferation

▶ Modeling, Cell Division and Proliferation

Homogeneous Structures

▶ Cellular Automata

Homologous Recombination

Myong-Hee Sung
Laboratory of Receptor Biology and Gene Expression, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA

Definition

Homologous recombination is a process in which two similar or identical DNA segments are swapped, giving rise to a newly combined sequence of DNA.

Homoplasy

▶ Convergent Evolution

Hopf Bifurcation

Xiaojuan Sun
Zhou Pei-Yuan Center for Applied Mathematics, Tsinghua University of Beijing, Beijing, China

Synonyms

Poincaré-Andronov-Hopf bifurcation

Definition

Hopf bifurcation is a local bifurcation in which a steady state of a dynamical system changes its stability, so that the appearance or disappearance of a periodic orbit occurs.

In a dynamical system that is described by ODE model (▶ Ordinary Differential Equation (ODE))

$$\frac{dX}{dt} = F(X). \tag{1}$$

Let X^* be a steady state such that $F(X^*) = 0$ and J the coefficient matrix (Jacobian matrix) of the system when it is linearized near X^*:

$$J = \left.\frac{\partial F(X)}{\partial X}\right|_{X=X^*}. \tag{2}$$

Suppose that all eigenvalues of J have negative real parts except one conjugate nonzero purely imaginary pair $\pm\beta$. A Hopf bifurcation arises when these two imaginary eigenvalues cross the imaginary axis because of a variation of the system parameters (Hale and Kocak 1991; Strogatz Steven 1994; Kuznetsov 2004). In the critical situation when all eigenvalues of J have negative real parts except one conjugate nonzero purely imaginary pair, the system is said to have critical parameter value.

References

Hale J, Kocak H (1991) Dynamics and bifurcations. Springer, New York
Kuznetsov Yuri A (2004) Elements of applied bifurcation theory. Springer, New York
Strogatz Steven H (1994) Nonlinear dynamics and chaos. Addison Wesley, Reading, Mass

Horizontal Genomics

Maureen A. O'Malley
Department of Philosophy, University of Sydney, Sydney, NSW, Australia

Synonyms

Lateral genomics

Definition

Horizontal genomics is the large-scale study of the pattern and process of gene transfer between organisms, whether closely or distantly related. These investigations have considerable implications for whether the evolutionary history of organisms, particularly microorganisms, should be represented by a unique tree of life.

Cross-References

► Metagenomics

References

Bapteste E, Burian RM (2010) On the need for integrative phylogenomics, and some steps towards its creation. Biol Philos 25:711–735
Doolittle WF (1999) Lateral genomics. Trends Cell Biol 9: M5–M8
Frost S, Leplae R, Summers AO, Toussaint A (2005) Mobile genetic elements: the agents of open source evolution. Nat Rev Microbiol 3:722–732

Host Adaptive Immune Response to HIV Infection

► Systems Immunology, Adaptive Immune Response to HIV Infection

Host Factor 1

► Hfq

Host–Pathogen Interactions

Karyala Prashanthi[1] and Nagasuma Chandra[2]
[1]Bioinformatics Centre, Indian Institute of Science, Bangalore, Karnataka, India
[2]Department of Biochemistry, Indian Institute of Science, Bangalore, India

Definition

The manifestation of a disease is critically dependent upon the interactions of pathogens with their hosts (HPIs), which are complex and dynamic in nature. HPIs are multifaceted, with each species having to recognize, respond, and adapt to each other. Specific interactions occur between macromolecules of the host and pathogen, some beneficial to the host as in pathogen elimination or suppression, while others are beneficial to the pathogen, such as during initiation or progress of an infection or even immune evasion. An understanding of the mechanisms of such interactions is necessary for targeted development of prevention and control measures against infectious diseases.

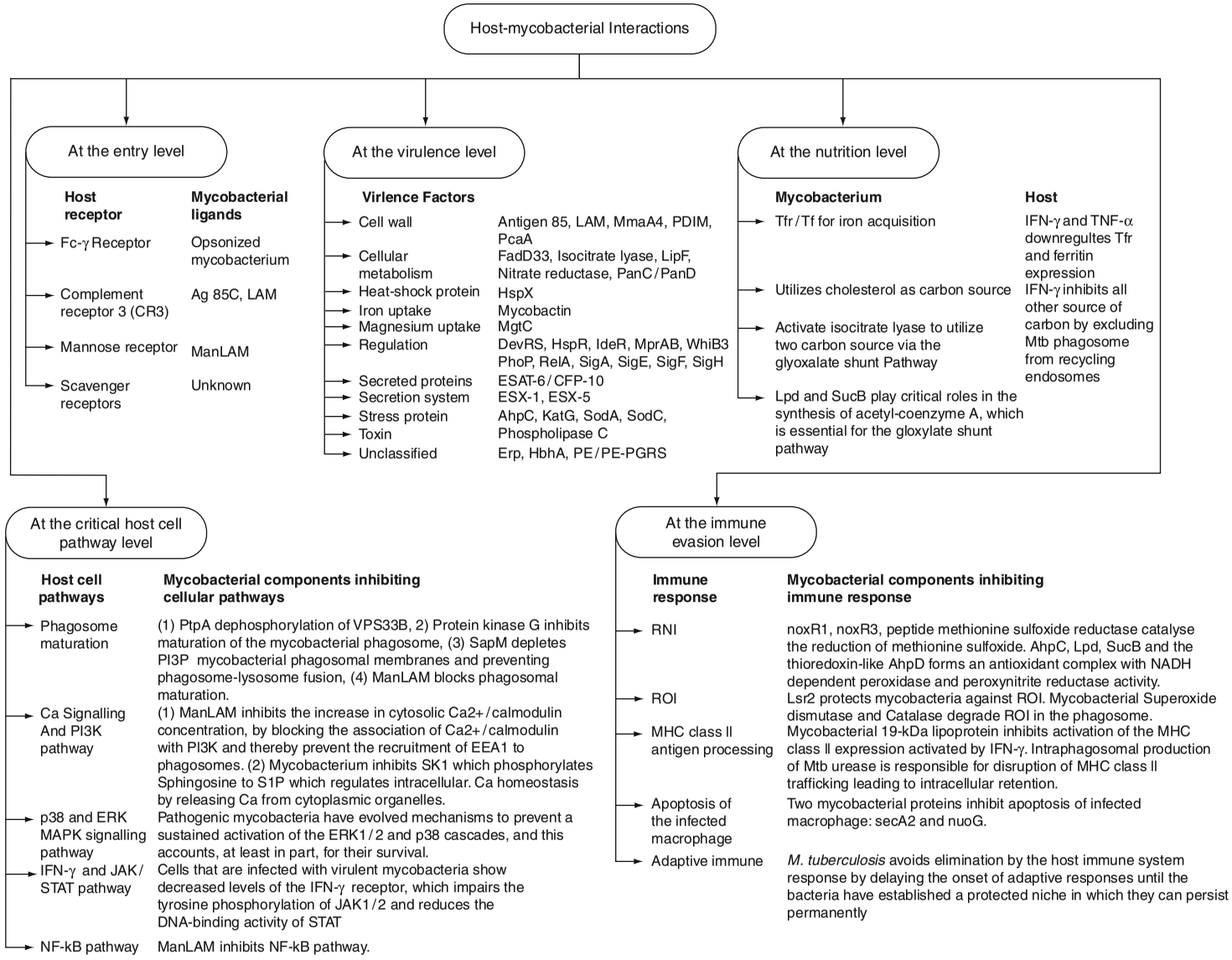
Upon entry into the host, pathogens interact with their hosts for replication or obtaining nutrition. In response, the host system attempts to either eliminate the pathogen by triggering its innate and adaptive immune responses or at least overcoming the exploitation of its resources by the pathogen, by modulating availability of nutrients and suppressing pathogenic virulence factors (Fig. 1).

The outcome that may range from pathogen clearance to asymptomatic carriage or active disease is a reflection of the properties of the microbe and the host's ability to respond to it.

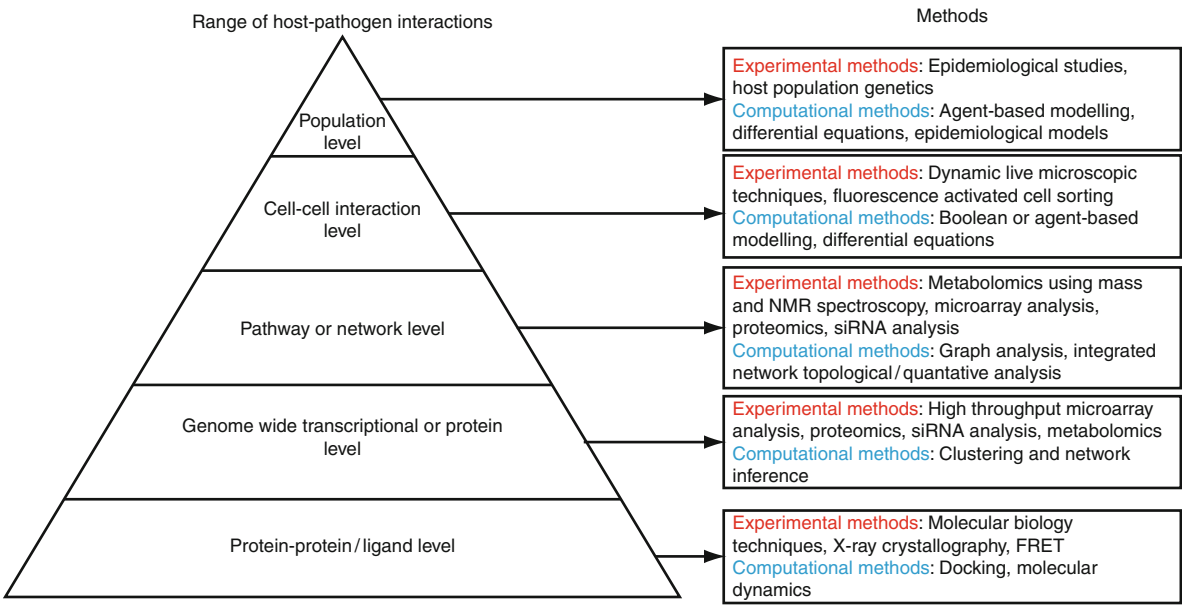
Characteristics

Systems Perspective of HPIs

Pathogens have a formidable task of surviving and infecting their target host cells, which they do, by manipulating the host's complex cellular networks. Host cells are equipped with their own armory against pathogens. Thus, it is no surprise that for every move a pathogen makes to exploit the host, a counter move has been observed in the host. Needless to say, the reverse is also true, which means, that every move by



Host–Pathogen Interactions, Fig. 1 An example of the multifaceted nature of host–pathogen interaction



Host–Pathogen Interactions, Fig. 2 Experimental and computational methods to study host–pathogen interactions at different levels of biological hierarchy

the host system in response to an infection can be countered by the pathogen as well. Thus, it becomes a complex interplay between the two species, making it often difficult to predict the winner. Appreciation of the enormity of HPIs and how they influence the outcome of an infection requires the study of the individual components as one connected system (Forst 2006). Recent developments in "omics"-scale experimental and computational technologies have facilitated the study of systems biology of HPIs. Figure 2 illustrates different levels at which HPIs are studied.

HPI in Virulence Mechanisms

Several pathogens utilize virulence factors in them to inhibit various host functions, to achieve colonization, immuno-evasion, or immune-suppression, through one or more of the following: (a) adherence to host cell involving adhesins (e.g., ManLAM in Mycobacteria) (b) invasion through invasins (e.g., collagenase produced by *Clostridium histolyticum* and neuraminidase by *Vibrio cholerae* and *Shigella dysenteriae*) that damage host cells to facilitate growth and spread of the pathogen, (c) avoidance of phagolysosome formation, (d) autophagy, (e) preventing complement activation, (f) nutrient acquisition through specialized receptor systems (e.g., Tfr/Tf in Mycobacteria), (g) quorum sensing and communication with other bacteria, and (h) biofilm formation to protect themselves against antibiotics and host defense mechanisms (Brogden et al. 2007). All of these processes are complex systems in themselves involving several molecules including toxins and extensive cross talk among them, some of which can be readily obtained from KEGG and other bio-systems' databases.

Pathogenesis of cholera by *V. cholerae* occurs through a series of spatio-temporally controlled events under the control of a gene cascade termed the ToxR regulon that encodes the virulence factors. A systems biology study of the temporal regulation of gene expression in *V. cholerae* using high-resolution time series genomic profiling (Kanjilal et al. 2010) provided insights not only into the temporal dynamics of the ToxR regulon but also identified potential new members of the process.

HPI for Adherence to or Entry into Host Cells

Pathogens must first adhere to host cells so as to colonize at appropriate sites. In its simplest form, attachment to the host cell requires two factors: a ligand and a host cell surface receptor. Alteration of the host's cellular morphology through actin polymerization or use of specialized structures in bacteria such as pili, fimbriae, or capsules is an example of mechanisms used by pathogens toward this (Pizarro-Cerdá and Cossart 2006). This initial phase then leads to a complex host–pathogen

molecular cross talk that enables them to reach their appropriate intracellular niches, subversion of cellular functions, and establishment of disease.

Endocytosis and phagocytosis are extremely dynamic processes and involve more than 200 proteins. Some pathogens trigger receptors that can activate these processes, while some others secrete toxins that enter the host cell and activate these processes. Specific sets of interactions not only trigger specific signaling cascades but also determine factors such as tissue tropism, species specificity, and genetic specificity. Specific signaling events bring about particular cellular responses such as stimulation of innate and adaptive immunity, growth, proliferation, survival, and apoptosis.

An example of a systems biology study of pathogen entry is reported for Chlamydia pneumonia, which illustrates the complexity of HPIs during entry (Wang et al. 2010). Nine functional modules that were activated upon exposure to the pathogen, consisting of 135 molecular components, involved in cell adhesion, transcription, endocytosis, and receptor systems, were incorporated into a network capturing known inter-pathway cross talks. The network was observed to be significantly resilient since intervention at any single point alone was insufficient to prevent entry. Instead manipulation of a combination of three key proteins (a chemokine receptor, an integrin receptor, and a platelet-derived growth factor) was found to inhibit pathogen's entry.

