
3 Systems Biology Approaches to Understanding and Predicting Fungal Virulence

LANAY TIERNEY¹, KATARZYNA TYC², EDDA KLIPP², KARL KUCHLER¹

CONTENTS

I. Introduction: Fungi as Model Organisms for SysBio Studies	45
II. High-Throughput and -Omics-Based Methods for Studying Fungal Virulence	48
A. Genomics	48
B. Transcriptomics	49
1. Clustering Gene Expression Data Sets ...	51
C. Proteomics	52
E. Metabolomics	55
F. Epigenomics	56
G. Data Mining Approaches and Genome-Wide Fungal Resources	57
1. Databases	57
2. Strain Collections	57
III. Modeling Biological Phenomena	58
A. Boolean Models	59
B. Petri Nets	61
C. Ordinary Differential Equation Models ...	62
D. Flux-Balance Analysis	63
E. Stochastic Modeling	64
F. Monte Carlo Simulation	64
G. Agent-Based Models	65
H. Game Theory	65
I. Model Parameters	66
J. Sensitivity Analysis	67
K. Standards for Modeling	67
IV. Conclusions and Future Perspectives	68
References	68

I. Introduction: Fungi as Model Organisms for SysBio Studies

Systems Biology (SysBio) is an overarching term used to describe a collection of methods that attempt to understand biological systems in a quantitative fashion, primarily using experimental data to formulate mathematical models that enable the prediction of the future behavior of a system or emergent properties in response to environmental stimuli. A “system” in this context can be defined as any group of interacting biological components that either directly or indirectly perform a specific and measurable function; it can include a small or large number of components. For example, a protein complex, a regulatory pathway, an organ, a cell or the entire organism could all be systems in different experimental contexts. This field is a recent addition to the scientific community. Although its exact origins are difficult to pinpoint, many ascribe its beginnings to the late 1960s when the first attempts to investigate metabolic pathways in bacterial cells were performed (von Bertalanffy 1969). The formulation of the metabolic control theory (MCT) by the groups of Kacser and Burns (1973) and Heinrich and Rapoport (1974) marks the birth of SysBio. MCT mathematically describes metabolic flux as an inherent system property, and that the flux control through different reactions is shared within the system, influencing one another.

“Scientific fields, like species, arise by descent with modification” (Kirschner 2005), and tellingly, to date, a consensus definition of what SysBio concretely means is intensely debated. Nevertheless, SysBio has some key

Lanay Tierney and Katarzyna Tyc contributed equally.

¹Christian Doppler Laboratory Infection Biology, Max F. Perutz Laboratories, Medical University of Vienna, 1030 Vienna, Austria; e-mail: karl.kuchler@meduniwien.ac.at

²Institute of Biology, Theoretical Biophysics, Humboldt-Universität zu Berlin, Invalidenstrasse 42, 10115 Berlin, Germany; e-mail: edda.klipp@biologie.hu-berlin.de

characteristics that differentiate it from classical reductionist-driven biological studies. Historically, scientific experiments have taken a reductionist approach, investigating a single gene under a single condition. However, biological systems are dynamic and nonlinear in nature, making the reductionist approach often unsuitable for investigating the behavior of a gene of interest. A **primary goal of SysBio is to understand the wiring and connectivity in biological networks that define genotype–phenotype relations** (Kohl et al. 2010). A “systems approach,” sometimes referred to as the “desk–bench–desk loop” (Mustacchi et al. 2006), combines computational and experimental tools to formulate a model based on prior knowledge of the system (i.e., from published literature, databases, and genome-wide data sets). This model is then tested under laboratory conditions and analyzed for biological insights that would have been otherwise difficult to gain without an interdisciplinary approach (Fig. 3.1). The recent introduction of whole-genome sequencing and genome-wide experimental platforms such as RNA sequencing (RNA-seq), metabolomics, microarray, and mass spectrometry (MS), among several others, form the core “omics-based” methodologies for data generation that drive SysBio studies.

The integration of data from these various technologies into SysBio models has remained a formidable challenge. Hence, SysBio approaches have been classified as “top-down,” “bottom-up,” or “middle-out” (Bray 2003; O’Malley and Dupre 2005; Bruggeman and Westerhoff 2007; Petranovic and Nielsen 2008). A **top-down** approach aims at extracting principles from experimental data representing molecular properties of the studied system. A top-down approach focuses on the comparison of genome-wide data sets, such as transcriptome and proteome, to formulate a focused and testable biological hypothesis. A **bottom-up** approach uses the knowledge of molecular properties of the system components to predict the behavior of the system as a whole. In short, the bottom-up approach connects smaller entities to predict, identify, and simulate the behavior of a bigger system. A bottom-up approach usually starts with prior knowledge of a specific gene, and a model is then generated based on these data to investigate the system as

a whole. A **middle-out** approach describes a process of starting at the level for which the best information for the process of interest is available, and then combining higher and lower levels of structural and functional information, essentially breaking out of a more strict top-down and bottom-up loop in order to validate the hypothesis at the current state of biological understanding (Brenner et al. 2001). Regardless of the approach used, SysBio strategies differentiate themselves from more classical biological methods by consciously taking into consideration different levels and dynamics of biological data (DNA, RNA, protein) simultaneously.

Fungal SysBio was ushered in with the completion of the *Saccharomyces cerevisiae* genome sequence, the first completely sequenced eukaryotic genome (Goffeau et al. 1996). Since then, the genomes of many of the most common fungal pathogens including, but not limited to, *Candida albicans*, *Aspergillus fumigatus*, and *Cryptococcus neoformans*, have become available (Loftus et al. 2005; Nierman et al. 2005; van het Hoog et al. 2007). Community-wide initiatives such as the Fungal Genome Initiative (Cuomo and Birren 2010) and the 1,000 Fungal Genomes Project (<http://1000.fungalgenomes.org/>) have been useful tools for studying the evolution of fungal virulence. The discovery of key genes positively and negatively regulated during the infection process, and understanding the function of their products, will drive the design of new strategies to combat fungal pathogens.

In this chapter, we provide a comprehensive overview of recent SysBio methods suitable for study of fungal virulence, including genome sequencing, -omics technologies, and bioinformatics tools, with an emphasis on computational and modeling-based approaches. We focus on the genera *Candida* and *Saccharomyces*; the latter stands out as a “workhorse” of fungal SysBio in which many of the methods described herein were originally established or tested (Mustacchi et al. 2006; Santamaria et al. 2011). These approaches are used to identify molecular wiring and dynamics in biological networks, with the goal of identifying their biological function and eventually identifying novel therapeutic options. We describe the

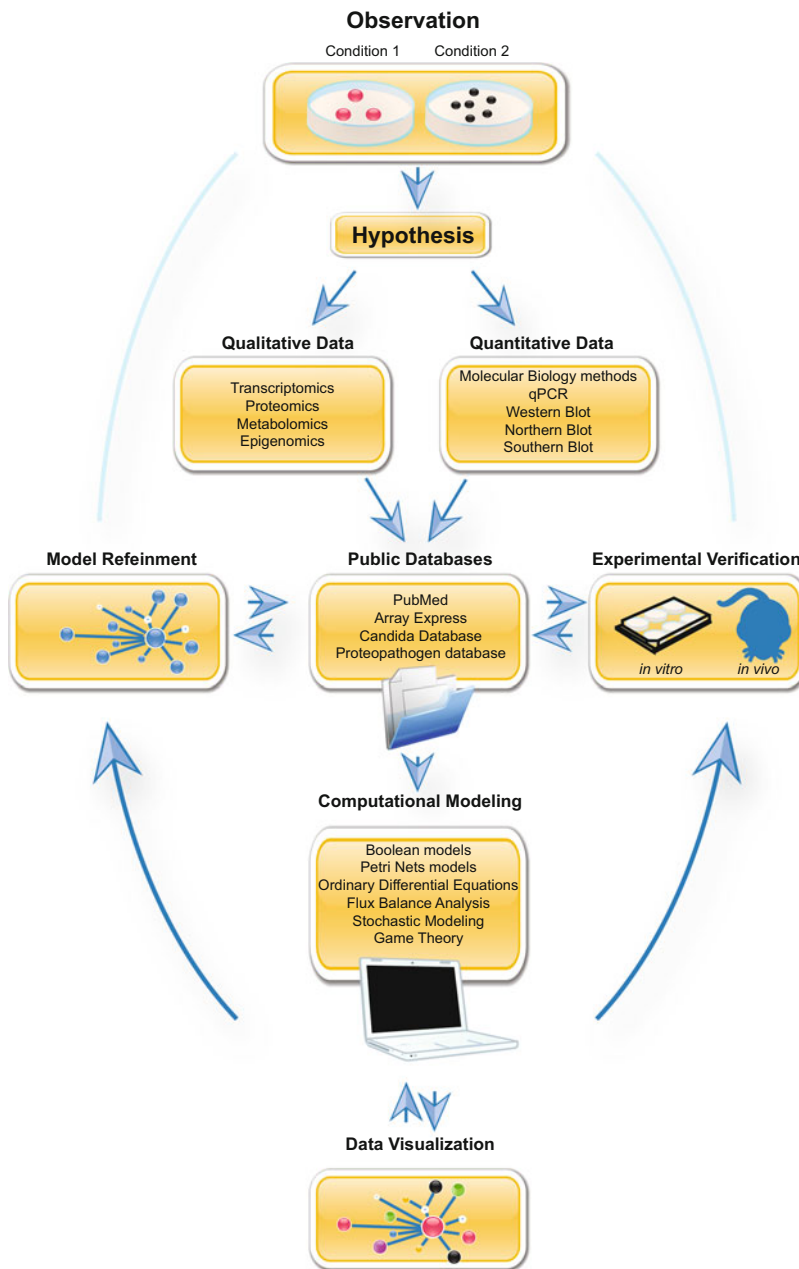


Fig. 3.1. The “bench–desk–bench” loop of SysBio. A typical SysBio workflow. Based on an observation, such as a difference in yeast colony size and phenotype under two different conditions, a hypothesis is formed based on the experimental results and further supported by prior knowledge in published literature. This hypothesis can be addressed using a number of qualitative and quantitative methods, the results of which are deposited

in publically available databases. With these data, modeling approaches attempt to mimic, predict, and visualize data. Once modeled, experimental verification and refinement of the model, creates the bench–desk–bench loop, where iterative cycles of prediction and verification are undertaken until the model and experiment validation are representative of one another

applicability of each method to specific experimental questions using numerous case examples and critically discuss some of the current pitfalls in the analysis of SysBio data sets.