HPI for Nutrition

The human host provides an appealing ecosystem for numerous microorganisms, with a variety of adaptations, to harness the existing nutrient resources. Bacterial pathogens that colonize extracellular niches often face environments with frequently changing physical conditions and nutrients. However, intracellular pathogens replicating in cytoplasm or phagosomal compartments encounter a more congenial environment for growth. Several nutrient acquisition adaptations are also known in intracellular bacteria (Schaible and Kaufmann 2005).

Iron being an essential growth factor for both host and pathogen is associated with coevolution of mutual high affinity iron uptake and retention systems. Mycobacteria, for example, block phagosomal maturation and are present in an early phagosome and accesses host cellular iron through the host transferrin receptor/transferrin (Tfr/Tf) system. The host counters this by Interferon-γ activation, which downregulates Tfr and ferritin levels.

Although most microorganisms can synthesize organic molecules that they need, they are critically dependent on the host for a few compounds. Due to the abundance of these resources in the host, microorganisms may have lost genes required for their biosynthesis. Chlamydia, for example, is entirely dependent on the host for supply of tryptophan, an essential nutrient. However, the host has devised a defense strategy to limit tryptophan through IFN-γ-mediated activation of indoleamine 2,3 dioxygenase (IDO), to catabolize L-tryptophan to n-formylkynurenine.

HPI in Critical Host Cell Pathways

Pathogens often hijack one or more host intracellular pathways in the host, for their survival (Bhavsar et al. 2007). Salmonella and Listeria, for example, utilize the host cytoskeleton to enter and move within host cells. *S. flexneri, Yersinia* spp, and *M. tuberculosis* provide examples of inhibition of a signaling cascade (NF-κB pathway), the latter having been activated in the host in response to the infections through the use of Toll-like receptors and other pattern recognition receptors.

The c-Met signaling network that is mediated by the hepatocyte growth factor (HGF) in normal physiology to achieve mitogenesis, motogenesis, and morphogenesis gets activated by the *Helicobacter pylori* virulence factor CagA as well. However, with the latter, it is highly correlated with gastric cancer. A systems level study based on comparative logical modeling identified intervention points for CagA-induced but not HGF-induced c-Met signaling, which were subsequently validated experimentally (Franke et al. 2008).

HPI for Elimination of Pathogen by Host or Immune Evasion by Pathogens

Exposure to a pathogen leads to layers of defense responses in the host. For each level of defense, pathogens have designed ways to evade them (Henderson and Oyston 2003). For example, countering innate immune responses can be seen in the following cases: *H. pylori* counteract acidic environment in the stomach by secreting urease which increases the pH surrounding the bacterium. Some bacteria such as *Streptococcus pneumoniae* have antiphagocytic substances in their surfaces to inhibit phagocytosis. Catalase and superoxide dismutase synthesized by bacteria such as *H. pylori* and *Staphylococci* scavenge reactive oxygen intermediates. Examples where adaptive immunity is evaded include cleaving and inactivation of IgA through IgA1 proteases

by *Streptococci* and *Neisseria spp*., and over-activation of T-cells by *Staphylococci* to produce toxic levels of the pro-inflammatory cytokines, through the use of super-antigens. There are also reports about the pathogen triggering complex interactions among the host subsystems that are detrimental to the host itself. For example, the mechanisms of systemic vascular dysfunction in Dengue Shock Syndrome were correlated with interplay of innate immunity, inflammation, and host lipid metabolism (Devignot et al. 2006).

Application in drug and vaccine discovery

Knowledge of the molecular mechanisms involved in HPIs can be utilized in disease diagnosis, treatment, or prevention, in a number of ways. For example, in Ebola virus, substitution of a single amino acid in the VP35 protein is sufficient to disrupt the viral inhibition of innate immune signaling, while maintaining the ability to replicate to wild-type levels in cell culture (Hartman et al. 2008), which has led to exploration of mutant varieties as vaccine candidates.

Specific host mechanisms are also being explored to target drug therapy, by (a) preventing exploitation of host proteins for pathogen's replication or (b) enhancing hosts' natural pathogen elimination mechanisms such as production of interferons, the latter is achieved through agonists of certain toll-like receptors such as 7-thia-8 oxoguanosine (TLR7 agonist) that is used for HCV treatment (Tan et al. 2007).

Although systems level studies and hence applications in drug and vaccine discovery emanating from them are not as yet commonplace, perhaps due to difficulties in comprehensive model building and experimental design, the need for studying them as whole systems has become firmly established. Some examples from literature are already showing the power of these approaches, with a very high potential to translate into clinically useful applications in the coming years.

Cross-References

► Host-Pathogen Systems, Target Discovery

References

Bhavsar AP, Guttman JA, Finlay BB (2007) Manipulation of host-cell pathways by bacterial pathogens. Nature 449(7164):827–834

Brogden KA, Minion FC, Cornick N, Stanton TB, Zhang Q (eds) (2007) Virulence mechanisms of bacterial pathogens, 1st edn. ASM Press

Devignot S, Sapet C, Duong V, Bergon A, Rihet P, Ong S, Lorn PT, Chroeung N, Ngeav S, Tolou HJ, Buchy P, Couissinier-Paris P (2006) Genome-wide expression profiling deciphers host responses altered during dengue shock syndrome and reveals the role of innate immunity in severe dengue. PLoS ONE 5(7):e11671

Forst CV (2006) Host-pathogen systems biology. Drug Discov Today 11(5–6):220–227

Franke R, Muller M, Wundrack N, Gilles ED, Klamt S, Kahne T, Naumann M (2008) Host-pathogen systems biology: logical modelling of hepatocyte growth factor and *Helicobacter pylori* induced c-Met signal transduction. BMC Syst Biol 2(1):4

Hartman AL, Ling L, Nichol ST, Hibberd ML (2008) Whole-genome expression profiling reveals that inhibition of host innate immune response pathways by Ebola virus can be reversed by a single amino acid change in the VP35 protein. J Virol 82(11):5348–5358

Henderson B, Oyston PCF (eds) (2003) Bacterial evasion of host immune responses, 1st edn, Advances in molecular and cellular microbiology (2). Cambridge University Press, Cambridge

Kanjilal S, Citorik R, LaRocque RC, Ramoni MF, Calderwood SB (2010) A systems biology approach to modeling *Vibrio cholerae* gene expression in virulence-inducing conditions. J Bacteriol 192(17):4300–4310

Pizarro-Cerdá J, Cossart P (2006) Bacterial adhesion and entry into host cells. Cell 124(4):715–727

Schaible UE, Kaufmann SHE (2005) A nutritive view on the host-pathogen interplay. Trends Microbiol 13(8):373–380

Tan SL, Ganji G, Paeper B, Proll S, Katze MG (2007) Systems biology and the host response to viral infection. Nat Biotechnol 25(12):1383–1389

Wang A, Johnston SC, Chou J, Dean D (2010) A systemic network for *Chlamydia pneumoniae* entry into human cells. J Bacteriol 192(11):2809–2815

Host–Pathogen Interactions, Mathematical Models

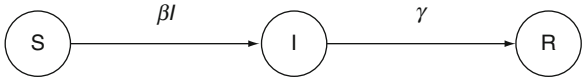
Sumanta Mukherjee[1] and Nagasuma Chandra[2]
[1]Bioinformatics Centre, Indian Institute of Science, Bangalore, Karnataka, India
[2]Department of Biochemistry, Indian Institute of Science, Bangalore, India

Definition

Systems approaches are essential to understand the complex web of host–pathogen interactions (HPIs) that determine the outcome of an infection. Mathematical modeling helps enormously to study emergent

properties of the system, dissect the role of individual components and their interactions, to understand systems difficult to study experimentally and to predict outcomes under a range of scenarios and ultimately to identify strategies for countering the disease.

HPIs constitute typical multicomponent interacting systems, making model-building a daunting task. Given that both host and pathogen are entire systems themselves, their interaction can be viewed as a system of systems. The complexity in HPIs, as in other biological systems, arises through feedback and feed forward controls, bistability, as well as activatory and inhibitory mechanisms. Evaluation of system parameters is often challenging, and it is a common practice to derive them from experimental observations. Currently available models of HPIs span across different levels of hierarchy in biological organizations (► Host–Pathogen Interactions, Fig. 2), the most well studied being epidemiological level models and molecular level models.



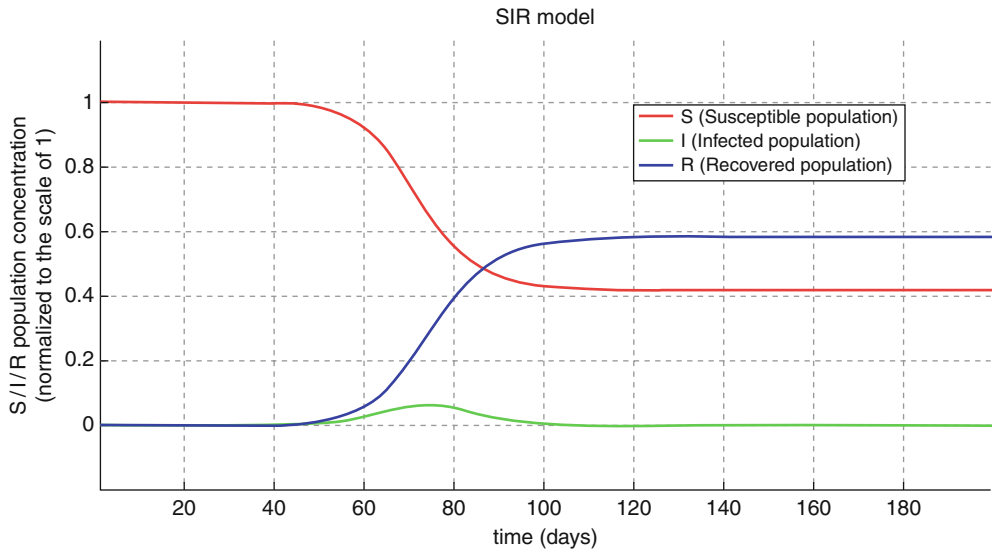
Host–Pathogen Interactions, Mathematical Models, Fig. 1 Compartmental view of the SIR model

Characteristics

Epidemiological Models

Models in this category provide a parametric representation of evolution of the geographical and population-wide spread of an epidemic, over a period of time, addressing questions such as (a) which population is more susceptible to a given disease (b) which strain of a pathogen is more likely to spread and cause disease and (c) which are the key parameters that control the epidemic, hence leading to identifying strategies for disease diagnosis and control.

The SIR model and its variants are the most commonly used models in this category, in which, the population is split into three compartments (Anderson and May 1992), which are (S (t)), susceptible to disease; (I (t)), actively infected population; and (R (t)), population recovered or not vulnerable to disease (Fig. 1). The infectivity rate (β) and the recovery rate (γ) are both assumed to be constants.



Host–Pathogen Interactions, Mathematical Models, Fig. 2 Simulation output of SIR model. The *Red curve* shows the time evolution of the susceptible population over time. *Blue* and *Green lines* show profiles of recovered and infected populations. The population axis has been normalized between 0 and 1 representing (S,I,R) concentrations as a fraction of the population

For short-period and low-fatality infections, the model equations are summarized in (Eq. 1), and a typical simulation output is shown in Fig. 2.

Equation 1 Differential equations representing a simple SIR model.

$$\frac{dS}{dt} = -\beta SI$$

$$\frac{dI}{dt} = \beta SI - \gamma \mathrm{I}$$

$$\frac{dR}{dt} = \gamma I$$

The extent of infectivity in the population will depend upon rates of infectivity, recoverability, and initial susceptible population. The parameter R_o (Basic Reproductive Ratio) is generally defined as $(R_o = \frac{\beta S}{\gamma})$, capturing the expected number of secondary infections arising from a single infected individual. A number of case studies illustrate the usefulness of the SIR model. Studies on the Bombay plague epidemic (1905–1906), the influenza epidemic in an English school (1978), the Eyam plague episode in England, all show very good correlation with predictions by their respective SIR models (Murray 2005).

Modeling Interacting Networks of Biochemical and Biological Species

This set of mathematical models capture interactions in and between cells at the molecular level. These mathematical models are used to predict the evolution of various system parameters over time. Each parameter in the model can either be continuous or discrete in nature. Depending on the nature of the parameter, different modeling strategies are employed, as illustrated in Table 1.

Host–Pathogen Interactions, Mathematical Models, Table 1 Choice of modeling strategies

System parameter	Time	Modeling strategy
Continuous	Continuous	ODE, PDE
Continuous	Discrete	Difference equation
Discrete	Continuous	Network model, integer programming
Discrete	Discrete	Agent-based modeling

Boolean Network

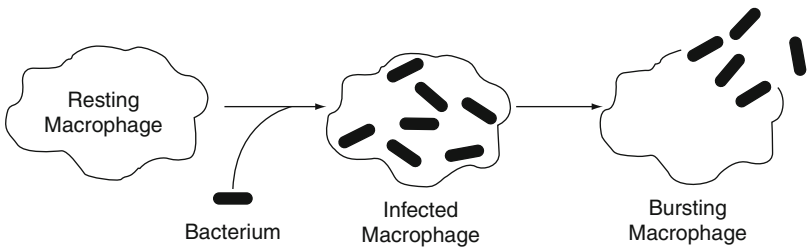
Although a quantitative modeling of the system dynamics provides accurate predictions, the required data for such modeling is often not available, making qualitative modeling the only feasible option. Though qualitative in nature, they are often sufficient to capture the topology of the network and provide significant insights into the system dynamics. Boolean logic functions are used for describing interactions. When the interacting components are represented in the vector form as, $\bar{x} = (x_1, x_2, \ldots, x_n) \in \{0, 1\}^n$, the time updates defined by a set of Boolean functions, time updates are represented as $x_i(t+1) = f_i(\bar{x}(t))\ \mathrm{V}\ i \in \{1 \ldots n\}$, then the dynamics of the network is represented by means of a transition graph. A Boolean network with N variables has 2^{N} distinct states, each state being a set of unique combination of system variables. Formally, $\rightarrow s\{0, 1\}^n \times \{0, 1\}^n$ shows the synchronous transition of the states. Boolean systems too exhibit *attractor* dynamics, similar to other deterministic dynamical systems. An attractor is a distinct set of states, occurring in the time evolution of the system. If x_s is an attractor point, then $x_s = f(x_s)$, where f is the Boolean time evolution rule, which implies that once attained, it remains in that state. Cyclic attractors on the other hand recur in sequence, once any of its states is visited. A given system can have multiple attractors. Different initial states eventually converge to one or the other attractors. The set of states which leads to a particular attractor is collectively called the basin of attraction for the given attractor (Albert et al. 2008). Different variants that are explored are the use of synchronous evolution of states, asynchronous firing of the state transition, and compartmentalization of biological components (Kauffman 1993).

A simplified Boolean model of phagocytosis is shown in Fig. 3, in which a resting macrophage M engulfs extracellular bacteria B_E, and becomes an infected macrophage M_I, rendering extracellular bacteria as intracellular bacteria B_I.

The Boolean rules are as in (Eq. 2).

Equation 2 Boolean rules representing state transition of simple phagocytosis model.

$$M_I = M \, and \, B_E$$



Host–Pathogen Interactions, Mathematical Models, Fig. 3 Macrophages ingesting bacteria and becoming infected. Internalized bacteria proliferate and burst out

$$B_I = (B_E \, and \, M) \; or \; (B_E \; and \, M_I)$$

The host immune response to a Bordetella infection, studied using this approach (Thakar et al. 2007), indicated that the dynamics were different for the freshly infected host, reinfection of the host, and fresh infection with antibody injections. A convalescent host immune response was predicted to be much faster than in a host with antibody transfused treatment, correlating well with experiments using mice (Thakar et al. 2007). In a separate study, a 75 node host–pathogen interactome model, built to study tuberculosis, indicated that the propensity for bacteria to persist was the highest as compared to those for clearance or active proliferation (Raman et al. 2010).

Ordinary Differential Equations (ODES)

Ordinary differential equations serve as powerful tools to build time-continuous models of any system. By first principle, differential equations describe the rate of change of value of a variable with respect to time. The symbolic representation of a differential equation is given as

$$\frac{dy}{dx} \; or \; \dot{y}$$

In notation using limits, $y = \frac{dy}{\lim\limits_{\Delta t \to 0}} = \lim \frac{y(t+\Delta t) - y(t)}{\Delta t}$, representing the slope of the curve $y(t)$ at any given time t.

HPI models generally contain many parameters, giving rise to coupled systems, as shown in the previous SIR model (Eq. 1). Coupled differential equations are a system of ODEs, where each equation signifies the time evolution of one parameter. The parameters are dependent on each other as specified by their respective equations. Hence, the system of equations needs to be solved simultaneously. A simple model from literature is described to illustrate the use of ODEs in studying HPIs. The set of ordinary equations representing the phagocytosis model (Fig. 3) are given in

Equation 3, where, λ is the average rate of macrophage production, γ is the pathogen intake probability by a resting macrophage per contact. δ represents the death rate of a normal macrophage and δ_1 is the mortality rate of infected macrophages. γ_N and γ_I are the average pathogen intake rates by normal resting macrophage and infected macrophages respectively. N is the average number of pathogens thrown out into the system per macrophage burst, and η represents the ingestion rate of phagocytosed pathogen.

Equation 3 System of differential equations representing interactions between macrophages and bacteria.

$$\frac{dM}{dt} = \lambda - \gamma \mathrm{M} B_E - \delta M$$

$$\frac{dM_I}{dt} = \gamma M B_E - \delta_1 M_I$$

$$\frac{dB_E}{dt} = \alpha B_E - \gamma_N M B_E - \gamma_I M_I B_E + \delta_1 N M_I$$

$$\frac{dB_I}{dt} = \gamma_N M B_E + \gamma_I M_I B_E - \delta_1 N M_I - \eta B_I$$

Partial Differential Equations and Compartment Models (PDEs)

Given the complexity and heterogeneity in the host system, it is often difficult to place them into a single differential framework. In order to address the temporal and spatial evolution together, multi-compartment models and partial differential equation based models are employed. In these, the entire space is subdivided into well-mixed smaller compartments, each of which has its own governing equations, while communication among them is allowed through interface equations.

Agent-Based Models

Agent-based models, also known as cellular automata, are hybrid computational models capable of addressing far more complex interaction patterns than any other mathematical tool can. The basic idea is to divide the complete system into multiple subgroups, each group represented by an individual entity on a computational grid. Different individuals are allowed to interact in the computational framework, subject to a set of bounds and predefined rules. The framework also allows introduction of stochastic perturbations, as observed in real systems.

Taking the same example of phagocytosis, agent interactions are defined in a two-dimensional grid with cyclic boundary conditions. The space is initially randomly filled with M macrophages and N bacteria. At time $t = 0,$ none of the macrophages are infected and all bacteria are extracellular. The maximum carrying capacity of each macrophage C and a probability that any bacteria and macrophage contact ends up in phagocytosis P_N are defined. Then the contact is defined with a distance threshold D between extracellular bacteria and macrophages. The different components which are the resting macrophage, the infected macrophage, extracellular bacteria and intracellular bacteria are considered as different types of agents. Resting macrophage has an initial bacterial count of zero. External bacteria are free to move and upon contact with macrophages, get internalized with a probability P_N, after which they are marked as internal bacteria. The internal bacteria have no free mobility and move with the macrophage that ingested it. When the internal bacterial count exceeds the infected macrophage's carrying capacity C, the latter bursts out, resulting in reducing the count of macrophages and all the ingested internal bacteria will be marked back as extracellular bacteria. Though the model proposed here is highly simplified, the strategy can be used to extend the model to simulate real systems very closely. IMMSIM is an example of an agent-based model tool, that has been used to simulate humoral and cellular responses (Kohler et al. 2000).

Game Theoretic Approach

Game theory provides a mathematical framework to address complex issues such as adaptability. For studying HPIs, host and pathogen can be considered as two players making moves based on strategies available to them. Adaptation of each strategy incurs a cost for the player, but geared for maximizing the respective payoffs (Blaser and Kirschner 2007). Pure strategies often do not exist, rendering the problem to one of identifying strategies that maximize payoffs for both players. An example of HPIs in Salmonella as well as mycobacterial infections illustrates the use of game theoretical approaches (Eswarappa 2009).

Summary

As evident from the above discussion, a system can be addressed by different mathematical models, each at a different level of granularity and providing insights from a different perspective. Multi-scale modeling, where different models can be built at different levels and all are threaded together into an integrated framework, appears to be a most plausible direction for future pursuits (Kirschner et al. 2007).

Cross-References

▶ Host–Pathogen Interactions

References

Albert I, Thakar J, Li S, Zhang R, Albert R (2008) Boolean network simulations for life scientists. Source Code Biol Med 3:16

Anderson R, May R (1992) Infectious diseases of humans: dynamics and control. Oxford University Press, Oxford

Blaser MJ, Kirschner D (2007) The equilibria that allow bacterial persistence in human hosts. Nature 449(7164): 843–849

Eswarappa SM (2009) Location of pathogenic bacteria during persistent infections: insights from an analysis using game theory. PLoS ONE 4(4):e5383

Kauffman S (1993) The origins of order: self-organization and selection in evolution. Oxford University Press, Oxford

Kirschner DE, Chang ST, Riggs TW, Perry N, Linderman JJ (2007) Toward a multiscale model of antigen presentation in immunity. Immunol Rev 216:93–118

Kohler B, Puzone R, Seiden PE, Celada F (2000) A systematic approach to vaccine complexity using an automaton model of the cellular and humoral immune system. I. Viral characteristics and polarized responses. Vaccine 19(7–8):862–876

Murray J (2005) Mathematical biology: I. An introduction. Springer, New York

Raman K, Bhat AG, Chandra N (2010) A systems perspective of host-pathogen interactions: predicting disease outcome in tuberculosis. Mol Biosyst 6(3):516–530

Thakar J, Pilione M, Kirimanjeswara G, Harvill ET, Albert R (2007) Modeling systems-level regulation of host immune responses. PLoS Comput Biol 3(6):e109

Host–Pathogen Systems, Target Discovery

Christian V. Forst
Department of Clinical Sciences, Southwestern Medical Center, University of Texas, Dallas, TX, USA

Synonyms

Disease marker identification; Drug targets; Host–pathogen interactions

Definition

Host-pathogen systems describe the interactions of a pathogen with a host. These interactions take place on different level of details, both in time and space, from fraction of a second to the lifetime of the host (e.g., decades), from molecules to whole organisms and societies (in the case of epidemiology).

Target discovery refers to the identification of drug targets, the naturally existing cellular or molecular structure involved in the pathology of interest that potential drugs are meant to act on (see ▶ Epigenetics, Drug Discovery).

Characteristics

Model System

Host-pathogen systems with respect to target discovery are predominantly modeled by interaction networks. The host-pathogen interactome is the most recent focus of genomic technologies. Together with interaction information, genome-wide studies of host-pathogen interactions using mRNA and microRNA transcriptomics, RNA interference (RNAi), proteomics, and other platforms are used to obtain important insights into components and pathways which are essential for pathogen infection, proliferation, and persistence.

The Discovery Process, a Historical Perspective

Target identification has been approached using a variety of genetic and biochemical methods (Terstappen et al. 2007). Historically, pharmacological activities were often discovered by testing plant extracts in complex living systems and observing changes of phenotypes. With the possibility to isolate pharmacologically active substances that are responsible for the observed effects at the beginning of the nineteenth century, a key step toward modern drug discovery was made. With the advances in molecular biology and biochemistry, the approach of testing defined substances in complex living systems was largely abandoned in the 1990s on favor of a more reductionistic target-based approach powered by the sequencing of the human genome and the spawning of a new era with the potential knowledge of all potential drug targets readily available (Overington et al. 2006).

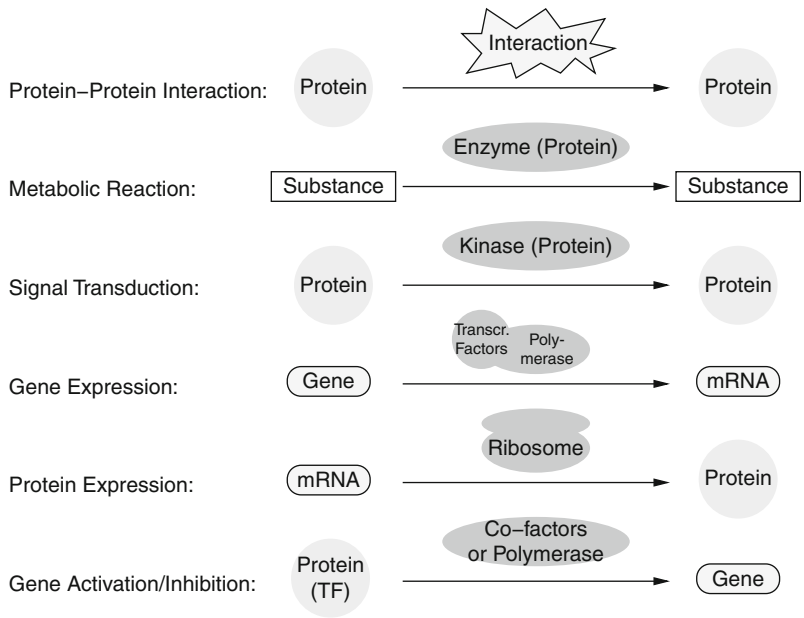
The failed promises of postgenomic target-base drug discovery has recently yielded to revisit a more systems-level approach involving the screening of test compounds under disease conditions to determine induced phenotypic changes (Sams-Dodd 2005). This approach, which goes beyond individual genes and proteins as it involves the investigation of biochemical networks by a systems approach (Butcher 2005) is referred to as chemical genetics. Such chemical genetic screens typically involve the use of genetic (gene deletion, knockdowns, overexpression) as well as nongenetic, environmental (infection with pathogens) perturbations while comparing the influence of lead molecules with and without such permutations. The retroactive identification of pathways and biological functions that underlie the observed phenotypic responses, termed *target deconvolution*, provides important insights into the biological mechanisms of disease and further facilitate drug development (Terstappen et al. 2007).

The Overall Process

Drug target/pathway deconvolution in host-pathogen system is embedded within the modern, holistic, drug-development pipelines including (1) assay development, (2) screening, (3) hits and leads, and (4) target deconvolution. It utilizes chemical genomic approaches and analyzes the experimental results in the context of the host-pathogen interactome. The resulting deconvoluted drug targets and key biological processes are realized as subnetworks of the host-pathogen interactome.

The Interactome

The host-pathogen interactome typically consists of nodes (vertices) and edges that connect the nodes.



Host–Pathogen Systems, Target Discovery, Fig. 1 Types of elementary interactions/reactions in biochemical networks

More complex network representations include hypergraphs (Forst et al. 2006) and rule-based descriptions (Hlavacek et al. 2006). In simple node/edge representations, genes, RNA/DNA, proteins, and chemicals are represented as nodes, binary interactions and reactions are described as edges (Fig. 1). Multiple interactions between chemicals in biochemical reactions are appropriately described by a graph generalization, a hypergraph, where an (hyper) edge can connect more than two nodes. Alternative representations require appropriate ontologies that relate biological concepts using a controlled dictionary (Karp 2000). Examples include network representations in KEGG Kanehisa et al. (2011) and Reactome (Matthews et al. 2008).

Interactome information is readily available on the Internet. The website Pathguide (http://www.pathguide.org) provides a comprehensive list of repositories and resources on protein-protein interactions, metabolic pathways, signaling pathways, pathway diagrams, transcription factors and gene-regulatory networks, protein-compound interactions, genetic interaction networks, as well as other interaction-based resources.