II. High-Throughput and –Omics-Based Methods for Studying Fungal Virulence

SysBio approaches, especially top-down analyses, incorporate genome-wide data sets such as comparative genomics, transcriptomics, and proteomics data. These approaches fit into the category of “–omics” or “–ome” studies, which attempt to analyze a genome-wide response to a specific condition. –Omics studies represent an important shift in the way biological data is both produced and interpreted, complementing traditional hypothesis-driven research (Weinstein 2001). In order to understand SysBio as a whole, it is important to understand the types of data sets that it utilizes to address a given question. Several important methods have established themselves in this field over the past decade and have been used extensively to investigate fungal virulence. For clarity, we have divided these methods into qualitative and quantitative approaches. We address some of the most popular methods used at different levels of biological understanding, including DNA, RNA, and protein, as well as epigenetic modifications and validation methods, and examine the key contributions they have made to the understanding of fungal virulence.

A. Genomics

The initial genomic sequencing of *S. cerevisiae* was a monumental international collaboration that included some 600 scientists worldwide (Goffeau et al. 1996).

This sequencing was performed using a series of hybrid plasmids, called “cosmids.” Cosmids had the advantage that a much longer DNA sequence stretches could be incorporated than using normal plasmids, and at the same time longer DNA stretches could be sequenced to build up the genomic library. Sequencing polymerase

chain reaction (PCR) fragments then filled the remaining gaps between sequence stretches of the assembled genomic library to complete the genome (Dujon 1993). The *Candida* Genome Sequencing project began directly after the *S. cerevisiae* sequencing in 1996, ending in 2004 with the *C. albicans* genome assembly known as Assembly 19 (Jones et al. 2004). This genome assembly was divided into 412 contigs (consensus stretches of DNA that are assembled to form the scaffold of the genome assembly) and sequenced with a shotgun-based sequencing strategy. In order to obtain a more complete view of the diploid sequence, Assembly 21 was created using a fosmid library, which is conceptually similar to a cosmid library, except that it is based instead on a bacterial F-plasmid and is more stable than a cosmid because of its low copy number (Hall 2004). These early sequencing projects took years because of the low throughput.

Today, genome and transcriptome sequencing has become routine, with ever-increasing stability, coverage (several fungal genomes can now be sequenced in a couple of days), and bioinformatics assembly tools publically available. As DNA sequencing technologies have become more efficient, there has been a surge in the number of sequenced genomes, with over 150 fungi sequenced so far (Marcet-Houben and Gabaldon 2009). These sequences facilitate functional and comparative genomics studies. **Functional genomics** aims to understand relationships between genotype and phenotype. **Comparative genomics** attempts to identify genes or genetic rearrangements between closely related species based on their DNA sequence; in the case of fungal pathogens, this often includes a highly virulent species compared to a significantly less or even avirulent species. This is done in order to identify genetic transitions that might explain the evolutionary divergence of pathogens or the identification of novel virulence factors.

Comparative genomics studies use two main techniques: comparative whole genome sequencing or hybridization-based microarrays. Comparative whole genome sequencing literally attempts to identify genetic elements present in one species and absent in another based on the genome sequence; this is done by overlapping the genome sequences and identifying outlier sequence stretches that do not match between them. Comparative genomic hybridization (CGH) arrays identify genome-wide variation in gene copy number. CGH experiments assume that the binding

ratio of the experimental sample to the control is proportional to the sequencing concentration in the samples. These methods provided significant insight into the evolution of pathogenicity for many fungal species. For example, early comparative studies of fungi identified a strong sequence homolog among 228 genes in *S. cerevisiae*, *Schizosaccharomyces pombe*, *Aspergillus niger*, *Magnaporthe grisea*, *C. albicans* and *Neurospora crassa* genomes for which no homology was found in the human or mouse genomes, representing potential targets for pan-fungal treatment (Braun et al. 2005).

Numerous studies have investigated the evolution of pathogenicity within a single fungal clade. For example, in the *Candida* clade, eight genomes were sequenced, including the *C. albicans* WO-1 strain (which is characterized for white-opaque switching and is associated with specificity to host tissues), along with the de novo sequencing of *C. tropicalis* and *C. parapsilosis*, *Lodderomyces elongisporus*, *C. guilliermondii* and *C. lusitaniae*, many of which are now classified as emerging fungal pathogens (Butler et al. 2009). These strains were compared to the previously sequenced genomes of *C. albicans* clinical isolate SC5314, the marine yeast *Debaryomyces hansenii*, and species from the *Saccharomyces* clade. Some 21 gene families emerged that were enriched in pathogenic species as compared to nonpathogenic fungi. A related study investigated the closest known relative of *C. albicans*, *C. dubliniensis*, which, despite its similarities, is significantly less virulent than *C. albicans*. Comparative sequence analysis has identified almost 200 species-specific genes in *C. albicans*, including the absence of the key *C. albicans* invasion gene *ALS3* in *C. dubliniensis*, and members of the aspartyl proteinase family *SAP4* and *SAP5* (Jackson et al. 2009). *ALS3* is among the most important virulence factors in *C. albicans*. It is a cell surface protein that plays a major role in adhesion to host cells and in maintenance of infection (Hoyer 2001; Hoyer et al. 2008). Notably, numerous translocations were identified in *C. dubliniensis*, especially in the *SAP* family, which is known to play a role in *Candida* pathogenesis. Comparative genomics has even lent itself to the investigation of genetic variations at chromosome level, using a single *C. albicans* isolate that had been passaged multiple times

in an in vivo model organism using CGH (Forche et al. 2009), showing the environmental impact on the host strain evolution. Together, these studies collectively demonstrate that even closely related species have significantly diverged at their genomic levels, suggesting mechanisms for the evolution of fungal virulence factors.

A number of resources for fungal genomics research have recently been made available. Large genome databases, including the Broad Institute (<http://www.broad.mit.edu/annotation/cgi/>), the Sanger Center (<http://www.sanger.ac.uk/Projects/Fungi/>), the Institute for Genomic Research (<http://www.tigr.org/tdb/fungal/>), and the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/genomes/FUNGI/funtab.html>), are all publically available.

B. Transcriptomics

Looking at the DNA level, one can investigate how the information stored in the genetic code is translated into protein molecules; however, it does not provide information on the diverse molecules actually produced in the cell normally or in response to different environmental conditions. This takes place at the RNA level by looking at the complete set of RNA species produced in a given population of cells at a specific time. This field is referred to as **transcriptomics**.

There are two main methods to investigate transcriptional dynamics, also known as expression profiling. These include genome-wide **microarrays** and, more recently, **next-generation sequencing** (NGS) technologies, most notably, RNA-seq. Microarray experiments employ special microarray chips carrying printed copies of the entire genome and are used for assessing the relative differences in gene expression between a control sample and a treated sample. A microarray chip is, in principle, able to measure the relative changes in expression levels of all known genes simultaneously. Southern blotting was the inspiration for the microarray technology (Maskos and Southern 1992). Southern blotting involves the

hybridization of a DNA probe to a specific DNA fragment on a solid substrate. Microarrays use the same principle but cover a genome-wide scale rather than single genes. The chip itself is made up of probes, corresponding to short DNA stretches for all genes in the genome. Depending on the specific experimental question, there may be multiple probes for each gene of interest and the number of spotted probes is often well into the thousands. In a typical microarray experiment, RNA is collected from an organism of interest, transcribed into cDNA (the sample is also referred to as the “target”), and hybridized to the chip where the target forms hydrogen bonds with the probes. In order to determine the relative abundance of the transcripts, the chip is then scanned with the hybridized sample. In theory, if a gene is expressed in the organism, it will have hybridized to the probe on the chip. The abundance of a gene product is then measured by the detection of chemiluminescent-labeled targets. Based on the intensity of the target–probe hybridization, the relative abundance of the RNA produced by the organism in response to a condition can be measured. Because microarray technology is chip-based, its ability to detect a specific gene or transcript is limited to the original spotting on the chip itself. This is especially important to keep in mind for certain organisms, where only incomplete genomic sequences are available, leading to low quality annotations and unknown alternative spliced products, which would remain undetected if not already taken into account in the original design of the microarray.

The first genome-wide array was developed for *S. cerevisiae* in 1997 (Lashkari et al. 1997). Since then, microarrays for fungi have evolved into high-density tiling arrays (Sellam et al. 2010) and splicing-sensitive exon-junction arrays (Inada and Pleiss 2010), among many others. Microarray technology has been extensively used to investigate global changes in gene expression in response to changing environmental conditions and genetic knockouts. It has also been used in conjunction with immune cell and animal models at different infection stages in vitro and in vivo to investigate different infection stages. For example, the transcriptional response of both *S. cerevisiae* and *Candida glabrata* to antifungal agents and other chemical stress agents in vitro was profiled (Lelandais et al. 2008). To identify

pathogen-specific responses on the side of *C. glabrata*, the authors compared the transcriptional profile of both species after treatment. Surprisingly, they found a high conservation among the regulated genes during infection, and a subpopulation of genes that were pathogen-specific. Further, in vitro infection experiments of human blood cells with *C. albicans* identified a number of differentially expressed genes, which may be important in the survival of *Candida* during bloodstream infections (Fradin et al. 2003). Transcriptional profiling has been used to identify effects of phagocytosis of *C. albicans* by immune cells, including neutrophils (Rubin-Bejerano et al. 2003) and macrophages (Lorenz et al. 2004). These studies identified the extent of the amino-acid-deficient environment within the phagosome, and characterized the dynamic starvation response of *Candida* over the time course of infection. The first dual transcriptional profiling using microarrays for a host and pathogen interaction was also performed with conidia of *A. fumigatus* during infection of human airway epithelial cells. This work confirmed the upregulation of inflammatory interleukin (IL)-6 and the immune response to conidia, as well as pathways whose activation had previously only been investigated from either the host or pathogen perspective alone (Oosthuizen et al. 2011).

Transcription profiling using **RNA-seq** is conceptually similar to microarray, insofar as the end result of the experiment is often a list of differently expressed genes. However, the sample is sequenced using a parallel sequencing approach referred to as next-generation sequencing (NGS) instead of using hybridization-based methods. Based on Sanger sequencing methods, high-throughput technology began with tag-based methods that were developed so that multiple sequencing reactions could be run in parallel. These included serial analysis of gene expression (SAGE) (Velculescu et al. 1995), cap analysis of gene expression (CAGE) (Kodzius et al. 2006) and massive parallel signature sequencing (MPSS) (Reinartz et al. 2002). In order to increase the scale of reactions taking place, a number of novel sequencing strategies and commercially available platforms have been developed. These included Roche/454, Illumina/Solexa, Life/APG Helicos BioSciences, and Pacific Biosciences. Each system has pros and cons, depending on the biological application (Metzker 2010).