Network Analysis

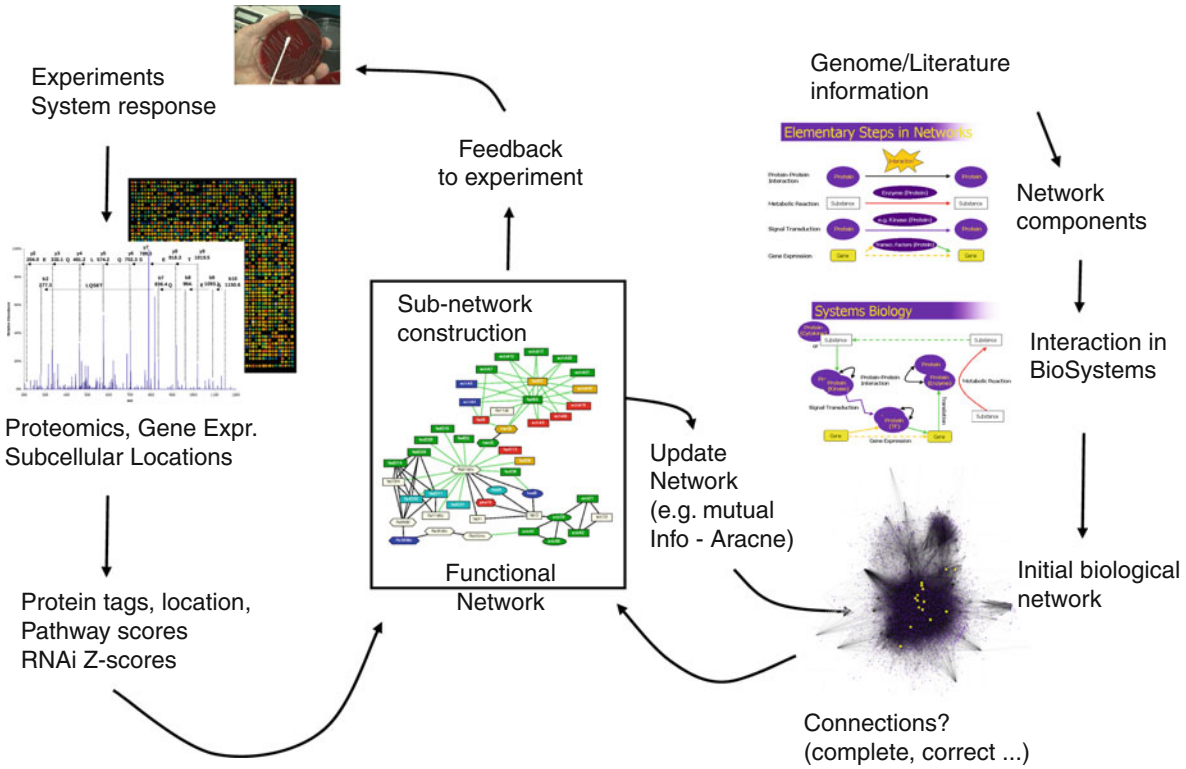
Genome-wide studies of host-pathogen interactions using mRNA and microRNA (miRNA) transcriptomics, global RNA interference (RNAi), proteomics, and other platforms, are revealing important insights into pathways and functional units that are essential for pathogen infection, proliferation, and persistence as well as for host defense.

Genomic screening data are analyzed in the context of biochemical networks. Figure 2 describes a principle approach to analyze host-pathogen networks for target deconvolution. On the left side, the corresponding experiments assays are developed, the experimental data collected and preprocessed. Data involves qualitative sequence tags, presence/absence of proteins, as well as quantitative values such as gene expression, Z-scores after RNAi screens, and drug concentrations. On the right side, interaction information is collected and synthesized into a large biochemical host-pathogen network.

The data is then analyzed in the context of biochemical networks. Typical analysis protocols involve

- Network biology and graph topology
- Network clusters, complexes, and modules
- Response networks
- Pathway enrichment

Network biology was coined by Barabasi and Oltvai 2004 and describes the graph-topological analysis of biochemical networks (Barabasi and Oltvai 2004). His research hypothesized that high connectivity of proteins in protein interaction networks indicate importance with respect to biological functions. Highly



Host–Pathogen Systems, Target Discovery, Fig. 2 Response network analysis. Screening data is analyzed in the context of biochemical networks

connected protein knockouts tend to be lethal for the organism as tested in the case of yeast.

Network clusters, complexes, and modules are used to group subsets of network components (genes, proteins, compounds) into connected subgraphs. Purpose of this exercise is to cluster tightly connected compounds which are thought to be functionally related (Aravind 2000).

The calculation of *response networks* involves of the superposition of local, component-based data, such as gene-expression values, with network information. Response networks of a system refer to best-scored sub-networks in large biochemical networks responding to specific environmental conditions measured by the corresponding experiment (Ideker et al. 2001; Ideker et al. 2002; Cabusora et al. 2005).

Pathway enrichment is a multistep process. It is based on set enrichment analysis identifying statistically significant genes that enrich a particular set. For example, hypergeometric distributions have been used to determine enrichment of gene-ontology nodes by genes. By including quantitative data, such gene-expression profiles, the Gene Set Enrichment Analysis method (GSEA; Subramanian et al. 2005) is capable to determine whether an a priori defined set of genes shows statistically significant, concordant differences between two biological states (e.g., phenotypes). Predefined gene sets (aka Molecular Signature Databases) were extracted, for example, from BioCarta, KEGG (Kanehisa et al. 2011), and Reactome pathways. In a next step, together with biochemical network information, enriched gene sets are used as scaffolds to construct enriched pathways. Thus, the network provides additional context information.

Target Deconvolution

Target deconvolution describes the retrospective identification of targets that underlie observed phenotypic responses. In the case of host-pathogen interactions, the phenotypic responses typically involve survival or

death of the host, virus replication, or host defense – with and without administered drugs. Recent studies involve the identification of human host factors required for influenza virus replication using genome-wide RNAi screens. Shapiro et al. utilize yeast two-hybrid analysis to identify physical associations between host and viral proteins (Shapira et al. 2009). After verification with RNAi screens, a core network that is enriched in RNA-binding proteins, components of WNT signaling, and viral polymerase subunits have been identified as potentially important in influenza infections. König et al. uses multiple analysis approaches including network context to calculate consensus scores for the identification of target host proteins required for WSN virus replication (König et al. 2010). Two hundred and nineteen factors were confirmed to be required for efficient wild-type influenza virus growth, including those involved in kinase-regulated signaling, ubiquitination, and phosphatase activity, and 181 factors assemble into a highly significant host-pathogen interaction network.

All of these studies employ human cell models. Thus the success of their deconvoluted targets have to be further verified in preclinical and clinical scenarios.

Cross-References

- Biological Network Model
- Functional Enrichment Analysis
- Gene Expression
- Gene Ontology
- Gene Set and Protein Set Expression Analysis
- Graph
- Host–Pathogen Interactions
- Hypergraph Theory
- Interactome
- MicroRNA
- Network Clustering
- Pathway
- Proteome
- Response
- RNA Interference
- Signal Transduction
- Topology of Metabolic Reaction Networks
- Transcription Factor
- Transcriptome

References

Aravind L (2000) Guilt by association: contextual information in genome analysis. Genome Res 10:1074–1077

Barabasi A-L, Oltvai ZN (2004) Network biology: understanding the cell's functional organization. Nat Rev Genet 5:101–113

Butcher EC (2005) Can cell systems biology rescue drug discovery? Nat Rev Drug Discov 4:461–467

Cabusora L, Sutton E, Fulmer A, Forst CV (2005) Differential network expression during drug and stress response. Bioinformatics 21:2898–2905

Forst CV, Flamm C, Hofacker IL, Stadler PF (2006) Algebraic comparison of metabolic networks, phylogenetic inference, and metabolic innovation. BMC Bioinformatics 7:67

Hlavacek WS, Faeder JR, Blinov ML, Poser RG, Hucka M, Fontana W (2006) Rules for modeling signal-transduction systems. Science STKE, 344:re6, doi:10.1126/stke.3442006re6

Ideker T, Ozier O, Schwikowski B, Siegel AF (2002) Discovering regulatory and signalling circuits in molecular interaction networks. Bioinformatics 18(Suppl 1):S233–S240

Ideker T, Thorsson V, Ranish JA, Christmas R, Buhler J, Eng JK, Bumgarner R, Goodlett DR, Aebersold R, Hood L (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. Science 292:929–934

Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M (2011) KEGG for integration and interpretation of large-scale molecular data sets. Nucl Acids Res doi:10.1093/nar/gkr988

Karp PD (2000) An ontology for biological function based on molecular interactions. Bioinformatics 16:269–285

König R, Stertz S, Zhou Y et al (2010) Human host factors required for influenza virus replication. Nature 463:813–817

Matthews L, Gopinath G, Gillespie M, Caudy M, Croft D, de Bono B, Garapati P, Hemish J, Hermjakob H, Jassal B, Kanapin A, Lewis S, Mahajan S, May B, Schmidt E, Vastrik I, Wu G, Birney E, Stein L, D'Eustachio P (2008) Reactome knowledgebase of human biological pathways and processes. Nucleic Acids Res 37:D619–D622

Overington JP, Al-Lazikani B, Hopkins AL (2006) How many drug targets are there? Nat Rev Drug Discov 5:993–996

Sams-Dodd F (2005) Target-based drug discovery: is something wrong? Drug Discov Today 10:139–147

Shapira SD, Gat-Viks I, Shum BOV et al (2009) A physical and regulatory map of host-influenza interactions reveals pathways in H1N1 infection. Cell 139:1255–1267

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA 102:15545–15550

Terstappen GC, Schlupen C, Raggiaschi R, Gaviraghi G (2007) Target deconvolution strategies in drug discovery. Nat Rev Drug Discov 6:891–903

Hough Transform

Virginio Cantoni[1,2] and Elio Mattia[3]
[1]Department of Computer Engineering and Systems Science, University of Pavia, Pavia, Italy
[2]Computational Biology, KTH Royal Institute of Technology, Stockholm, Sweden
[3]Centre for Systems Chemistry, Stratingh Institute for Chemistry, Rijksuniversiteit Groningen, Groningen, The Netherlands

Definition

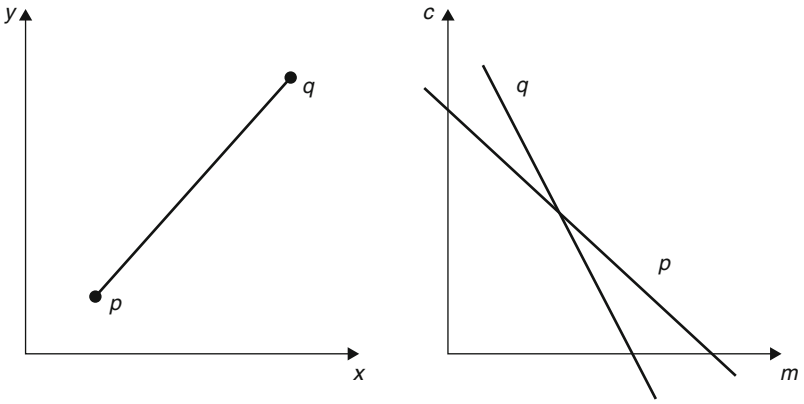
The Hough transform (HT) is a coordinate transformation introduced by Hough (Hough 1962). It is useful in computer vision as a method for retrieving shapes within digital images. It was first conceived for lines, circumferences, and simple polygons (Duda and Hart 1972) and later generalized to arbitrary shapes (Ballard 1981).

HT is best explained for the case of retrieval of lines. HT maps every meaningful point of an x–y image space into a line of an m–c parameters space. Meaningful features of an image for the purpose of HT are identified by retrieving points that have high gradient values, as these points could possibly belong to the contour of, e.g., a straight or curved line; computationally, they can be determined by edge detection operators. Figure 1 depicts the HT of points p and q of the image on the left into lines in the parameters space on the right. The various points of the HT line of a point p give the set of slope (m) and intercept (c) values of the bundle of lines which contain the point p.

In the example of Fig. 1, applying HT to the points in the segment $[p, q]$ results in the mapping of the latter segment into a bundle of lines which will intersect into a particular point (m_0, c_0). This point will correspond to the actual slope and intercept the line of the image space which contains the segment $[p, q]$. Computationally, the HT is performed by voting (m, c) values in a quantized parameters space; in the example of Fig. 1, (m_0, c_0) will result in a higher final vote, both indicating success in retrieving a line and providing its actual slope and intercept.

The number of points into which one single point of the image space is mapped can be enormously reduced if edge detection operators not only provide information about the location of meaningful points in an image, but also about the approximate orientation of the curve on which points are possibly located. This variation is the so-called gradient method (GM). Although the GM makes line retrieval by HT a trivial process, it only *facilitates* retrieval of more complex shapes such as polygons – for which even with the aid of the GM, segments have to be voted.

In the generalized Hough transform (GHT), retrieval of shapes of any nature – e.g., a wrench – is accomplished with the aid of the GM, to help establish relative orientations. One *reference* point internal to the object to be retrieved is established, then a table containing information about distance and orientation of selected *control* points of the object – e.g., 10 points located on peculiar features of the object – with respect to the reference point is built. GHT is then performed on a given image by performing HT, in a parameters space corresponding to the image space; HT is performed multiple times per meaningful point – e.g., 10 times per point of the image, in our example – supposing that it



Hough Transform, Fig. 1 The Hough transform duality between edge points and straight line parameters: two points of a segment of the image space (*left*) and two lines of the parameters space (*right*)

might correspond to any control point previously established. If the object of interest is present in the image, one point in the image space will be highly voted; it will correspond to the reference point of the object and will be voted *once* by every control point of the object actually in the image.

The advantages of HT are efficiency, rotation- and translation-insensitiveness, parallelizability, noise and occlusion insensitiveness, and relative tolerance to approximate descriptions. Among its disadvantages, its high computational complexity, which results in high time and memory requirements. Randomized and deterministic HT variations help dealing with these downsides.

Cross-References

▶ Hough Transform, Structural Motif Retrieval

References

Ballard DH (1981) Generalizing the Hough transform to detect arbitrary shapes. Pattern Recognit 13:111–122

Duda RO, Hart PE (1972) Use of the Hough transformation to detect lines and curves in pictures. Commun ACM 15:11–15

Hough PVC (1962) Methods and means for recognizing complex patterns. US patent 3069654

Hough Transform, Structural Motif Retrieval

Virginio Cantoni[1,2] and Elio Mattia[3]
[1]Department of Computer Engineering and Systems Science, University of Pavia, Pavia, Italy
[2]Computational Biology, KTH Royal Institute of Technology, Stockholm, Sweden
[3]Centre for Systems Chemistry, Rijksuniversiteit Groningen, Groningen, The Netherlands

Definition

The Hough transform can be used efficiently in the context of protein structural comparison and motif retrieval, thereby constituting another fundamental tool among the available heuristics that allow to estimate the presence of particular structural features within the structure of a protein.

Characteristics

Systems biology benefits enormously from automatic methods aimed at sorting out meaningful information and overall trends from the huge amount of data about biological systems that has been, is being and will be collected with modern high-throughput techniques.