In a typical RNA-seq experiment, cDNA is first fragmented; these templates are then attached to a substrate

(which will vary with the technology used) with the aid of adaptor sequences. The immobilization of the template samples gives the advantage of allowing billions of simultaneous sequencing reactions, differentiating itself from first generation sequencing technology in terms of capacity and cost (Metzker 2010). Templates can be sequenced either from one end (single-end sequencing) or both ends (paired-end sequencing). The resulting sequencing reads can vary in length, depending on the technology used, from less than 30 bp to over 300 bp (Wang et al. 2009; Metzker 2010). Reads are then mapped back to the reference genome to determine gene expression and, when compared to other samples, differential gene expression.

RNA-seq has rapidly gained in broad popularity over the past few years, especially because of its ability to sequence to a high depth and also because it detects low abundance transcripts, offering a more complete view of the transcriptional profile of an organism than microarrays.

The sequencing technology has significant advantages over microarray, especially for non-model organism species, **as the detection of expressed genes is not dependent on having a priori knowledge of the gene investigated.** Moreover, RNA-seq does not have intrinsic limitations to the dynamic range of detection (Royce et al. 2007). RNA-seq has been especially important in the detection of novel noncoding RNA species and small RNAs, as well as for de novo annotation (Wang et al. 2009). Under in vitro conditions, a de novo annotation of the *C. albicans* transcriptome under nine different environmental conditions was recently performed and was able to identify over 600 novel transcriptionally active regions and introns from a total of 177 million uniquely mapped reads (Bruno et al. 2010). Similarly, with *A. fumigatus*, RNA-seq was used to investigate planktonic and biofilm growth to identify differences in pathological and morphological characteristics in these two stages. Numerous biofilm-specific genes were identified as being regulated, representing targets for biofilm development.

Most recently, the first **dual-species RNA-seq** approach, sequencing RNA mixture comprising both host and fungal pathogen transcriptomes over a time course of infection, has been accomplished. Furthermore, this study predicted, using mathematical approaches, and experimentally

verified novel host–pathogen regulatory networks implicated in the interaction. The use of **a combination of sequence analysis and network inference enabled this dual-systems approach** (Tierney et al. 2012). This study presents the first adaptation of network inference to model host–pathogen interactions, validating the use of network inference for the analysis of multiple species data sets.

1. Clustering Gene Expression Data Sets

The most common output of transcriptomics is a list of differentially expressed genes in one condition versus a control condition. Differential gene expression analysis begins with a testing for the statistical significance of the variation within the sample. Statistical approaches for determining differential expression have been extensively reviewed elsewhere (Cui and Churchill 2003), as have a number of freely available tools to aid in statistical analysis for both microarray (Steinhoff and Vingron 2006) and RNA-seq data sets (Sun and Zhu 2012). This method reduces a genome-wide comparison down to only those genes significantly affected under a specific condition. Convenient analysis pipelines, especially for RNA-seq data, have been recently created to help non-computational biologists in the analysis of sequencing from the raw data file to a list of differentially expressed genes (Oshlack et al. 2010; Garber et al. 2011). High-dimensionality data such as microarray or RNA-seq samples complicate data analysis due to the inequality of variables measured compared to the sample number. Because the list of differentially expressed genes can still be on an order of magnitude of several hundred genes, **additional methods to reduce complexity** are often necessary.

Partitioning expression data into subgroups of genes, called clusters, facilitates data visualization and interpretation underlying a biological process of interest. Depending on the approach used, the groupings can then be visualized by scatter plots, histograms, dendrograms, or heat maps. Genes are clustered into specific categories, which can be functional,

structural, temporal, or a combination of the above. A number of clustering approaches have been developed, including principle component analysis (PCA), hierarchical clustering, fuzzy clustering, biclustering, and mutual information analysis, each of which tackle different potential bias aspects of the data set (Eisen et al. 1998; Kerr et al. 2008).

PCA identifies data trends within samples, called principal components, such that very large data sets can be graphically represented using a smaller number of dimensions (Ringner 2008). This technique is especially useful for visually identifying batch effects or noise between samples, which may otherwise negatively affect downstream analysis.

Hierarchical clustering aims to create a hierarchy of gene groups, whereby relationships among genes are represented by a dendrogram. The shorter the length of dendrogram branches between objects, the more closely related the gene expression patterns are. These differences are assessed by pair-wise similarity functions. In this way, the method builds a hierarchy of gene groups by progressively merging clusters (Eisen et al. 1998). One of the major limitations of hierarchical clustering is that the decision-making for gene assignments is focused locally, without considering the data set as a whole, which can affect downstream interpretation (Tamayo et al. 1999).

Fuzzy clustering (also referred to as **soft clustering**) was developed to partially counteract the local bias of hierarchical clustering approaches. Fuzzy clustering allows for data elements to simultaneously belong to multiple groups with respect to a given criteria. Each data element has a “degree of belonging” to a cluster, instead of being assigned to an individual cluster and this degree represents how close the fit is in multiple clusters (Dembele and Kastner 2003; Fu and Medico 2007). This is in contrast to **hard clustering**, where data elements only are allowed to belong to one group.

Some of the newest clustering approaches have attempted to incorporate prior biological knowledge into the clustering algorithm. This has been attempted with a form of **biclustering** (Madeira and Oliveira 2004), a matrix-based clustering approach that includes both genes

and conditions in the algorithm. One example algorithm, called cMonkey, was used to identify and cluster sequence motifs in *Helicobacter pylori*, *S. cerevisiae*, and *Escherichia coli* based on microarray data sets (Reiss et al. 2006). A similar clustering approach that incorporates prior knowledge, called **mutual information analysis**, has also been shown to identify transcriptional interactions with a high fidelity in mammalian cells (Margolin et al. 2006). Finally, a number of standardized tools and analysis techniques are already publically available (Table 3.1) to facilitate transcriptional data analysis. To date, they have been able to provide detailed views of changing transcriptional landscapes in response to different environmental conditions on a functional level, and have been highly beneficial for the identification and prediction of virulence factors in fungi.

C. Proteomics

The term “proteome” was coined in 1996 to describe the complete set of proteins that is synthesized by a cell (Wilkins et al. 1996). The proteome provides the highest level of functional information of a cell, revealing the end product of the transcription and downstream transcriptional processing. The use of proteomics data sets is also becoming a popular approach for studying proteins involved in virulence.

The major areas in proteomics research include identification of proteins and their posttranslational modifications as well as protein–protein interactions.

These areas are investigated using two main methods, traditional two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) and, increasingly, mass spectrometry (MS). In 2D-PAGE, protein samples are resolved by two intrinsic properties: first in one dimension on an SDS gel, and then in the second dimension, at a 90° rotation (O’Farrell 1975); These properties can include their isoelectric point, protein complex mass in the native state, and protein mass. The properties chosen will depend on the specific experiment. Proteins are then visualized by staining of gels, often using silver, Coomassie Blue or Ponceau S staining techniques. Once visible, spots can then be picked out by hand or more often using automated detection software based on

Table 3.1. OMICS resources

-Omics	Methods	Standards	Databases	Analysis resources and tools	Fungal-specific resources
Genomics	CGA, NGS	MIGS	1000 Genomes http://www.1000genomes.org	ClustalW2 http://www.ebi.ac.uk/Tools/msa/clustalw2	Candida Database http://www.candidagenome.org
			ENSEMBL http://www.ensembl.org	UCSC Genome Browser http://genome.ucsc.edu	Aspergillus Database http://www.aspgd.org
Transcriptomics	Microarray, RNA-seq	MIAME, MINSEQE	GOLD http://www.genomesonline.org	IGV http://www.broadinstitute.org/igv	Saccharomyces Database http://www.yeastgenome.org
			Array Express www.ebi.ac.uk/arrayexpress	Galaxy https://main.g2.bx.psu.edu	YEASTRACT http://www.yeasttract.com
Proteomics	2D-PAGE, MS	MIAPE	GEO www.ncbi.nlm.nih.gov/geo	Bioconductor http://www.bioconductor.org	Filamentous Fungal Gene Expression Database http://bioinfo.townsend.yale.edu
			PRIDE http://www.ebi.ac.uk/pride	InterProScan http://www.ebi.ac.uk/Tools/ipfs/iprscan	Proteopathogen http://proteopathogen.dacya.ucm.es/
Metabolomics	NMR, HPLC-MS	MIAMET, CIMR	MassBank http://www.massbank.jp	APEX http://pfgrc.jvri.org/index.php/bioinformatics/apex	NetPhosYeast http://www.cbs.dtu.dk/services/NetPhosYeast
			HMDB http://www.hmdb.ca	Arcadia http://arcadiapathways.sourceforge.net	YEASTNET http://www.comp-sys-bio.org/yeastnet
Epigenomics	ChIP-chip, ChIP-seq	MIAME, MINSEQE	GOLM http://gmd.mpimp-golm.mpg.de	MetaboAnalyst http://www.metaboanalyst.ca	FunSecKB http://proteomics.yasu.edu/secretomes/fungi
			Roadmap http://www.roadmapepigenomics.org	STAR Genome Browser http://tabit.ucsd.edu/sdec	ChromatinDB http://www.bioinformatics2.wsu.edu/ChromatinDB
			ENDCODE http://genome.ucsc.edu/ENCODE	RMAP http://rulai.cshl.edu/rmap	Nucleosome Acetylation and Methylation in Yeast http://younglab.wi.mit.edu/nucleosome

The table includes a nonexhaustive list of -omics methods and resources with an emphasis on those available for fungi. Reporting standards are abbreviated as follows with their corresponding reference: *MIGS* minimum information about a genome sequence (Field et al. 2008); *MIAME* minimum information about a microarray experiment (Brazma et al. 2001); *MINSEQE* minimum information about a high-throughput sequencing experiment (<http://www.mged.org/minseqe/>); *MIAPE* minimum information about a proteomics experiment (Taylor et al. 2007); *MIAMET* minimum information about a metabolomics experiment (Bino et al. 2004); and *CIMR* core information for metabolomics reporting (<http://msi-workgroups.sourceforge.net/>).

their location on the gel. The identified spots are then excised, proteolytically digested, and then subjected to MS analysis. Briefly, MS measures the mass-to-charge ratio of charged particles such as peptides, and this information can then be used to identify the composition of the peptide and the gene it is derived from. Experimentally, MS samples are first vaporized and then ionized using an electron beam. The produced ions are then detected by the mass analyzer, which sorts the ions by their masses, and then processed into mass spectra where the detector measures the quality and quantity of the ions present. Variations of MS, including liquid chromatography tandem MS (LC-MS/MS) (Yates et al. 1999) and gel-free proteomics techniques (2012; Stastna and Van Eyk 2012) are also widely used approaches. These facilitate the analysis of proteins that are not easily separated in 2D gels due to their high hydrophobicity or high molecular weight, as in the case of many integral membrane proteins (Aebersold and Mann 2003; de Godoy et al. 2008). MS/MS involves additional rounds of ionization; however, the reproducibility between technical replicates of a sample remains in the range of 35–60% overlap (Tabb et al. 2010). Unfortunately, absolute protein quantification remains out of reach at the moment (Peng et al. 2012). Major hurdles remain to improve the reproducibility and standardization of the MS-based methods (Kniemeyer et al. 2011). Nonetheless, since 1996, the percentage of protein-coding genes in *S. cerevisiae* for which some biological function has been identified has increased to over 80%, greater than for any other sequenced eukaryotic genome (Botstein and Fink 2011). Proteomics studies have been highly beneficial in achieving this.