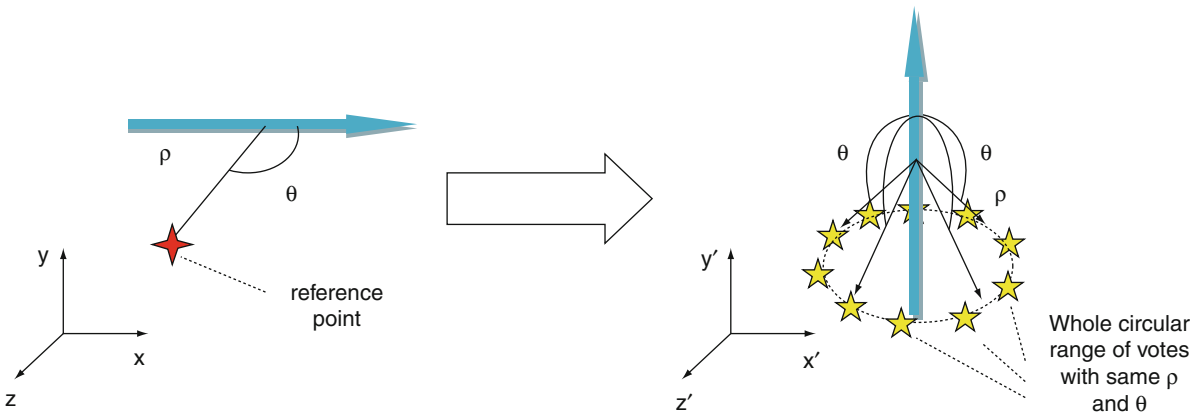
Among these methods, it is widely acknowledged that a substantial role in carrying out data mining nowadays is played by protein structural comparison. Known functional units such as structural motifs can be retrieved in recently discovered proteins and similarities between known and new structures can be established. This allows inferring protein function, explaining the role of specific sequences in biochemical pathways and biological networks, building up phylogenetic trees and ultimately creating new database annotations, which in including the results of the completed predictions will be helpful for the structures that will be discovered in the future.

Various approaches have been developed for protein structure comparison, based on either distance matrices (Taylor and Orengo 1989; Holm and Sander 1993), graph theory, or geometric hashing (Nussinov and Wolfson 1991; Comin et al. 2004). The various available algorithms are heuristics.

It is fundamental, in this context, to be acquainted with the notion of heuristic: An experience-based or intuition-guided problem-solving technique, which is generally fast at the price of suboptimality, i.e., there is generally no theorem to fully guarantee the reliability of the results they provide. Heuristics are employed either when just no optimal approach is available at all or when, despite availability of an optimal method, the heuristic approach is way faster and yields results with an acceptable accuracy given the inherent gain in execution time.

The reason for the existence of multiple solutions of the same protein structural comparison problem is that disparate inspirations lead to diverse approaches, very often based on very different lines of reasoning.

It turns out that the existence of multiple heuristics for protein structure comparison is actually vital and the reason for this is that it is very difficult to



Hough Transform, Structural Motif Retrieval, Fig. 1 The principle of applying the Hough transform to protein structure comparison. *Left*: model protein; *right*: object protein (voting space)

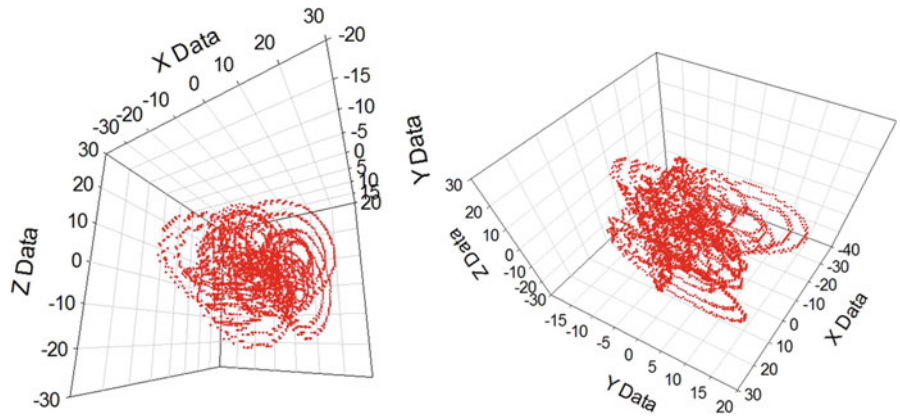
establish quantitatively the distance from optimality of suboptimal structural comparison results. Therefore, only when such methods yield results which are in accordance with one another, despite the fact that the nature of the calculations can be profoundly different, can the predictions be deemed accurate.

The purpose of this entry is to illustrate a recently developed heuristic approach for assessing protein structure similarity and for retrieving structural motifs, which is inspired by a computational method imported from the field of computer vision: the ▶ Hough transform (Mattia 2010).

In this method, protein similarity is determined by performing a comparison between pairs of protein structures and calculating a comparison score. Motif retrieval is allowed for by letting one of the proteins in the comparison be the structural motif which is to be searched for. The comparison score is calculated in a vote space, which generally corresponds to the coordinate space in which one of the proteins is described, by appropriately aligning couples of secondary structures of the two proteins (e.g., alpha-helices and beta-strands), one from each of the latter, and voting selected reference points. Either the vote of the mostly voted point in the voting space or other functions of the votes in the voting space can be used as the final comparison score.

The method can be easily tuned so that voting is performed only between secondary structures of similar type, e.g., alpha-helices with alpha-helices and beta-strands with beta-strands. Higher information content in the voting procedure always results in a cleaner voting space and in a better signal-to-noise ratio.

The principle of the method is illustrated in Fig. 1. On the left is the xyz space of the so-called model protein, while on the right is the $x'y'z'$ space of the so-called object protein. Voting is performed in the latter space. In order to do so, for every secondary structure of the model protein (illustrated as a bold arrow on the left of Fig. 1), the characteristic parameters rho and theta must be computed; these correspond to the distance of the secondary structure to a well-defined reference point, such as the geometric center of the protein, and the angle that the direction of the secondary structure forms with the segment joining the latter with the reference point. The parameters rho and theta allow to vote reference points in the voting space. Knowledge of only two parameters in a 3D space results in incomplete definition of the position of the point to vote, making voting of an entire circumference indispensable. This circumference is the rim of a cone centered in the object protein's secondary structure, as Fig. 2 illustrates. If the secondary structures of the object proteins have a similar spatial arrangement as the ones of the model protein from which the rho and theta parameters are taken from, then, after many voting steps, i.e., when each of the rho and theta values from all of the secondary structures of the model protein has been used to vote circumferences around every secondary structure of the object protein, as many circumferences as the number of secondary structures in the model protein will all intersect, in the voting space, in one highly voted point,



Hough Transform, Structural Motif Retrieval, Fig. 2 Example of a filled voting space, in which circumferences from different secondary structures interfere to create voting peaks (two different perspectives)

which can be used as a score for the comparison. This will not happen whether the two proteins have different structures, resulting in a low score. Figure 2 illustrates a sample voting space, in which successful comparison results in the formation of vote peaks due to the voting circumferences intersecting with one another.

Fundamental for the implementation of the method is discretization. Both the geometric space where voting is performed and the voting circumference are discretized, the former in "voxels" (cubic volume elements), the latter in an integer number of steps. The entity of the discretization deeply influences the quality of the results: A high-detail mesh produces more reliable scores, although at the price of longer execution times, making a compromise between parameter settings and acceptable execution times necessary. Overwhelming memory usage is avoided by maintaining information of only the voxels that contain a nonzero vote.

The most important aspect of the implementation regards the way of dealing with the voting space after it has been filled. Since structural similarity does not mean structural identity, the circumferences in the voting space will not actually overlap perfectly, although they will come to close proximity around one point. This results in the votes not forming the actual peaks that are wanted in order to compute a score. For this reason, smoothing of the vote space by accumulation of all of the votes sufficiently near to every point in the vote space is needed. Performing a smoothing of the vote space by merely scrolling the vote space and for every point summing every vote that happens to be sufficiently close to it (thereby scanning the vote space again for every point in it) is computationally costly, to the point that it easily becomes too heavy to be executed in reasonable times for ordinary purposes. The algorithmic complexity is in this case of $O(N^2)$, where N is the number of votes in the vote space, which is in turn proportional to the number of secondary structures in the model protein and in the object protein and to the number of steps in which the voting circumference is divided.

The problems associated with the high computational requirements of the smoothing algorithm are solved if a particular data structure is introduced in the implementation, i.e., the ▶ range tree. The use of range trees in the smoothing step allows the computational complexity to fall down to $O(N\log^3 N)$. Balancing of the range trees guarantees that the computational complexity stick to the logarithmic order above, without worst-case scenarios.

The execution times vary strongly depending on the size of the input, i.e., how many proteins to compare and how many secondary structures they contain, and the parameters settings. A typical execution time for a one-to-one comparison is from a fraction of second to a few seconds on a standard laptop. Noteworthy is the change in execution times that the use of range trees brings about: A 33-to-33 proteins comparison took 8 h to complete with the standard algorithm (without range trees), while only 11 min with the optimized algorithm (with range trees).

The algorithm is embarrassingly parallel. This means that, since it requires very little communication between independent voting steps, it is easily split into parallel tasks, resulting in faster execution, which is fundamental for operation in the context of database annotation.

All in all, as it has been pointed out in the introductory paragraphs that results of structural alignment from different methods which are in reciprocal agreement help to confirm positive predictions, the Hough transform method, which has proven successful, is therefore a fundamental component of the suite of protein structural comparison algorithms available to date.

The method has its own peculiar advantages, too, which are inherited directly by the algorithm it is based upon, the ▶ Hough transform. Among them, efficiency, rotation- and translation-insensitiveness, parallelizability, noise and occlusion insensitiveness, and relative tolerance to approximate descriptions. The method is also highly parameterized, so that it can be adapted to specific cases in order to obtain the best results. The main downside is its suboptimality, as is the case with any heuristic. Other specific disadvantages of the Hough transform, such as high time and memory requirements, and of range trees, such as the inherent difficulties in treating complex data structures, have been successfully dealt with and solved.

Cross-References

▶ Hough Transform
▶ Range Tree

References

Comin M, Guerra C, Zanotti G (2004) PROuST: a comparison method of three-dimensional structures of proteins using indexing techniques. J Comput Biol 11–6:1061–1072
Holm L, Sander C (1993) Protein structure comparison by alignment of distance matrices. J Mol Biol 233:123–138
Mattia E (2010) Protein structure comparison and motif retrieval by the generalized Hough transform. IUSS Thesis, Pavia, Italy
Nussinov R, Wolfson H (1991) Efficient detection of three-dimensional motifs in biological macromolecules by computer vision techniques. Proc Natl Acad Sci USA 88:10495–10499
Taylor WR, Orengo CA (1989) Protein structure alignment. J Mol Biol 208:1–22

HP1

▶ Heterochromatin Protein 1

HPC

▶ Large-Scale and High-Performance Computing

HTLV

▶ HTLV, Cellular Transcription

HTLV, Cellular Transcription

Christian Schönbach
Department of Bioscience and Bioinformatics, Kyushu Institute of Technology, Iizuka, Fukuoka, Japan

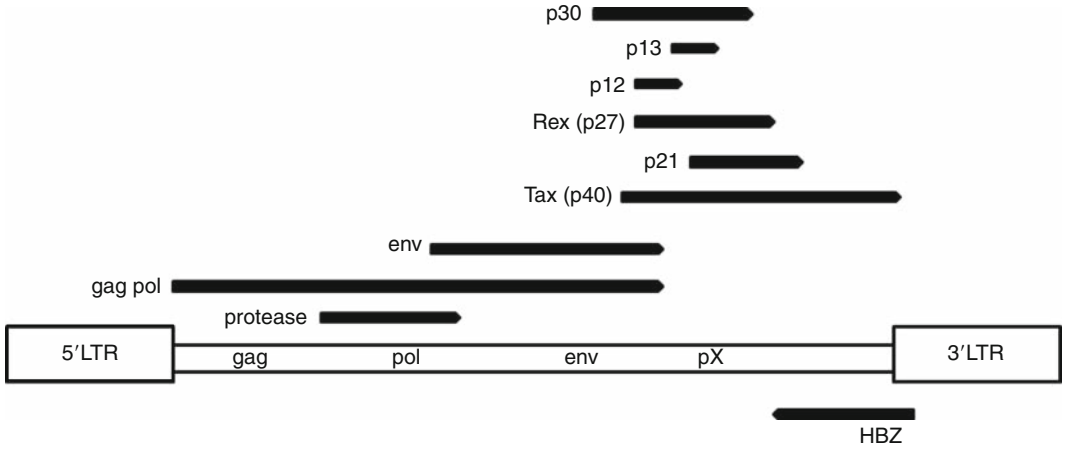
Synonyms

HTLV; Human T-lymphotropic virus

Definition

Human T-lymphotropic viruses (HTLV) are members of the ▶ Deltaretroviridae. HTLV-1, the only known human oncogenic retrovirus is the etiological agent of ▶ adult T-cell leukemia/lymphoma (ATL) and ▶ Human T-Lymphotropic Virus Type-I-associated Myelopathytropical Spastic Paraparesis (HAM/TSP). HTLV-2 is associated with HAM/TSP-like illness, whereas the pathogenicities of HTLV-3 and HTLV-4 are unknown. HTLVs share a similar proviral genomic organization. Long terminal repeats (LTR) flank the structural genes *gag*, *pol*, and *env* (Fig. 1). The *pX* region encodes the regulatory protein Tax and so-called accessory proteins Rex, p21, p12, p13, and p30. The basic leucine zipper factor HBZ is translated from an antisense transcribed mRNA of the 3′LTR region (Matsuoka and Jeang 2007; Higuchi and Fujii 2009).

$CD4^+$ T-cell transformation by HTLV-1, viral persistence, and immune response modulation are driven by the highly pleiotropic oncoprotein Tax1 (Tax of HTLV-1) in conjunction with HBZ, p12, p13, Rex1, and p30. Env, Rex, Tax, and dendritic cells are associated with HTLV-1 tropism for infecting $CD4^+$ T-cells (Jones et al. 2008; Boxus and Willems 2009).