Proteomics approaches have led to the identification of a number of fungal virulence factors. Using an in vitro approach, the proteome of *C. albicans* yeast-form cells in the exponential or stationary growth phase was investigated in response to nutrient limitation using 2D-PAGE. The authors aimed to identify metabolic response patterns in these two cell types that might confer a tolerance phenotype (Kusch et al. 2008) similar to that observed in *S. cerevisiae* in response to stress (Herman 2002). They observed that the stationary phase cells upregulated a number of proteins, including those involved in the defense against reactive oxygen species and heat stress, as compared to exponentially growing cells. The ability to undergo morphological transitions between yeast and hyphal cells is an important virulence trait of many but not all *Candida* spp. This is especially important as the cell wall itself is always subjected to recognition by the host

cell surface and is thus exposed to immune recognition.

For example, a number of proteins are expressed in the yeast or hyphal stage only, suggesting a potential mechanism for secretion of cytosolic proteins, which may contribute its overall virulence in these different morphological states (Ebanks et al. 2006). These data further support the idea that the regulation of Hsp90, an essential chaperone protein that is activated in response to stress, is posttranscriptional in hyphal cells.

Additional variations of MS, such as matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) MS has been shown to be a useful tool in drug susceptibility screening of *C. albicans* to fluconazole (Marinach et al. 2009). A complete map of the yeast proteome using MS has most recently been completed, using a combination approach of high-throughput peptide synthesis in conjunction with MS for the *S. cerevisiae* proteome. This study provides insight into the evolution of yeast proteins and protein complexes (Picotti et al. 2013).

For *A. fumigatus*, the conidia mediates the initial contact with the immune system of the host and therefore is an interesting target for proteomics studies looking for fungal virulence factors and posttranscriptional responses upon host recognition. For example, comparison of the proteome profiles of *A. fumigatus* conidia and mycelial cells revealed some 50 conidia-specific proteins (Teutschbein et al. 2010). Interestingly, the data suggested that many proteins that are not needed during the resting stage are stored, perhaps for a rapid response to the activation of metabolic processes or in response to recognition by the immune system.

In vitro co-culture approaches with immune cells have also been influential in investigating differential protein expression in response to infection conditions. Using a time course of interaction between *C. albicans* and macrophages, a combination of proteomics and transcriptomics techniques highlighting specific pathways related to the virulence of *Candida* spp, including the regulation of apoptosis (Fernandez-Arenas et al. 2004; Fernandez-Arenas et al. 2007), was performed. The authors used a *C. albicans* strain of attenuated virulence, in which the kinase HOG1, important for the

oxidative stress response, was absent. They were able to identify several novel *C. albicans* antigens and further characterized the protective antibody response of mice against *C. albicans* infection. The use of proteomics methods, in general, has been useful in validating transcriptional data sets. However, they also revealed a number of discrepancies between the transcriptome and the proteome, which remains an active area of research in the validation of fungal virulence factors. Finally, **many resources have recently become available**, including the Proteopathogen Database for studying host–pathogen interactions (<http://proteopathogen.dacya.ucm.es>) with *C. albicans*, Compluyeast (<http://compluyeast2d.page.dacya.ucm.es/cgi-bin/2d/2d.cgi>), which catalogues 2D-PAGE data sets from *C. albicans*, *Mus musculus*, and *S. cerevisiae* for comparative proteomics.

E. Metabolomics

Metabolites are the products of metabolism or reaction intermediates and are usually small molecules serving a number of functions within the cell, including signaling and inhibition or stimulation of enzymes, among a number of other functions. As reaction intermediates, **metabolites provide the “missing-link” between DNA, RNA, and protein interactions within a cell**. One of the major themes of metabolomics is to investigate the influence of metabolites on cellular phenotypes. The metabolome is composed of intracellular metabolites and the exo-metabolome, also referred to as the secretome, which contains all small molecules secreted from a cell. It has been estimated that over 70% of metabolites participate in more than two biological reactions, and therefore represent interesting molecules for SysBio approaches (Nielsen 2003). Furthermore, from an evolutionary perspective, it is expected that a number of the filamentous fungi share their primary metabolism with their yeast ancestor *S. cerevisiae*, suggesting a broad applicability of the metabolomics research in the fungal research community. In fungal cells, there is an estimated number of more than 1,000 metabolites in the steady state (Smedsgaard and Nielsen 2005), some of which

are extremely short-lived or of low abundance, making their quantification a formidable challenge.

A number of methods are in use to identify metabolite profiles in cells. The most common are nuclear magnetic resonance (NMR) spectroscopy, MS (see Sect. II.C) as well as metabolic labeling with radioactive isotopes (Niittylae et al. 2009; Zamboni and Sauer 2009). Another method for investigation of metabolomics is gas chromatography coupled to mass spectrometry (GC-MS). GC-MS utilizes GC with detection by MS. GC is used in analytical chemistry to separate and identify molecules based on their migration within a capillary system. The sample is vaporized and travels through the capillary using an inert carrier gas. The time it takes for each molecule to elute from the column will vary according to its molecular properties and therefore can be used to identify compounds. Combining this elution with MS gives a highly detailed description of the molecule. High performance liquid chromatography (HPLC) is often used in combination with MS (HPLC-MS). HPLC is a chromatographic purification technique using a high-pressure capillary tube system, allowing for the fine separation of molecules. These methods, among others, provide a comprehensive way to identify the structure of metabolites on a genome-wide scale.

The identification and function of metabolites is highly relevant for a better understanding of fungal virulence. Fungi, more so than other pathogenic species, are notoriously known for the diversity of metabolites produced in response to host immune defense, and are thus useful organisms for studying metabolic diversity (Jewett et al. 2006). Notably, about a dozen *A. fumigatus* **secondary metabolites** have been implicated in niche adaptation and virulence (Galagan et al. 2005). To date, significant progress has only been made in metabolic profiling of fungi such as *S. cerevisiae*. The first metabolic network reconstruction of *S. cerevisiae* used an extensive data-mining approach of previous literature in combination with mathematical techniques to identify approximately 600 metabolites (Forster et al. 2003). Shortly thereafter, GC-MS methods were able to verify the presence of approximately 100 of these metabolites under standard laboratory growth conditions (Villas-Boas et al. 2005). **Metabolic flux in over 30 *S. cerevisiae* mutants demonstrated robustness and inherent redundancies built into yeast metabolism** (Blank et al. 2005).

In *C. albicans*, LC-tandem MS was used to profile the regulation of the secretome under standard laboratory conditions (Sorgo et al. 2010) and in response to the antifungal agent fluconazole (Sorgo et al. 2011), identifying numerous immunogenic peptides as novel vaccine candidates for antifungal therapy. Recently, the metabolome of *A. fumigatus* was investigated using 1H-NMR metabolomics under infection conditions (Grahl et al. 2011). Using this technique, the authors detected ethanol in the lungs in a murine model of invasive pulmonary aspergillosis, suggesting a role for fungal alcohol dehydrogenase in pathogenesis (Grahl et al. 2011). 1H-NMR metabolomics also enabled the identification of pneumococcal or cryptococcal meningitis without prior sample culture, which if implemented in a clinical setting would speed up the time it takes for patients to be diagnosed (Himmelreich et al. 2009).

F. Epigenomics

Among the biologically relevant -omics approaches, the most recent addition, epigenomics, has entered center stage. **The epigenome describes the global epigenetic modifications that take place within a cell.** Epigenetic modifications take place on the DNA, histones, and chromatin in its various functional states. They use numerous posttranslational modifications, including, but not limited to, the addition of single or multiple methyl residues, ubiquitination, acetylation, phosphorylation, or adenylation just to name the most common modifications [for review see (Hnisz et al. 2011)]. Most importantly, many modifications are reversible, providing an additional and even heritable level of cellular regulation.

The most common methods for investigation of the epigenetic landscape are studies on the variation of chromatin states using chromatin immunoprecipitation (ChIP).

Combinations of ChIP with microarray technology, known as ChIP-Chip or ChIP-on-Chip, and a similar combination of ChIP with NGS technology, termed ChIP-seq, have been recently introduced. ChIP identifies transient in vivo protein-DNA complexes by

crosslinking DNA and associated proteins within a cell lysate. The DNA is then fragmented either by sonication or nuclease digestion. The proteins of interest are then selected using an antibody, precipitated, purified, and the associated DNA is either sequenced or placed on a microarray, depending on the technology used.

ChIP-Chip has been used to investigate genome-wide changes in patterns of histone methylation in the fission yeast *S. pombe*. A complex composed of two proteins, Swm1 and Swm2, mediates demethylation of lysine 9 in histone H3 (H3K9) (Opel et al. 2007). Epigenetic regulation via this complex, in concert with additional histone deacetylases and chromatin remodelers, is a major factor in the transcriptional regulation of *S. pombe* (Opel et al. 2007). In *C. albicans*, Nobile and colleagues identified the transcriptional network for controlling biofilm formation using a combination of ChIP-Chip and in vivo animal models. The six identified core transcriptional regulators, regulating over 1,000 target genes, provide insight into biofilm formation during host infection (Nobile et al. 2012). In yeast, ChIP-Chip was used to investigate histone and gene deletion mutants during environmental stress, highlighting the importance of epigenetic regulation in this process (Weiner et al. 2012).