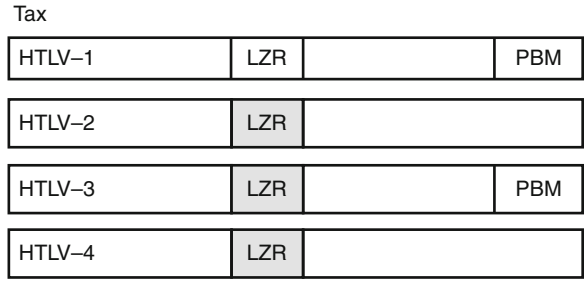
HTLV, Cellular Transcription, Fig. 1 HTLV proviral genome and transcripts

Tax1 can interact with hundreds of cellular proteins to transform $CD4^+$ cells (Boxus et al. 2008). Protein complex formation of Tax1 with cellular proteins CREB2, ATF2, EP300, and KATB2 is essential for Tax1-mediated viral and cellular transcription initiation. Major Tax1 targets are the non-canonical NFκB and AKT1 signaling pathways including transcription factors AP1, SRF, and TP53 (Matsuoka and Jeang 2007). Various feedback mechanisms on viral and cellular transcription level ensure HTLV-1 persistence and/or immune escape from recognition by Tax1-specific cytotoxic T-cells. For example, Tax1-mediated T-cell immortalization is suppressed by HBZ heterodimers with CREB2 or JUN, whereas HBZ transcription is activated by Tax1 (Boxus and Willems 2009; Matsuoka 2010). p30 binding to EP300 can modulate transcription of both, viral and cellular genes. Depending on the level of available p30, it may either stabilize Tax1-CREB2-ATF2-EP300-KATB2 complex or compete for EP300 and suppress transcription initiation (Bai et al. 2010). Upon transformation, mutations in the Tax1-coding region, deletions and/or methylation of 5′LTR prevents Tax1 expression.

Characteristics

Tax-Mediated Transcriptional Changes

Tax1 leucine zipper–like region (LZR) mediates the activation of the non-canonical NFκB pathway, which is essential in T-cell transformation. Complex formation



HTLV, Cellular Transcription, Fig. 2 Comparison of HTVL Tax motif/domain organization

of Tax1 with CHUK, IKBKG, and NFKB2 leads to nuclear translocation of RELB and NFKB2 heterodimers which induce the expression of various cell proliferation–promoting cytokines (IL1B, IL2, IL6, IL9, IL13, and IL15) and their corresponding receptors (Boxus and Willems 2009). The LZRs of HTLV-2, -3, and -4 are more similar to each other than to HTLV-1 LZR (Fig. 2) (Higuchi and Fujii 2009). Accordingly, HTLV-1 but not HTLV-2 in vitro infection of IL2-dependent cell lines results over time in IL2-independent growth (Boxus et al. 2008).

HTLV-1 complements host cell transformation with Tax1-driven aneuploidic or improper chromosomal segregation effects and clastogenic or mismatch repair–associated DNA damage (Matsuoka and Jeang 2007; Boxus and Willems 2009). The Tax1 C-terminal PDZ domain–binding motif (PBM) has been associated not only with cell proliferation but also genomic instability–promoting properties. Tax proteins of HTLV-2 and -4 lack PBM (Fig. 2).

Modulation of the cell cycle by Tax1 and HBZ in conjunction with epigenetic changes promotes the clonal expansion of transformed $CD4^+$ T-cells via mitotic proliferation. On the other hand, Tax1-specific cytotoxic $CD8^+$ T-cells select against clonal proliferation. Overall, the relatively low infection efficacy of HTLV-1 and multifactorial dependence on $CD4^+$ T-cell transformation and proliferation capacity of abnormal cells results in latency periods of several decades and up to 5% ATL incidence. In case of HAM/TSP, the latency period is shorter, and patients show a vigorous HTLV-1-specific cytotoxic $CD8^+$ immune response accompanied by pro-inflammatory cytokine production, which contribute to the neuropathogenicity (Matsuoka and Jeang 2007).

Effects of Accessory Proteins

While Tax1 is sufficient and necessary to transform T-cells, accessory proteins are equally important for the survival of HTLV-1 in the host cells and progression toward ATL or HAM/TSP as outlined in excellent reviews of p12 (Van Prooyen et al. 2010), p13 (Silic-Benussi et al. 2010), p30 (Bai et al. 2010), and HBZ (Matsuoka 2010) in a special issue of *Molecular Aspects of Medicine*. Some of the multifunctional accessory proteins have Tax1 synergistic and antagonistic functions to promote immune escape and modulate cell proliferation, viral replication, and persistence.

p12

p12 protein usually localizes to the endoplasmic reticulum (ER) and cis-Golgi. Yet, proteolytic cleavage of p12 at a non-canonical ER retention signal produces p8, which localizes to lipid rafts of the immunological synapse and affects T-cell receptor (TCR) signaling. In addition, interaction of p8 with linker for activation of T-cells (LAT) decreases LAT phosphorylation. As a result, suppressed T-cell receptor signaling downregulates the activity of NFAT transcription factor family members which decreases cytokine production and T-cell proliferation (Van Prooyen et al. 2010).

ER-located p12 promotes in synergy with Tax1 T-cell proliferation and viral persistence by activating both IL2 and IL2R expression. p12 interaction with CALR and CANX increases cytoplasmic Ca^{2+} levels and dephosphorylation of NFATC1, which locates to the nucleus to activate IL2 transcription. In turn, the increase in IL2 activates Tax1 transcription via CREB2 and ATF2. At the same time ER-located p12 also binds to the immature forms of the IL2Rβ and γ_c chains, which enhance STAT5 activation and TCR signaling in response to IL2. To avoid immune recognition of the infected proliferating T-cells, p12 binds in the ER to HLA class I molecules which reroutes HLA class I-trafficking to the proteasome for degradation rather than to the cell surface (Van Prooyen et al. 2010).

p13

p13 has dual subcellular localization potential. Mainly, p13 localizes to mitochondria where it increases reactive oxygen species production and reduces mitochondrial Ca^{2+} uptake, which influences the proliferation and death of T-cells. Tax1-mediated ubiquitylation facilitates sorting of p13 to the nucleus. Subsequent heterodimerization with Tax1 inhibits CREBBP and EP300 co-activator binding to Tax1 and attenuates Tax-mediated transcription of HTLV-1 genes in a potential negative feedback loop (Silic-Benussi et al. 2010).

Rex1 (p27)

The RNA-binding post-transcriptional regulator Rex1 (p27) controls the splicing and export of HTLV RNAs transcribed from the *pX* regions, including its own RNA. Rex1 is essential for the production and assembly of viral particles which enables active infection (Boxus and Willems 2009; Bai et al. 2010).

p30

p30 antagonizes the effects of Tax1 on both, transcriptional and post-transcriptional levels. Direct interaction of p30 with CREBBP and EP300 suppresses the transcription of HTLV-1 from the 5′LTR and of cellular genes that are dependent on CREBBP/EP300 activation. Another cellular target of p30 is transcription factor SPI1 also known as PU.1. Binding of p30 to the ETS domain of SPI1 prevents its binding to DNA and therefore transcriptional activation of its target genes (e.g., *TLR4*) including itself. Downregulation of TLR4 expression interferes with the innate immune response, induction of pro-inflammatory cytokines (e.g., IL8, TNFA, CCL2, etc.) and production of anti-inflammatory IL10 in TLR4-primed dendritic cells (Bai et al. 2010).

Post-transcriptionally p30 interaction with the large ribosomal subunit protein L18a and binding to Tax1 and Rex1 mRNAs were found to increase their nuclear retention, thereby suppressing viral replication and

possibly prolonging viral latency. On cellular level, p30 increases via an unknown mechanism the phosphorylation of GSK3B kinase in macrophages, which leads to an increase in IL10 production (Bai et al. 2010).

HBZ

HBZ gene is transcribed from the 3′LTR which is not known to undergo methylation and/or deletions. Therefore, HBZ is expressed at all stages of infection. Both HBZ mRNA and protein modulate cellular targets. HBZ protein forms heterodimers with CREB, CREB2, and p300/CBP which attenuate Tax-mediated transcription of HTLV-1 genes from the 5′LTR. A stem-loop structure of HBZ mRNA promotes the proliferation of infected and leukemic cells, and increases the transcription of *E2F1*. Possibly, the increased cell proliferation is associated with the induction of E2F1 target genes. In addition, HBZ mRNA may modulate the phenotype of T-cells (Matsuoka 2010).

HTLV and Transcription of Cellular miRNAs

HTLV-1 infection affects also the expression of small RNAs. Ruggero and co-workers (Ruggero et al. 2010) reviewed reports of numerous miRNAs that were found to be up- or downregulated in infected T-cells and ATL cells. For example, in ATL cells, miR-93 and -130b downregulate TP53INP1, an apoptosis- and cell cycle arrest–promoting tumor suppressor. In HTLV-1-infected T-cell lines, Tax1 upregulates miR-146a and miR-130b via the NF-κB pathway. Other Tax1-modulated small RNAs include DNA-directed RNA polymerase III–transcribed transfer RNAs (tRFs) and miRNAs, which affect cell proliferation and cell cycle progression.

Conclusions

In the past 5 years, HTLV-1 and -2 studies have considerably contributed to the understanding of the molecular mechanism of T-cell transformation and viral manipulation of cellular pathways at both transcription and protein levels. Potential therapeutic tools (e.g., PI3K inhibitors, allogeneic hematopoietic stem cell transplantation, or mutant Tax vaccination) and diagnostic markers such as miRNAs have emerged (Matsuoka and Jeang 2007; Ruggero et al. 2010), but efficient therapies and diagnostic markers have yet to be developed.

Cross-References

- ▶ Adult T-Cell Leukemia/Lymphoma
- ▶ Deltaretroviridae
- ▶ Human T-Lymphotropic Virus Type-I-associated Myelopathytropical Spastic Paraparesis

References

Bai XT, Baydoun HH, Nicot C (2010) HTLV-I p30: a versatile protein modulating virus replication and pathogenesis. Mol Aspects Med 31(5):344–349

Boxus M, Willems L (2009) Mechanisms of HTLV-1 persistence and transformation. Br J Cancer 101(9):1497–1501

Boxus M, Twizere JC, Legros S, Dewulf JF, Kettmann R, Willems L (2008) The HTLV-1 Tax interactome. Retrovirology 5:76

Higuchi M, Fujii M (2009) Distinct functions of HTLV-1 Tax1 from HTLV-2 Tax2 contribute key roles to viral pathogenesis. Retrovirology 6:117

Jones KS, Petrow-Sadowski C, Huang YK, Bertolette DC, Ruscetti FW (2008) Cell-free HTLV-1 infects dendritic cells leading to transmission and transformation of CD4(+) T cells. Nat Med 14(4):429–436

Matsuoka M (2010) HTLV-1 bZIP factor gene: Its roles in HTLV-1 pathogenesis. Mol Aspects Med 31(5):359–366

Matsuoka M, Jeang KT (2007) Human T-cell leukaemia virus type 1 (HTLV-1) infectivity and cellular transformation. Nat Rev Cancer 7(4):270–280

Ruggero K, Corradin A, Zanovello P, Amadori A, Bronte V, Ciminale V, D'Agostino DM (2010) Role of microRNAs in HTLV-1 infection and transformation. Mol Aspects Med 31(5):367–382

Silic-Benussi M, Biasiotto R, Andresen V, Franchini G, D'Agostino DM, Ciminale V (2010) HTLV-1 p13, a small protein with a busy agenda. Mol Aspects Med 31(5):350–358

Van Prooyen N, Andresen V, Gold H, Bialuk I, Pise-Masison C, Franchini G (2010) Hijacking the T-cell communication network by the human T-cell leukemia/lymphoma virus type 1 (HTLV-1) p12 and p8 proteins. Mol Aspects Med 31(5):333–343

HTML

Tim Clark
Department of Neurology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA

Synonyms

Hypertext markup language

Definition

HTML is the principal markup language of the ▶ World Wide Web. It specifies the structural semantics for text in Web documents. Its elements indicate to Web browsers how to display content on a Web page, including text, images, video, and audio. It also can embed programming instructions to the browser using JavaScript, or other scripting languages.

A separate language, CSS (Cascading Style Sheets), is used to specify the presentation of the page elements, such as typefaces and styles, layout, colors, etc.

Elements of HTML consist of tags enclosed in angle brackets, inserted into the content, like this example in HTML5:

```
<!DOCTYPE html>
<html>
      <head>
          <title>Definition of HTML
          </title>
      </head>
      <body>
          <p>HTML is the markup
          language of the World Wide
          Web.</p>
      </body>
</html>
```

Tags plus CSS indicate to the browser how the content should be rendered.

HTML 4.01 is the most recent fully standardized W3C version of HTML. HTML5, with a number of attractive new features, is currently usable in "working draft" state. It is under parallel and coordinated development by both the W3C (http://w3.org) and the WHATWG (http://www.whatwg.org).

Cross-References

▶ World Wide Web

References

Hickson I (ed) (2011) HTML5: a vocabulary and associated APIs for HTML and XHTML, W3C Working Draft 25 May 2011. World Wide Web Consortium

Hickson I, Schwartz B (eds) (2012) HTML5: a technical specification for Web developers. WHATWG. http://developers.whatwg.org/

Raggett D, Hors AL, Jacobs I (eds) (1999) HTML 4.01 specification, W3C recommendation 24 December 1999. World Wide Web Consortium. http://www.w3.org/TR/REC-html40/

HTTP

Tim Clark
Department of Neurology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA

Synonyms

Hypertext transfer protocol

Definition

HTTP is a generic stateless Internet application protocol for distributed and collaborative ▶ hypertext- and hypermedia-based information systems. It provides the foundation for data communications on the *World Wide Web*. It is currently the dominant Internet protocol.

The HTTP protocol has a circumscribed set of allowed actions to access its target, which is a distributed resource specified by the ▶ URL or Uniform Resource Locator string. Resources are resolved, interpreted, and acted upon according to client request, by computers running Web server software (e.g., Apache httpd).

The actions supported by HTTP are constrained to a relatively limited set. Programmatic Web services that conform to this set of allowed actions (GET, HEAD, PUT, POST, DELETE, TRACE, OPTIONS, and CONNECT) are called "resource-oriented" services and are said to conform to the REST (representational state transfer) architectural pattern – they are said to be "RESTful."

Secure communications on the Web for such purposes as financial transactions are implemented using the ▶ HTTPS protocol, which encrypts the Internet transport layer data transmitted by HTTP.

HTTP standards development has been a global collaborative engineering effort, coordinated by the World Wide Web Consortium (http://w3.org) and the Internet Engineering Task Force (http://www.ietf.org).

Cross-References

► World Wide Web

References

Fielding RT, Taylor RN (2002) Principled design of the modern web architecture. ACM Trans Internet Technol 2(2):115–150

Fielding R, Getty J, Mogul J, Frystyk H, Masinter L, Leach P, Berners-Lee T (1999) Hypertext transfer protocol – HTTP/1.1, IETF RFC 2616. Internet Engineering Task Force. http://www.ietf.org/rfc/rfc2616.txt

Jacobs I, Walsh N (2004) Architecture of the World Wide Web, vol 1. In: W3C recommendation. World Wide Web Consortium. http://www.w3.org/TR/webarch/

HTTP Secure

► HTTPS

HTTPS

Tim Clark
Department of Neurology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA

Synonyms

HTTP secure; Hypertext transfer protocol secure

Definition

► HTTPS is an encrypted version of the ► HTTP application-layer Internet protocol for enacting secure transactions on the ► World Wide Web, such as in financial transactions and for confidential corporate and other data. HTTPS syntax is identical in all respects to the standard HTTP protocol except that resource requests, as well as the returned data, are encrypted by the SSL (Secure Sockets Layer) or TLS (Transport Layer Security) protocols using symmetric encryption technologies such as AES or RC4.

Before establishing a secure link, communicating nodes must authenticate themselves – they must present credentials that reliably establish their identities. In HTTPS, Web browsers and services determine whether or not to trust servers, by authenticating an X.509 public-key cryptographic identity certificate provided by the server. X.509 certificates are signed by a trusted Certification Authority (e.g., Verisign, Microsoft), or self-signed by the server's own organization.

The Electronic Frontier Foundation (https://www.eff.org/) recommends encryption of Web traffic whenever possible. "HTTPS Everywhere" (https://www.eff.org/https-everywhere) browser extensions are now available to support this recommendation.

References

Apache Software Foundation (2011) SSL/TLS strong encryption: an introduction. Apache Software Foundation. https://httpd.apache.org/docs/2.3/ssl/ssl_intro.html

Dierks T (2008) Rescorla E: IETF RFC 5246: the transport layer security (TLS) protocol, version 1.2. In: IETF Requests for Comment, vol 5246. Internet Engineering Task Force. https://www.ietf.org/rfc/rfc5246.txt

Rescorla E (2000) HTTP over TLS: RFC 2818. Internet Engineering Task Force. https://www.ietf.org/rfc/rfc2818.txt

Hub

Junhua Zhang
Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China
Key Laboratory of Random Complex Structures and Data Science, Chinese Academy of Sciences, Beijing, China
National Center for Mathematics and Interdisciplinary Sciences, Chinese Academy of Sciences, Beijing, China

Synonyms

Date hub; Party hub

Definition

Complex networks (including most biological networks) are known as scale-free networks and are characterized by a power-law degree distribution. This means that most nodes of the network have a lower degree, whereas a small percentage of nodes possess a large number of links in the network. These high-degree nodes are called hubs. In protein-protein interaction (PPI) networks hubs represent proteins with a large number of interactions, called hub proteins. In ▶ gene regulatory networks sometimes a single ▶ transcription factor can regulate many target genes; such a ▶ transcription factor is called a regulatory hub. Functional analysis showed that hubs were enriched in kinase and adaptor domains acting primarily in signal transduction (▶ Signal Transduction Pathway) and ▶ transcriptional regulation, whereas non-hubs had more DNA-binding domains and were involved in catalytic activity. Moreover, hub proteins were more likely to be essential than non-hub proteins.

Based on a computational analysis of yeast microarray data, Han et al. (2004) proposed two types of hubs, i.e., party hubs and date hubs. Date hubs display low ▶ co-expression with their partners, while party hubs have high co-expression. These two kinds of hubs were further discussed in Batada et al. (2006) and Agarwal et al. (2010).

References

Agarwal S, Deane CM, Porter MA, Jones NS. Revisiting date and party hubs: novel approaches to role assignment in protein interaction networks. PLoS Comput Biol. 2010;6:e1000817

Batada NN, Reguly T, Breitkreutz A, Boucher L, Breitkreutz BJ, et al. Stratus not altocumulus: a new view of the yeast protein interaction network. PLoS Biol. 2006;4:e317

Han JD, Bertin N, Hao T, Goldberg DS, Berriz GF, et al. Evidence for dynamically organized modularity in the yeast protein-protein interaction network. Nature. 2004;430:88–93

Human Clock

▶ Circadian Rhythm

Human Leukocyte Antigens (HLA)

▶ Major Histocompatibility Complex (MHC), Applications

Human Proteome Organisation

Sandra Orchard
EMBL Outstation, European Bioinformatics Institute, Hinxton, Cambridge, UK

Synonyms

HUPO

Definition

The Human Proteome Organisation (www.hupo.org) is an international scientific organization representing and promoting proteomics through international cooperation and collaborations by fostering the development of new technologies, techniques, and training. The organization organizes an annual congress and a number of initiatives, largely aimed at identifying the proteome content of a number of healthy and diseased tissues and cell types.

Human T-Lymphotropic Virus

▶ HTLV, Cellular Transcription

Human T-Lymphotropic Virus Type-I-associated Myelopathy/Tropical Spastic Paraparesis

▶ Human T-Lymphotropic Virus Type-I-associated Myelopathytropical Spastic Paraparesis

Human T-Lymphotropic Virus Type-I-associated Myelopathytropical Spastic Paraparesis

Christian Schönbach
Department of Bioscience and Bioinformatics, Kyushu Institute of Technology, Iizuka, Fukuoka, Japan

Synonyms

Human T-lymphotropic virus type-I-associated myelopathy/tropical spastic paraparesis; HAM/TSP

Definition

HAM/TSP is a slow-onset, chronic-progressive central nervous system disease that develops in less than 5% of HLTV-1 infected individuals and leads to corticospinal tract degeneration. In endemic areas, the onset of HAM/TSP is biased toward females at a 3:1 ratio and occurs between 40 and 50 years of age (Culcea and Sandbrink 2009). The proviral load and genetic background influence the disease onset and progression. Yet, the molecular mechanisms of onset and progression are not fully understood. Proinflammatory mediators found in lesions are associated with the infiltration of cytotoxic $CD8^+$ T-cell in the central nervous system. In addition, high proviral load and the presence of extracellular Tax1 have been implicated in the initiation of TNFA-mediated destruction of neuronal cells (Irish et al. 2009). The inflammation of the spinal cord causes spastic paraparesis of both legs, impaired position sense of the feet, lower lumbar pain, hyper-reflexia of upper extremities, and urinary incontinence (Culcea and Sandbrink 2009). The inflammation can trigger autoimmune conditions such as uveitis and myositis. Infections with HTLV-2 are also associated with HAM/TSP, but symptoms were reported to be milder and the progression slower. Ataxic HAM is only associated with HTLV-2 (Roucoux and Murphy 2004). Treatment of HAM/TSP with INFA2 and IFNB1 ameliorates symptoms and slows the progression of the disease during the course of the therapy. Effective HAM/TSP therapies or HTLV vaccines are not available.

Cross-References

▶ HTLV, Cellular Transcription

References

Culcea E, Sandbrink F (2009) Tropical Myeloneuropathies. eMedicine. http://emedicine.medscape.com/article/1166055-overview

Irish BP, Khan ZK, Jain P, Nonnemacher MR, Pirrone V, Rahman S, Rajagopalan N, Suchitra JB, Mostoller K, Wigdahl B (2009) Molecular mechanisms of neurodegenerative diseases induced by human retroviruses: a review. Am J Infect Dis 5(3):231–258

Roucoux DF, Murphy EL (2004) The epidemiology and disease outcomes of human T-lymphotropic virus type II. AIDS Rev 6(3):144–154

HUPO

▶ Human Proteome Organisation

Hybrid Simulation Strategies

Ruiqi Wang
Institute of Systems Biology, Shanghai University, Shanghai, China

Definition

Hybrid simulation strategies aim to combine different approaches into one calculation scheme. The essential idea is to partition reactions or species into two or more groups: e.g., a group of low-copy number species and a group of high-copy number species, and then treat them in different ways. For reactions, a system can be decomposed into two subsystems containing fast and slow reactions, respectively. Fast reactions often involve high-copy number species, e.g., in metabolism. Slow reactions or reactions involving low-copy number species can frequently be found in signal transduction or gene expression systems. The two subsystems are then simulated by using different methods,

e.g., exact methods and approximate algorithms, respectively.

In the slow/discrete subset of reactions, the fast subset evolves due to the action of the fast reactions, which means that, during that time, the behavior of fast subset of reactions can be approximated by using ordinary differential equations (ODEs), stochastic differential equations (SDEs), or approximate stochastic simulation methods, independent of the slow subset. Therefore, hybrid simulation strategies can be roughly divided into discrete/ODE method, maximal timestep algorithm, and discrete/Langevin method, depending on how the high-copy number reactions are approximated.

The discrete/ODE method begins with a partitioning of species and combining discrete event simulations with ODE models. This method approximates the high-copy number species with continuous deterministic techniques. The basic idea is that an ODE integration step will take place on the assumption that no discrete reaction takes place. If discrete reaction has occurred, the time of the event should be identified, the discrete updating should take place, and the continuous variables should be updated over this shorter timestep.

The hybrid methods also use a combination of an exact updating procedure for the low concentration species with various approximate simulation methods, e.g., τ-leap method, for the other species. There are various exact/approximate simulation methods, e.g., maximal timestep method proposed by Puchalka and Kierzek, which combines the next reaction method with τ-leap method.

Other methods also use a combination of stochastic simulation and numerical integration of SDEs. In these hybrid simulation methods, slow reactions are simulated based on discrete updating, while fast reactions are updated based on Langevin approach. An automatic partitioning and dynamic repartitioning of fast and slow reactions has been proposed by Salis and Kaznessis in 2005.

Actually, many different hybrid simulation strategies have been proposed. They integrate different partitioning policy and various exact/approximate simulation techniques. Hybrid simulation strategies are important because of the existence of very complex and heterogeneous models that integrate signaling, metabolism, and gene expression.

Cross-References

- ▶ Langevin Equation
- ▶ Law of Mass Action
- ▶ Master Equation
- ▶ Stochastic Simulation Algorithms

References

Alfonsi A, Cances E, Turinici G, Ventura BD, Huisinga W (2005) Exact simulation of hybrid stochastic and deterministic models for biochemical systems. ESAIM Proc 14:1–13

Chen L, Wang R, Li C, Aihara K (2010) Modeling biomolecular networks in cells: structures and dynamics. Springer, London

Puchalka J, Kierzek AM (2004) Bridging the gap between stochastic and deterministic regimes in the kinetic simulations of the biochemical reaction networks. Biophys J 86:1357–1372

Salis H, Kaznessis Y (2005) Accurate hybrid stochastic simulation of a system of coupled chemical or biochemical reactions. J Chem Phys 122:054103

van Kampen NG (1981) Stochastic processes in physics and chemistry. Elsiever, Amsterdam

Wilkinson DJ (2006) Stochastic modelling for systems biology. Chapman & Hall/CRC, Boca Raton

Hypergeometric Distribution

Yong Wang
Academy of Mathematics and Systems Sciences, Chinese Academy of Sciences, Beijing, China

Synonyms

Binomial distribution without replacement

Definition

In probability theory and statistics, the hypergeometric distribution is a discrete probability distribution that describes the number of successes in a sequence of a number of draws from a finite population without replacement. So in essence the hypergeometric distribution is the binomial distribution without replacement.

Suppose we have the following hypergeometric experiment:

- N: The number of items in the population
- N_2: The number of items in the population that are classified as successes
- N_1: The number of items in the sample

Then random variable X, the number of items in the sample that are classified as successes, follows a hypergeometric distribution:

$$P(X = i) = \frac{\binom{N_1}{i}\binom{N - N_1}{N_2 - i}}{\binom{N}{N_2}}$$

Here $\binom{N}{N_2}$ is the number of combinations of N things, taken N_2 at a time.

References

http://en.wikipedia.org/wiki/Hypergeometric_distribution

Hypergraph Theory

► Subset Surprisology and Toponomics

Hyperplasia

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

Definition

Hyperplasia is an excessive, orderly growth in which cells have a "normal increase in number," maintain uniformity and polarity, and stop growing after cessation of the stimulus that evoked the growth.

Cross-References

► Cancer Pathology

Hypertext and Hypermedia

Tim Clark
Department of Neurology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA

Definition

Hypertext is a nonlinear system of digital text organization, conceptually similar to footnotes, in which text contains pointers, called "hyperlinks," to other text segments, which may be accessed by acting (e.g., mouse clicking) on the pointers. These other text segments may be within the same document, or not.

Hypermedia is the multi-media generalization of hypertext. It forms the conceptual basis for navigational user experience on the ► World Wide Web.

Ted Nelson coined the terms "hypertext" and "hypermedia" and was an early advocate, developing the Xanadu system concept. Douglas Engelbart was one of the first to actually explore and demonstrate hypertext and hypermedia computer systems, resulting in his famous online multimedia "NLS" system demonstration of December 9, 1968, at the Stanford Research Institute. (Engelbart received the 1990 ACM Software Systems Award along with William K. English for NLS.) These and other researchers were initially inspired by the vision of Vannevar Bush in his July 1945 Atlantic Monthly article, "As We May Think" (Nelson 1965; Engelbart and English 1968; Conklin 1987).

Hypertext and Hypermedia are the subjects of significant ongoing research as topics in their own right, on which the Association for Computing Machinery (ACM) sponsors annual conferences (http://ht2010.org/, http://www.ht2011.org/, etc.). Many developments in distributed hypermedia that were originally omitted from the simplified initial design of the web, such as stand-off link services, are now being addressed and brought back into currency on the modern Web, through ► semantic web technologies.

Cross-References

► World Wide Web

References

Conklin J (1987) Hypertext: an introduction and survey. Computer 20(9):17–41

Engelbart DC, English WK (1968) A research center for augmenting human intellect. In: Proceedings of the December 9–11, 1968, fall joint computer conference, part I, 1476645. ACM, San Francisco, pp 395–410

Nelson TH (1965) Complex information processing: a file structure for the complex, the changing and the indeterminate. In: ACM '65, proceeding of the 1965 20th national conference. ACM Press, New York

Hypertext Markup Language

▶ HTML

Hypertext Transfer Protocol

▶ HTTP

Hypertext Transfer Protocol Secure

▶ HTTPS

Hypothesis Space

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

Definition

In machine learning, the goal of a supervised learning algorithm is to perform induction, i.e., to generalize a (finite) set of observations (the training data) into a general model of the domain. In this regard, the hypothesis space is defined as the set of candidate models considered by the algorithm.