In *C. neoformans*, the size of the capsule increases under infection conditions and is a well-established virulence factor of the species. The direct targets of Ada2 in *C. neoformans* were recently investigated using ChIP-seq (Haynes et al. 2011). Ada2 is a member of the Spt-Ada-Gcn5 acetyltransferase (SAGA) complex, which regulates transcription by histone acetylation. The authors identified a relationship between the function of Ada2 and capsule size, linking this epigenetic modification and its targets to the overall virulence of the species (Haynes et al. 2011). Most recently, in *C. albicans*, a role of chromatin-modifying enzymes in the inhibition of the yeast-to-hyphal transition was discovered using a combined approach of ChIP-seq and RNA-seq. The authors identified a role for the histone deacetylase Set3/Hos2 complex (Set3C) as a transcriptional cofactor of metabolic and morphogenesis-related gene expression. They found that the acetylation status of

C. albicans chromatin influences transcription kinetics at target genes, showing that the epigenetic regulation supersedes a core transcriptional factor circuit involved in morphogenesis, a circuit that might be shared among other fungal pathogens (Hnisz et al. 2012).

G. Data Mining Approaches and Genome-Wide Fungal Resources

1. Databases

High-throughput molecular biology techniques have enormously increased the sheer volume of data generated and the need for proper data storage has never been higher (Kersey and Apweiler 2006). The value of this biological information is dependent on the ability of researchers to access and extract the information in a quick and reliable format, but also requires **high-level curation**.

Databases classify, organize, and systematize information. The maintenance of databases is essential in disseminating biological data to the community. The early development of two excellent databases for *S. cerevisiae*, the *Saccharomyces* Genome Database (www.yeastgenome.org/) and the Yeast Proteome Database (<http://www.proteome.com/YPDhome.html>) led to the rapid use of *S. cerevisiae* as a functional genomics tool and model organism (Botstein and Fink 2011). Despite their importance, maintaining high-quality, reliable databases is a constant struggle (Baker 2012), partly because of inherent high costs required for the curation of ever-changing biological data sets. Systems biologists extensively use databases for retrieval of prior knowledge, both qualitative and quantitative, on the biological question to increase the strength predictability, and identifiability models.

Major database initiatives include PubMed (<http://www.ncbi.nlm.nih.gov/pubmed/>), Kyoto Encyclopedia of Genes and Genomes (KEGG) (<http://www.genome.jp/kegg/>), and Gene Ontology (GO) (<http://www.geneontology.org>), all of which have established themselves as staple websites for researcher. Specialized databases are also becoming increasingly popular, such as the database of virulence factors in fungal pathogens (<http://sysbio.unl.edu/DFVF/>), which enable inclusion of more

in-depth information about a specific topic that may not be sufficiently covered in larger databases.

An additional benefit to including scientific data into databases is the ability to standardize the reporting format, facilitating both the integration of distinct data sets from different laboratories and the development of analysis tools. Standardization has been greatly aided with the push for minimum reporting guidelines for biological and biomedical information (Taylor et al. 2008) (<http://mibbi.sourceforge.net/>). Reporting guidelines now exist for all major -omics methodologies (Table 3.1) and there has been a general push from the scientific community to adhere to and popularize these standards for biological information.

2. Strain Collections

Genome-wide profiling at the RNA and protein level has been greatly aided by **publicly available strain libraries** in the form of loss-of-function (deletions) and gain-of-function (overexpression) collections. They have provided an efficient screening tool for scientists worldwide to investigate transcriptional and posttranscriptional changes in response to external stimuli, such as drug treatment and environmental variation, or exposure to host immune surveillance. Specifically for fungi, they have increased the throughput of virulence factor screening.

Of all fungi, *S. cerevisiae* has contributed the most number of strain collections. Starting with the yeast knockout (YKO) strain collection, this set methodologically deletes open reading frames (ORFs) by substituting the gene of interest with a selectable drug-resistance cassette, allowing for the systematic screening of the effects of gene loss (Winzeler et al. 1999; Giaever et al. 2002). More than 20,000 strains are currently available from the *Saccharomyces* Genome Deletion Project (http://www-sequence.stanford.edu/group/yeast_deletion_project/), including both homozygous and heterozygous diploid deletions, *MAT α* and *MATa* haploids, green fluorescent protein (GFP)-tagged (Huh et al. 2003), even essential temperature-sensitive collections (Li et al. 2004, 2011; Yan et al. 2008).

To investigate the posttranscriptional regulation of a gene and the number of protein complexes in the *S. cerevisiae* by proteomics, a Tandem Affinity Purification (TAP) collection was created in which each ORF is tagged with a high-affinity epitope expressing the protein at its native locus (Ghaemmaghami et al. 2003). To investigate protein–protein interactions with *S. cerevisiae*, a yeast two-hybrid collection was created, where hybrid proteins were derived from over 6,000 transformations and a Gal4 transcription–activation domain vector was inserted to create a hybrid protein for each ORF (Uetz et al. 2000). Data sets generated using this collection have also been made publically available (<http://portal.curagen.com/>). Several overexpression libraries for *S. cerevisiae* are also available, including a yeast GAL-GST library of over 5,000 strains, containing inducible overexpressed tagged ORFs from the GAL1 promoter, covering over 80% of the genome (Sopko et al. 2006). Additionally, the overexpression transformable plasmid based library of *S. cerevisiae* has been created, including over 13,000 entries with over 95% functional coverage of the genome (Jones et al. 2008). Approximately three-quarters of the *S. cerevisiae* proteome are also covered by the chromosomally C-terminal-tagged GFP fusion proteins strains (Huh et al. 2003). Using this library of 4,159 yeast–GFP clones, the localization of proteins in response to external stimuli can be easily visualized by live-cell fluorescence microscopy.

After the early success of strain collections in *S. cerevisiae*, more focused collections of pathogenic fungi have been created in order to specifically address fungal virulence. In *C. albicans*, a homozygous deletion library of approximately 670 homozygous deletion strains affecting 11% of the *C. albicans* genome was used to screen for virulence in a mouse model of infection, identifying 115 infectivity-attenuated mutants (Noble et al. 2010). A knockout collection of *C. albicans* transcriptional regulators includes over 100 strains, which were screened in 55 different growth conditions (Homann et al. 2009). Among the phenotypes identified, a number of them showed altered

susceptibility towards antifungal treatment. These results also support the theory that there is a high redundancy among transcriptional regulatory circuitry, where a single knockout does not greatly affect the strain’s overall virulence. In *C. neoformans*, a knockout collection of 1,201 genes was screened in an in vitro model of murine lung tissue for virulence phenotypes (Liu et al. 2008). Using these collections, a number of previously uncharacterized genes were identified as virulence factors, including those involved in growth at body temperature and in melanization, and those dependent and independent of capsule formation.

Smaller arrayed mutant collections of *Neurospora crassa* and *A. fumigatus*, among other pathogenic fungi, are available from the Fungal Genetics Stock Center (<http://www.fgsc.net/>) for screening. Finally, a single gene deletion collection comprising around 650 haploid *C. glabrata* genes will become available to the community shortly (Schwarzmueller, unpublished data).

III. Modeling Biological Phenomena

Most, if not all, biological processes follow a **dynamic, nonlinear pattern**. Nonetheless, a biological process can be approximated via a set of mathematical expressions to form a mathematical model. In this context, a “model” is referred to as a description of a biological process using mathematical expressions of quantitative data rather than a graphical representation. Although SysBio studies do not exclusively rely on either high or low throughput data sets, and lean towards a combination of both when possible, mathematical models have become increasingly useful ways of representing the information. These studies integrate computational approaches with experimental data to gain a more complete picture of how cells, tissues, and organs of species function and how the entire genetic information is wired and connected. Although current experimental techniques allow detailed measurements, it is impossible to gain full information about the system from discrete data sets without considering the topology and dynamics of the interacting

components. Depending on what is known a priori about the specific system under investigation, and what one wants or needs to learn about it, different modeling approaches are required. Identifying the proper modeling approach is a critical point, as not all methods will be appropriate for each experimental question. The pros and cons for each approach should be weighed for each biological question (Di Ventura et al. 2006; Karlebach and Shamir 2008). This needs a close interaction between mathematicians and experimentalists because very often it is impossible to generate the experimental data required for a particular model. Visual examples and a summary of the pros and cons for the use of Boolean models, ordinary differential equations (ODEs), and Petri Nets (PNs) are provided in Fig. 3.2.

In short, mathematical models represent a **simplified and abstract view** of the studied phenomena. They are extremely useful in understanding dynamic and multifaceted systems and their perturbations. These models often encompass different levels of biological understanding. Although it is important to know the consecutive action of the individual components of the system, an understanding of the time scale under which these interactions take place is essential for the whole picture.

Processes such as the ability of the cell to respond and adapt to a stimulus can be modeled using a set of ODEs. These processes are triggered by molecular interactions, which are not spontaneous and any environmental signal will result in a cellular response. A cellular response involves consecutive activation of a list of proteins that together establish a signaling pathway. These proteins pass the signal to the transcriptional machinery via a transcription factor (TF) that sometimes shuttles between cytoplasm and nucleus. TFs regulate gene expression and define the amplitude, magnitude, and duration of the cellular response. Such interactions can be graphically represented by “networks” that connect the interacting molecules. For example, gene regulatory networks (GRNs) are types of pathways that consist only of genes; in the network, gene A is connected with gene B if its product regulates the activity of gene B. A Boolean approach is often applied to

study the topology of GRNs. Thus, based on the specific experimental question, different modeling approaches will be applicable. Here, we discuss in detail several computational approaches commonly applied to model experimental data sets, emphasizing how key attributes of fungal virulence have so far been investigated using modeling approaches.

A. Boolean Models

Boolean models are often used to infer GRNs from microarray data or other types of expression analysis (Hickman and Hodgman 2009). Boolean models were first introduced by Kauffman and colleagues (Glass and Kauffman 1973). In contrast to ODE models, a Boolean approach is a discrete type of modeling where time and states are represented by discrete values. Boolean models are suitable for studying biological problems that can be interpreted using a rather simple on/off behavior, such as gene transcription. For example, an algorithm called REVEAL (Liang et al. 1998) infers the network topology from expression data. A measure of element interactions is used to derive logic functions that define them. Boolean models are useful for studying the existence of steady states or whether a given network topology is robust (Li et al. 2004). Furthermore, the approach can be used when the precise network topology is uncertain and our primary goal is to understand the wiring of the interactions in the system. The topology of the system of interest must be known before implementing kinetic models. As an example, the Boolean approach was used to infer a *Drosophila* segment polarity GRN (Ay and Arnosti 2011).

A Boolean model is defined by n entities interconnected via k edges, forming a directed graph. Each model entity is in a state either “on” (1) or “off” (0). Using the example of gene transcription, each gene can be considered as either expressed, 1, or not, 0. In the **synchronous Boolean model**, the states of the model entities are simultaneously evaluated and updated at time $t+1$, according to the regulatory functions and variables states at time t . Such Boolean models are purely deterministic.

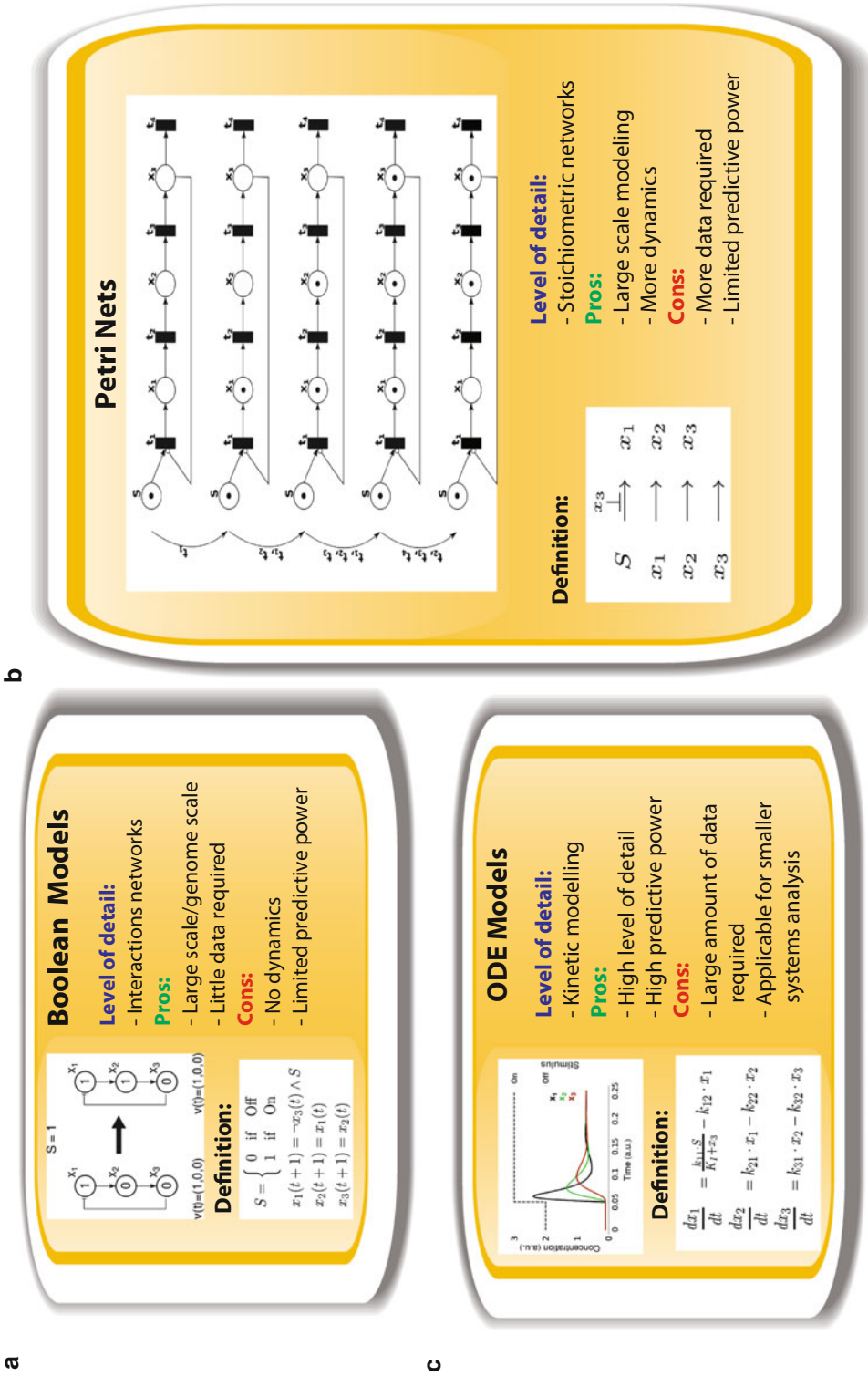


Fig. 3.2. Simulation of a negative feedback loop with different modeling techniques. (a) Boolean models incorporate only two values: 0 for the inactive state and 1 for the active state of a variable. They represent only an interaction network where no kinetic information can be included. (b) Petri Nets are an extension to the Boolean approach such that they include the stoichiometric information of the considered reaction network. The reaction can take place only if the right amount of variable is present. (c) ODE model. Parameters used: $x_1^0 = 0.1$, $x_2^0 = 0.1$, $x_3^0 = 0.1$, $S^{on} = 1$, $S^{off} = 0.1$, $k_{11} = 50$, $k_{12} = 99$, $K_I = 0.001$, $k_{21} = k_{22} = 50$, $k_{31} = k_{32} = 50$. Both the Boolean modeling approach and PN indicate oscillatory behavior of the system, whereas the ODE model, which incorporates higher level of detail, suggests that a system adapts to the external stimuli by reaching a new steady state

Regulatory relations are described via logic functions, such as the Boolean operators “and,” “or,” “not.” For vi being a vector representing a state of the model, we call a state space a set of all possible vectors vi . Thus, the state space has 2^n elements for n entities in the network, and vi are vectors of 0s and 1s. The elements of the state space are connected via arrows indicating the flow of model states.

In the **asynchronous Boolean model**, one node at a time is chosen and updated, and the evaluation of the next selected node state takes this change into account. If the order of choosing the nodes is fixed, then the model is called a **deterministic asynchronous Boolean model**; if the nodes are chosen at random, then the model is termed a **stochastic asynchronous Boolean model**. The state space typically contains single point attractors, which are fixed points (also called steady states) towards which the systems evolve into both synchronous and asynchronous Boolean models. However, the time needed to reach the fixed point can vary between synchronous and asynchronous modeling variants. Cyclic attractors, e.g., limit cycles, can be lost in the stochastic Boolean model (Wang et al. 2012). In either case, the identification of point and cyclic attractors of large-scale Boolean models is not a trivial task but there are algorithms that deal with this problem (Wang et al. 2012). Because stochastic processes influence any biological process, Boolean networks have been further developed to account for noise in the system and for making the approach suitable for the study of stochasticity and uncertainty. These include development of probabilistic Boolean networks (Shmulevich et al. 2002) or Boolean models where stochasticity is implemented by reversing a node’s state at some probability rate, or by implementing stochasticity of a biological function that fails to be executed (stochasticity in function, SIF, models) (Garg et al. 2009).

Although a Boolean approach only allows for a very simplified representation of a biological system, it can be a powerful method for studying its underlying nature. For instance, a Boolean model can be used for systematic screening of possible networks that reproduce

a pattern of interest (Giacomantonio and Goodhill 2010). Recently, Boolean modeling was used to study the interplay between gene expression, chromatin modifications, and DNA methylation, where the authors linked the epigenetic landscape with the probability state space (Flöttmann et al. 2012). This innovative application shows that there are other possible ways of analyzing Boolean models that are waiting to be explored.

B. Petri Nets

There is a growing interest in applying Petri Nets (PNs) to modeling and analysis of biological networks. **In principle, a PN is a directed bipartite graph, whose nodes are called either “places” or “transitions”.** Places indicate resources and they are indicated with circles; transitions are events or biochemical reactions, which are shown as boxes. Both types of nodes are connected by arrows. An arrow (called an “arc”) from a place (input place) to a transition indicates that a compound is necessary for the reaction. Further arcs point from the reaction to its products (see Fig. 3.3).

Each place holds a nonnegative number of tokens, which indicate resources of a given substance in the system. The state of a system is represented by an allocation of tokens at a given time point, which is called marking (M). Initial marking takes place at the time point zero. A transition will fire if there is a sufficient number of tokens on the input places, and it is at least equal to the edge’s weight. Once a transition fires, tokens are transferred into the respective output places and the number of tokens in the output places is again indicated by the weight of the arcs. In summary, formally a PN is a tuple $PN = (P, T, F, W, M_0)$, where P is a set of places, T is a set of transitions, F is a set of arcs, W is a map that assigns each arc with a specific weight, and M_0 is initial marking (Ackermann 2011). For the analysis of PNs, concepts like P-invariants and T-invariants are introduced. For instance, for a stoichiometric matrix N of a PN, a P-invariant, is any vector x where $x^T N = 0$ and it holds $\langle x, M \rangle = const.$ for any marking M

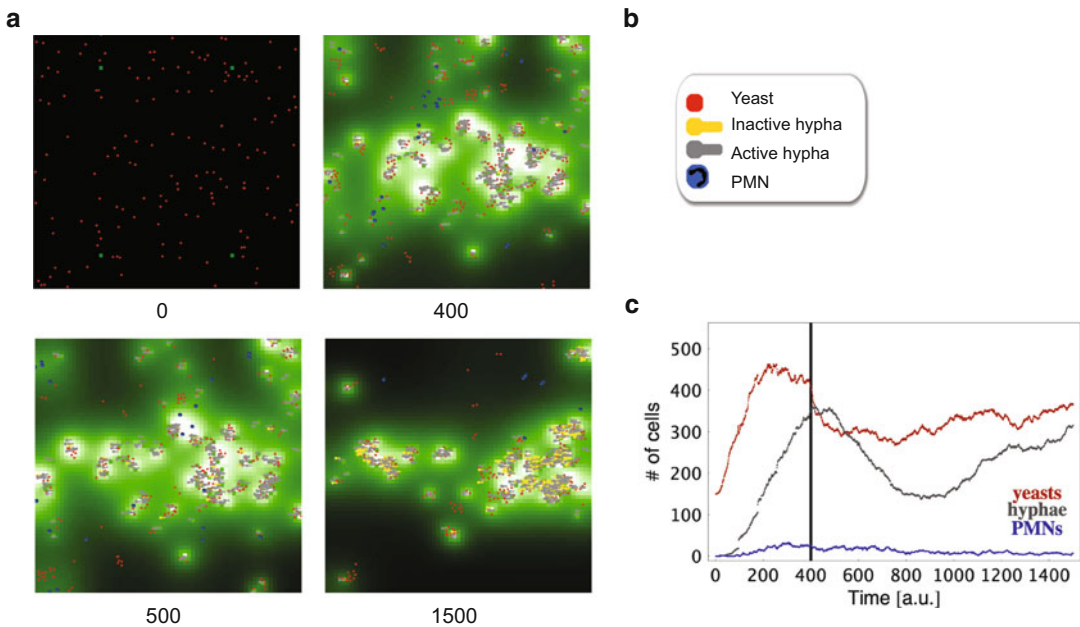


Fig. 3.3. Model simulation of the system treated with a low drug dose. Drug is applied at time $t=400$ (arbitrary units). Treatment does not clear the infection. After initial reduction, $t=500$, the fungal population recovers from the drug stress, $t=1,500$. (a) Model simulation output at different time points. (b) Agents considered

in the model: yeast cells (*red*), inactive hyphal cells (*yellow*), active hyphal cells causing damage to the host (*gray*), polymorphonuclear neutrophils (*blue*). (c) Graphical representation of the model simulation over time; y -axis gives the total cell number

that appears during the simulation of PN. The scalar product $\langle x, M \rangle = \text{const.}$ represents the conservation relation, e.g., $\text{ATP} + \text{ADP} = \text{const.}$ A T-invariant on the other side is any vector x where $Nx = 0$. These provide a decomposition of the network, and a transition that does not belong to any T-invariant can be removed from the PN. A set of T-invariants corresponds to the elementary flux modes operating in the system. The marking graph of a PN is a directed graph that represents the evolution of markings during PN simulation. This concept is similar to state space in Boolean models.