More specifically, consider the problem of learning a mapping (model) $f \in F = Y^X$ from an input space X to an output space Y, given a set of training data $D = \{(x_1, y_1), ..., (x_n, y_n)\} \subset X \times Y$. A learning algorithm **A** takes D as an input and produces a function (model, hypothesis) $f \in H \subset F$ as an output, where H is the hypothesis space. This subset is determined by the formalism used to represent models (e.g., as logical formulas, linear functions, or non-linear functions implemented as artificial neural networks or ▶ decision trees). Thus, the choice of the hypothesis space produces a *representation bias*, which is part of the algorithm's ▶ *inductive bias*.

Cross-References

▶ Classification
▶ Decision Tree
▶ Inductive Bias

Hypothesis Testing

Roger Higdon
Seattle Children's Research Institute,
Seattle, WA, USA

Synonyms

Frequentist hypothesis testing; Parametric hypothesis testing

Definition

Conventional statistical hypothesis testing validates whether a hypothesis about a quantity of interest (a parameter) is true or false based on the likelihood that the observed data could have been generated if the hypothesis were true.

Characteristics

Statistical hypothesis testing has a long history (Fisher 1925 and Neyman and Pearson 1933). Hypothesis testing begins with a question about a parameter or parameters of interest that describe aspects of a population of interest. The parameters define the probability distribution for generating data from the population. In order to test a hypothesis about a parameter, a sample of

Hypothesis Testing, Table 1 Outcomes of a statistical hypothesis test

	Fail to reject H_0	Reject H_0
H_0 true	Correct	Type I error
H_0 false	Type II error	Correct

data is taken from the population and a test statistic is calculated from the data (Lehmann and Romano 2005). Below are a number of definitions associated with hypothesis testing.

- Simple hypothesis – A hypothesis based on a single parameter value (i.e., $\theta = 0$)
- Composite hypothesis – A hypothesis based on a range of parameter values (i.e., $\theta > 0$)
- Null hypothesis (H_0) – A simple hypothesis associated with the theory one would like to disprove (status quo)
- Alternate hypothesis (H_A) – A hypothesis (often composite) associated with a theory one would like to prove
- Test statistic – A function of the data with which decisions about the hypothesis are made
- Decision rule – Values of the test statistic that lead to rejecting or failing to reject the null hypothesis
- Acceptance region – The set of values for the test statistic which we fail to reject the null hypothesis
- Rejection region – The set of values for the test statistic which we reject the null hypothesis
- Critical value(s) – The value(s) on the border of the acceptance and rejection regions

In a statistical hypothesis test, either H_0 or H_A will be true, and either H_0 will be rejected or we will fail to reject it. This leads to the four possible outcomes of a statistical hypothesis test shown in Table 1.

These outcomes lead to two types of errors and the probability of those errors informs the decision rule defined by a statistical hypothesis test. Below are the definitions of these errors and probabilities.

- Type I error – Wrongly rejecting H_0
- Type II error – Wrongly not rejecting H_0
- Level of significance/size of the test (α) – The probability of a type I error
- Power – one minus the probability of a type II error for a given parameter value in Ha ($1-\beta$)
- p-value – smallest alpha value for which H_0 is rejected

Given the above definitions, the basic steps of a statistical hypothesis test are as follows:

1. The first step is to describe the null and alternative hypotheses that relate to the research objective. This is important so that the results of the hypothesis test are relevant to the research objective. Specifically, the null hypothesis needs to be defined in such a way that its rejection allows for the conclusion that research objective has been achieved. For example, rejecting the null hypothesis that the difference in means between a treatment and control is 0 allows for the conclusion that the treatment had an effect.
2. The second step is to consider the statistical assumptions being made about the sample in doing the test. Is the data continuous or discrete? Are the data independent? Are they paired? Collected over time? Is the sample random? This will help determine which methods to use and whether results are valid or generalizable.
3. Determine the appropriate test statistic given the assumptions.
4. Derive the distribution of the test statistic under the null hypothesis based on the assumptions. In most cases, there exists a test statistic with a known distribution, such as the two-sample t-test, ANOVA F-tests, chi-squared tests for independence, t-tests for regression parameters, and so on.
5. The distribution of the test statistic and the level of significance create acceptance and rejection regions.
6. Compute the value of the test statistic.
7. Decide to either fail to reject the null hypothesis or reject it in favor of the alternative hypothesis.

One problem with this approach is that it leads to an absolute reject or does not reject outcome and fails to separate uncertain conclusions from definitive ones. It also leads to arbitrary thresholds of significance such as 0.05. Calculating a p-value improves upon this by providing a quantitative scale rather than absolute decision. P-values near 0 provide definitive statistical proof against H_0, while values near alpha imply evidence against H_0 but are uncertain. Finally p-values near 1 would give no reason to believe the null hypothesis was not true. Also calculating p-values is simple since they are based on the distribution of test statistic under the null hypothesis. All this still assumes that people properly interpret p-values rather than use them as another way to threshold and that they are calculated in a valid manner (Jones and Tukey 2000).

Another issue that is not addressed by this approach is practical significance versus statistical significance. If the power of a test is very high, then often statistical significance is achieved without any practical significance. If the power of the test is low, then statistical significance may be difficult to achieve no matter how large the observed test statistic is. Power for a specific test versus a specific simple alternative hypothesis is generally controlled by the sample size; so it is important to choose a large enough sample size to achieve statistical significance for practically significant alternative hypotheses, but not so large that it yields statistical significance for uninteresting results thereby wasting resources.

Other issues to consider in hypothesis testing, particularly with respect to systems biology, are the impact of multiple hypothesis testing and the validity of results when assumptions are violated. If violating assumptions is a concern, then use of nonparametric hypothesis testing should be considered.

Cross-References

▶ Hypothesis Testing, Parametric vs Nonparametric
▶ Multiple Hypothesis Testing

References

Fisher RA (1925) Statistical methods for research workers. Oliver and Boyd, Edinburgh, p 43
Jones LV, Tukey JW (2000) A sensible formulation of the significance test. Psychol Methods 5(4):411–414
Lehmann EL, Romano JP (2005) Testing statistical hypotheses, 3Eth edn. Springer, New York
Neyman J, Pearson ES (1933) On the problem of the most efficient tests of statistical hypotheses. Philos Trans R Soc Ser A 231:289–337

Hypothesis Testing, Bayesian vs Frequentist

Roger Higdon
Seattle Children's Research Institute,
Seattle, WA, USA

Definition

Bayesian hypothesis testing, similar to Bayesian inference and in contrast to frequentist hypothesis testing, is about comparing the prior knowledge about research hypothesis to posterior knowledge about the hypothesis rather than accepting or rejecting a very specific hypothesis based on the experimental data.

Characteristics

Just as with Bayesian inference, Bayesian hypothesis testing is based on posterior distribution (Box and Tiao 1973):

$$p(\theta|X) = p(X|\theta)p(\theta)$$

A p-value-like quantity can be generated for a one-side hypothesis like $\theta < 0$ by integrating the posterior:

$$p = \int_{-\infty}^{0} p(\theta|X)$$

However, for a two-sided hypothesis testing $\theta = 0$ it is not feasible since the integral for a continuous posterior would be 0. An ad hoc solution would be to integrate in a region around 0 that would represent a region of no practical significant difference:

$$p = \int_{-a}^{a} p(\theta|X)$$

An alternative to direct integration of the posterior is the use of Bayes factors (Kass and Raftery 1995). The Bayes factor is the ratio of posterior odds of the null hypothesis ($\theta = \theta_0$) versus an alternative ($\theta = \theta_1$) to the prior odds, where the posterior odds is the ratio of posterior distributions for the two hypotheses and the prior odds is likewise defined.

The Bayes factor is the analog to the frequentist likelihood ratio test. In fact, if the prior odds is equal to 1, then they are equivalent. Large Bayes factors would favor the null hypothesis while small Bayes factors would favor the alternative. Bayes factors do not explicitly make probability statements and so tend to rely on rules of thumb for significance. Also, they can be difficult to calculate.

More formal comparisons of models can be carried out using the Bayesian Decision Theory, where cost functions are added to the Bayesian model (Berger 1985).

Cross-References

- ▶ Bayesian Inference
- ▶ Hypothesis Testing

References

Berger JO (1985) Statistical decision theory and Bayesian analysis, 2nd edn. Springer, New York

Box GEP, Tiao GC (1973) Bayesian inference in statistical analysis. Wiley, New York

Kass RE, Raftery AE (1995) Bayes factors. J Am Stat Assoc 90(430):791

Hypothesis Testing, Parametric vs Nonparametric

Roger Higdon
Seattle Children's Research Institute,
Seattle, WA, USA

Synonyms

Distribution-free tests; Nonparametric tests; Rank tests

Definition

Nonparametric hypothesis testing, in contrast to parametric hypothesis testing, does not rely on assumptions that data come from a particular parameterized distribution such as a normal distribution.

Characteristics

Nonparametric or distribution-free tests are widely used for statistical hypothesis testing, particularly when there is doubt as to whether data can be easily modeled by standard probability distributions (Gibbons and Chakraborti 2003 and Wasserman 2007). Most commonly this occurs when there is doubt about the data having a normal distribution. The most commonly used nonparametric tests are based upon substituting ranks for the original data. Table 1 gives nonparametric analogs to many commonly used statistical hypothesis tests.

Hypothesis Testing, Parametric vs Nonparametric, Table 1 Statistical tests and their nonparametric analogs

Parametric test	Nonparametric test
One sample T-test	Sign-rank test
Two-sample T-test	Rank-sum or Mann-Whitney test
One-way ANOVA	Kruskal-Wallis test
One-way randomize complete block ANOVA	Friedman test
Pearson correlation	Spearman rank correlation or Kendall's tau

There are many other tests such as the Kolmogorov-Smirnov test for comparing two distributions or the Wald-Wolfowitz runs test for randomness.

An alternative to rank tests is to apply randomization (Edgington 1995), permutation, or other resampling approaches such as the bootstrap (Efron 1982) to conventional test statistics. These approaches utilize the random assignment of the data to treatment groups or explanatory variables in order to estimate p-values. Randomization tests are based on calculating all possible randomizations of the data without replacement. A permutation test approximates a randomization test by taking random permutations of the data with replacement. Bootstrapping resamples the original data with replacement. Randomization tests are commonly used in systems biology due to the doubt about parametric assumptions in biological data. Some common examples are the significance analysis of microarrays tests in relative expression analysis (Tusher 2001) and gene set enrichment tests.

Nonparametric tests offer a number of advantages including protection from the violation of distributional assumptions, increased power when assumptions are violated, and the ability to preserve complex correlation structures. Disadvantages include decreased power when distributional

assumptions can be met, requirements for large sample sizes, and difficulty in applying tests to complex statistical models.

Cross-References

- ▶ Hypothesis Testing
- ▶ Relative Expression Analysis

References

Edgington ES (1995) Randomization tests, 3rd edn. Marcel-Dekker, New York
Efron B (1982) The jackknife, the bootstrap, and other resampling plans. Soc of Ind and Appl Math CBMS-NSF Monogr 38
Gibbons JD, Chakraborti S (2003) Nonparametric statistical inference, 4th edn. CRC Press, Boca Raton
Tusher V, Tibshirani R, Chu G (2001) Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci 98:5116–5121
Wasserman L (2007) All of nonparametric statistics. Springer, New York

Hypoxia

Marsha A. Moses
Department of Surgery/Harvard Medical School, Vascular Biology Program/Children's Hospital Boston, Boston, MA, USA

Definition

Hypoxia refers to the reduced oxygen tension caused by low oxygen availability or insufficient oxygen delivery within an organism. Under hypoxic conditions, the organism must exert a series of adaptive responses to ensure cellular survival, such as changes in metabolism, regulation of pH, increased red blood cell production and oxygen transport, and initiation of angiogenesis (Cassavaugh and Lounsbury 2011). These processes are mediated by dimeric protein complexes, the hypoxia-inducible factors, which regulate the transcription of many genes that are critical for these cellular responses (Semenza 2009). Hypoxia-induced cellular signaling is essential during embryonic development for the formation of multiple tissues including the central nervous system and the vasculature. Hypoxia also occurs in diseases such as cardiac ischemia and cancer; therefore, targeting signaling pathways initiated by hypoxia represents a promising therapeutic strategy for these diseases (Cassavaugh and Lounsbury 2011).

Cross-References

- ▶ Regulation of Tumor Angiogenesis

References

Cassavaugh J, Lounsbury KM (2011) Hypoxia-mediated biological control. J Cell Biochem 112(3):735–744
Semenza GL (2009) Regulation of oxygen homeostasis by hypoxia-inducible factor 1. Physiology (Bethesda) 24: 97–106

Hypoxia-inducible Factor-1

Marsha A. Moses
Department of Surgery/Harvard Medical School, Vascular Biology Program/Children's Hospital Boston, Boston, MA, USA

Definition

Hypoxia-inducible factor-1 (HIF-1) is a dimeric protein complex comprises of two subunits, HIF-1α and HIF-1β, both of which are members of the bHLH-PAS family of transcription factors. It regulates the transcription of many genes whose functions are critical for the survival of cells under hypoxic conditions. Under normoxic conditions, HIF-1α is hydroxylated by prolyl hydroxylase domain–containing (PHD) proteins loaded with oxygen as cofactors. The hydroxylated HIF-1α then interacts with the von Hippel-Lindau tumor suppressor protein (VHL), a member of the E3 ubiquitination complex, and is subjected to protein degradation by the 26s proteasome. Under hypoxic conditions, the activity of PHD is decreased due to low availability of oxygen in the cells. Therefore, HIF-1α is relieved from protein degradation, translocates to the nucleus, and forms a dimer with HIF-1β.

There, the HIF-1 complex binds to the hypoxia response element (HRE) within the promoters of its target genes and activates their gene transcription (Semenza 2007). Due to its ability to upregulate the transcription of proangiogenic factors such as vascular endothelial growth factor (VEGF), HIF-1 is one of the key mediators of hypoxia-induced angiogenesis (Levy et al. 1997).

Cross-References

► Regulation of Tumor Angiogenesis

References

Levy AP, Levy NS, Iliopoulos O, Jiang C, Kaplin WG Jr, Goldberg MA (1997) Regulation of vascular endothelial growth factor by hypoxia and its modulation by the von Hippel-Lindau tumor suppressor gene. Kidney Int 51(2):575–578

Semenza GL (2007) Hypoxia-inducible factor 1 (HIF-1) pathway. Sci STKE 407:cm8

Hysteresis

► Cell Cycle Dynamics, Irreversibility