PN approaches are applicable for discrete modeling, or modeling of GRNs. For GRNs, like with Boolean models, one can consider both synchronous and asynchronous modes. PNs can also be applied to perform quantitative analysis and establish stochastic models. Fuzzy modeling has also been introduced based on PNs, and details on the application of the techniques and examples have been published (Ackermann 2011).

A stochastic version of a Petri Net (SPN) has been applied to the study of the cell cycle in budding yeast (Mura and Csikasz-Nagy 2008). The authors provide a fair comparison of the results from SPN with the results of the deterministic version of the corresponding ODE model, as proposed earlier (Novak 2002). Recently, a PN-based technique has been applied to integrate signaling, metabolic, and regulatory events participating in the *S. cerevisiae* HOG signaling pathway (Tomar et al. 2013).

C. Ordinary Differential Equation Models

Dynamics of biological processes are most often described via ODE models or partial differential equations (PDEs) when space is included. Optimally, ODE models are used for systems that can be considered “well-stirred” and that comprise large molecule numbers. When this condition is met, changes in molecule numbers can then be considered as

continuous (Di Ventura et al. 2006). ODE models can also be used to address questions that regard changes in cellular phenotypes (Le Novere et al. 2005; Karlebach and Shamir 2008). Each equation in an ODE model describes temporal changes of one variable such as molecule concentration, phosphorylation levels, cell mass, or volume. If we consider a system with n variables, characterized by k , and positively valued parameters p , then the equation is:

$$\frac{dx_i}{dt} = f_i(x_1, \dots, x_n, p_1, \dots, p_k, t) \quad (3.1)$$

where t represents time. The equation describes dynamical changes of the variable x_i . A mathematical expression describing an increased production of a compound or its activation enters Eq. 3.1 with a positive sign; expressions for degradation or deactivation enter with a negative sign. Taken together, the equation evolves, having m reactions determined by the topology of our network of n state variables, which may or may not influence each other. Thus, we can also represent Eq. 3.1 in a matrix form:

$$\frac{dx}{dt} = Nv \quad (3.2)$$

where N is a $n \times m$ stoichiometric matrix and $v = (v_1, \dots, v_m)^T$ is a vector that stores rates for all the reactions taking place in our system. For practical examples, refer to detailed earlier descriptions (Klipp et al. 2009). The solution of an ODE model is a time course simulation. Given the initial state in a deterministic model, all the future states can be computed. Simulations of the model reproducing experimental data are used for understanding the time course dynamics of the interacting molecules, generating testable predictions, and for design of new experiments. Understanding of the system through these simulations has proven useful for the study of fungal virulence (Chen et al. 2004; Klipp et al. 2005; Leach et al. 2012). Only using the analysis of the model simulations can reveal the purpose of integrating by the cell certain molecular circuits, e.g., by negative or positive feedback loops.

Another example is a molecular autoregulatory loop integrated by *C. albicans* adapting to heat stress (Leach et al. 2012). In this case, the authors developed a dynamic model of the heat stress response in *C. albicans* using a set of ODEs supported with experimental data. The model reveals several features of the system such as a memory for acquired thermotolerance. For example, when pretreated with a mild heat shock, the system becomes more resistant to a severe shock. Moreover, the simulations of the model indicate a transient molecular memory in the system that is mediated through phosphorylation of heat shock transcription factor Hsf1.

D. Flux-Balance Analysis

Flux-balance analysis (FBA) is a mathematical framework that is widely used for the analysis of the **flow of metabolites** throughout a metabolic network (Orth et al. 2010). This structural modeling approach solely requires knowledge of the stoichiometric matrix N of the biological network. This is generally a well-known property for metabolic networks and, hence, opens the way for genome-scale studies (Edwards and Palsson 2000; Price et al. 2003; Yus et al. 2009). The FBA approach aims to identify the optimal distribution of fluxes in the steady state, i.e., fluxes v that satisfy the following equation:

$$\frac{dx}{dt} = Nv \stackrel{!}{=} 0 \quad (3.3)$$

FBA aims to find fluxes where a given objective function reaches extreme values, for example, fluxes leading to the maximal growth rate (Feist and Palsson 2010) or the minimal production of toxic metabolites. The problem takes the form of linear programming. Here, problem constraints are included on the basis of the experimental results in the steady state (for instance, thermodynamics, biomass produced, or energy availability). These data are required to reduce the degree of freedom in the solution space.

The major advantage of using FBA is that there is no need to know reaction kinetics and metabolite concentrations, because FBA addresses steady-state conditions. However, the approach is inherently deprived of quantitative information (e.g., enzyme concentrations). Because of this restriction, FBA cannot

be used to predict, for example, how specific levels of a certain enzyme must be changed in order to achieve a desired effect on flux. It is important to keep in mind that FBA does not help us to understand the dynamics of the system because it can only reveal steady state properties.

To enable the re-use of FBA models, they should be provided in the Systems Biology Markup Language (SBML) format, similar to the case of ODE models. FBA models can then be imported and solved using algorithms such as Matlab or COBRA Toolbox (http://systemsbiology.ucsd.edu/Downloads/Cobra_Toolbox) (Becker et al. 2007).

An extension of FBA, the regulated flux-balance analysis (rFBA) has been extensively reviewed (Karlebach and Shamir 2008) and aims at integrating the metabolic network with regulatory processes that are expressed by Boolean logic (Covert et al. 2001). Regulatory events can reflect situations when, for instance, certain regulatory proteins are not expressed and then appropriate fluxes will be shut down and thus set to zero.

E. Stochastic Modeling

ODE models, which are deterministic, are suitable for the analysis of dynamical behavior of the population on average. These, however, do not provide information about whether and how stochastic switches or noise impact the outcome of the biological process. For instance, the lysogenic and lytic cycle of λ phage (Arkin et al. 1998) could not be explained by deterministic modeling. Stochastic models describe **random processes that evolve and change over time**. It is convenient to use stochastic modeling in cases where one wants to investigate processes in which a molecule with small copy number affects key components of the model or if steady states are unstable. In general, simulating stochastic models is computationally more expensive. Moreover, to enable statistically significant conclusions, many simulation results have to be analyzed together. Stochastic simulations can be performed using tools such as COPASI (Hoops

et al. 2006) or Cain (<http://cain.sourceforge.net/>). For a thorough introduction to stochastic modeling, we refer to recent work in yeast (Klipp et al. 2009). The ideal type of data for stochastic modeling are time-resolved measurements of single molecules, e.g., by microscopic measurements. In practice, experiments can rarely be repeated for the same single cells.

F. Monte Carlo Simulation

Monte Carlo (MC) simulation is a hit-or-miss sampling method. It is typically applied to find extremes of a function in a restricted region of possible parameter space. The MC sampling method can be viewed as randomly choosing parameters (x,y) and keeping the pair that gives, for example, the highest value of $f(x,y)$. This method, however, is not a systematic approach to approximate the optimal solution; simply put – each time the simulation is performed, we either hit or miss the solution.

For each run, parameter values are randomly changed as, for example, initial conditions or kinetic constants. Then, it is recorded whether such perturbations influence the final result of the model and, if so, how. The parameters to vary are those where a significant uncertainty is encountered. From the analysis of the created range of estimated final values of the model, one can estimate how likely it is that a certain outcome will occur. MC simulation usually evaluates the model from hundreds to tens of thousands of times to estimate the solution to the model.

The MC simulation method was applied to study the dependence of drug dosage treatments on host resistance against disseminated candidiasis (Hope et al. 2006). In another study, the MC method was applied to study anticancer drug target inhibition strategies on the epidermal growth factor signaling pathway. The authors investigated the influence of changes in kinetic parameters by comparing parallel simulation runs (Wierling et al. 2012). MC simulation was also proposed as a method for assessing the degree of completeness of GRNs, where information on gene interactions is often missing (Kuhn et al. 2009). Among other tools for performing MC simulations are PyBios (<http://pybios.molgen.mpg.de/>), MATLAB, and Statistics Toolbox or Simulink.

Finally, MC Markov Chain (MCMC) is an optimization method in high dimensional spaces. It is designed to randomly search the parameter space such that the optimal value can be approximated. The way that the parameter space is searched and tested for whether or not parameters are accepted is often defined by methods such as Gibbs sampling or Metropolis–Hastings algorithms (Hastings 1970). MCMC is applied to solve integrals in high dimensional spaces when the traditional numerical methods fail. A complete description of MCMC is beyond the scope of this chapter. For further details on the method, we refer readers elsewhere (Gamerman and Lopes 2006).

G. Agent-Based Models

Agent-based models (ABMs) are counterparts to ODE models, where simple rules for each agent's action can lead to complex dynamics in the population of interacting agents. **Agents are autonomous entities that can represent molecules, cells, and organisms.** ODE models are often applied to study the dynamical properties of regulatory pathways, leaving out information on spatial distribution of the molecules in the cell. This can be tackled using ABMs (Pogson et al. 2006, 2008). For example, ABM was applied to examine the pathogenesis of gut-derived sepsis using the example of *Pseudomonas aeruginosa* interaction with its host (Seal et al. 2011). In general, ABMs are suitable for study of systems where the spatial and temporal distribution of agents influences the systems dynamics, such as chemotaxis, while sensing a gradient of quorum molecules (Netotea et al. 2009; Fozard et al. 2012), inducing biofilm formation (Mitri et al. 2011), or pheromone concentration gradients during mating. ABM has been applied to study the functionality of the immune system (Folcik et al. 2011), granuloma formation (Segovia-Juarez et al. 2004) and for predicting the outcome of different immunotherapy strategies (Pappalardo et al. 2011).

All ABMs are systems of agents, whose actions and decisions are specified by the user. The agents are typically living in a 2-D world, which is a square divided into a grid of patches

(or triangulated space). These patches, representing the environment, can influence an agent's actions, and agents can affect the attributes of the environment themselves as well. ABM is a discrete and stochastic modeling approach. At each step, an agent makes a decision according to its status at the time. Dynamics of biological systems are complex and although ABMs appear conceptually clear, the code of ABMs tend to have particularly long lines of code, numbering even into the thousands, which can make them difficult to handle.

ABMs have gained popularity, particularly in modeling the immune system, disease dynamics, and as a tool for elucidating the nature of host–pathogen interactions. Agent-based modeling is a modeling technique where each agent's action can be defined with simple rules. Furthermore, using ABM simulations, **it is possible to track dynamics of a single agent rather than the averaged behavior of a population.** Tools have already been developed for simulating immune dynamics whereby the user can specify the rules for an agent's interactions, including IMMSIM, SIMMUNE, SIS, and reactive animation [reviewed in (Bauer et al. 2009)]. Computational-oriented studies on host–pathogen interactions can be performed using CyCells, PathSim, MASyV (Bauer et al. 2009), or BSim (Gorochowski et al. 2012), which are freely available at <http://bsim-bccs.sf.net>.

ABMs have been implemented for the study of fungal pathogenesis using an additional ABM tool referred to as NetLogo (freely available at <http://ccl.northwestern.edu/netlogo/docs/>). NetLogo was used to study *C. albicans* interactions with its human host (Tyc and Klipp 2011). The authors explored the rules that determine the dynamics of a fungal population influenced by host phagocytic cells (Fig. 3.4). The model was then used to investigate the effects of potential drug treatments on fungal populations and their clearance. Another example is a study on *A. fumigatus* population clearance by neutrophils, where different rules defining neutrophil movement were examined, such as chemotaxis along a chemokine gradient, random walk, and communication between the phagocytes (Tokarski et al. 2012).

H. Game Theory

Humans have developed numerous strategies to protect themselves against invading patho-

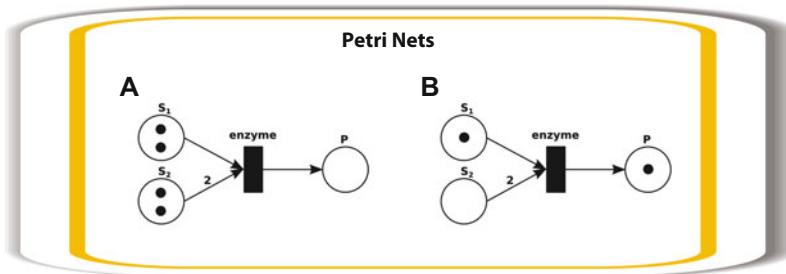


Fig. 3.4. Petri Nets. An example PN of the reaction $S_1 + 2S_2 \xrightarrow{\text{enzyme}} P$ is shown. Marking of the PN is presented (a) before the reaction takes place and (b) after enabling the reaction

gens, such as controlling pathogen growth and dissemination by clearing pathogens from the body through activation of the innate and adaptive immune system. Host-microbe interactions are therefore processes that can be viewed as shifting the balance between different populations of cells that will lead to either a healthy or diseased host state. Populations evolve and they establish equilibrium states in accordance with the equilibrium of the other populations, and this process repeats *ad infinitum* or until one partner is removed from the interaction, i.e., by death of the host or through pathogen clearance. Evolutionary game theory is an approach suitable for modeling dynamics of evolutionary processes that are assumed to modify the population fitness landscape. This can be assumed because normal optimization methods for population fitness or flux rates are not affected by interactions between individuals and the environment (Pfeiffer and Schuster 2005). In short, game theory (GT) is a framework suitable for modeling biological systems where **distinct strategies can be assigned to the individual partners.**

GT has been applied to the study of survival strategies of *C. albicans* in macrophages (Hummert et al. 2010) and to growth strategies adapted by individual cells when different carbon sources are given (Friesen et al. 2004). GT was also exploited to study distinct utilization of metabolic pathways by cell populations (Pfeiffer and Schuster 2005; Ruppin et al. 2010), to study cooperation during evolution (Nowak et al. 2004), and as an application for optimizing altruistic behavior in microbial populations (Schuster et al. 2010). GT was used in the analysis of cancer cells (Gatenby and Vincent 2003), for multiple knockout analysis in *S. cerevisiae* (Kaufman

et al. 2004), and for the analysis of microarrays data via defining coalitional game sets (Albino et al. 2008; Moretti et al. 2007, 2008).

I. Model Parameters

Simulation of a dynamic model requires knowledge of the kinetic parameter values. These, however, are often unavailable or extremely difficult to obtain from experimental data. Thus, a process often termed parameter estimation or more precisely, regression (Jaqaman and Danuser 2006) has to be employed. Parameter estimation is a typical inverse problem (i.e., deducing from effects to their causes) and its objective function is the set of parameter values that best represent the data. There are many algorithms suitable for solving such optimization problems that minimize the distance between experimental data points and simulation results of the model, which focus both on local and global optimization methods (Moles et al. 2003; Baker et al. 2010). Although global optimization methods search for the solution to the problem by scanning the entire parameter space, they are computationally more expensive and time-consuming than local optimization methods. Comparing a global optimization method and a local one, the latter method is faster; however, it is limited to providing a suboptimal solution, which in some cases might be only a local optimum. Regardless of the method utilized, algorithms tend to minimize a sum of squared residuals (RSS) given by:

$$\min_p \sum_{i=1}^n |y(t_i) - M(x(t_i), p)|^2 \quad (3.4)$$

where $y(t_i)$ are experimental data points and $M(x(t_i), p)$ are the values of the simulated model; $x(t_i)$ is a corresponding model variable and p is a set of parameter values.

J. Sensitivity Analysis

Most mathematical models require parameters to be estimated from experimental data sets. The quality of the data fit, however, depends both on reproducibility of experimental data and model structure itself. We can study how the changes in model parameters influence its output – a technique termed **sensitivity analysis**. Parameters with marginal effects on a model output cannot be estimated from experimental data, nor will their numerical values significantly affect the quality of model predictions. Therefore, in order to estimate a set of parameters for a given model, one needs to **focus on generating experimental data for the components that are strongly influenced by these parameters**. Sensitivity analysis can be performed either locally or globally. Using local sensitivity analysis, we can analyze the effects of comparatively small perturbation changes in the parameter (p) on the model output (O), such as at steady state (SS), and can represent the effects using the following mathematical equation:

$$R_p^O = \frac{p}{O^{SS}} \frac{\partial O^{SS}}{\partial p} \quad (3.5)$$

Local sensitivity analysis does not consider multiple parameter interdependencies. It also tends not to be robust, meaning that results will be partially affected by the parameters used in the model. By contrast, global techniques for performing sensitivity analysis search the entire parameter space, taking into account many parameters values rather than only one. Global techniques not only consider

larger variability in parameter values, but they also provide a measure for parameter interactions (Frey and Patil 2002; Marino et al. 2008).

The matrix of coefficients obtained from sensitivity analysis can be analyzed to address the question of structural identification of our model. Specifically, whenever (i) each column contains a large absolute value (i.e., each parameter has a strong effect on at least one model variable), and (ii) columns are linearly independent, then the model is structurally identifiable (Jaqaman and Danuser 2006). In summary, sensitivity analysis can help the process of parameter estimation. Although a strategy for estimating parameters is very model-dependent, there are no golden standards and, hence, it will rely on the modeler's experience and the biological context.

K. Standards for Modeling

Standards in computational biology are necessary to ease the exchange of the results of research between scientific communities. SBML has been proposed for describing models of signaling pathways and GRNs (Hucka et al. 2003). SBML is used for formatting ODE models or a system of ordinary differential and algebraic equations (DAE) such that they can be re-used in other software tools. The need for standardization of models is evident from an ever-increasing number of software tools for model implementation. Lack of standards will inevitably lead to incompatibility of the models between different tools, and make their broad use impossible. **Minimum information requested in the annotation of biochemical models (MIRIAM) has been proposed** (Le Novère et al. 2005). Proper annotations of model components (using ontology names and unique database identifiers) allow for the comparison of different models but also enable model merging. Model annotation can be done either manually or using tools such as semanticSBML (<http://www.semanticsbml.org>, (Krause et al. 2010)).

IV. Conclusions and Future Perspectives

As experimental techniques improve and evolve, the dimensionality of the biological problems under investigation increases in parallel. Our knowledge about system properties increases exponentially with the amount of biological data. Initially, studies were focused on understanding the functionality of single genes and followed strictly reductionist schemes. The availability of high-throughput and genome-wide data sets (genomics, metabolomics, proteomics epigenomics, etc.) has dramatically extended the size and complexity of biological problems because we have remained almost paralyzed in our efforts to integrate data to gain a better understanding of systems. Hence, there is an emerging need to come up with creative and efficient solutions for integrating large-scale data into meaningful biological context.

The ever-increasing amount of genome-wide data sets has significantly aided our understanding of the nature of cellular responses and how these have evolved. It has helped us to identify novel network components so that it is possible to identify what makes, for example, one species more resistant to an antifungal drug compared to another.

Modeling techniques across the board have been useful in the visualization of both large- and small-scale data sets. For each technique, it is important to keep in mind the kind of information that is being included, and what the expected outcome of the system is. For instance, protein-protein interactions between molecules do not always correlate with a conserved GRN (Roguev et al. 2008). One should also interpret these models cautiously, especially when perturbing a mathematical model by including gene mutations or alternations in the system, since such alterations may also change physical interactions of the proteins and therefore may not fit the proposed model structure (Goh et al. 2007; Zhong et al. 2009). There are also benefits and limitations to modeling a single cell versus a cell population. It is imperative to weigh the pros and cons of each of these limitations when modeling biological data.

The generation and cataloguing of further data sets and models will become increasingly important in fostering our understanding of fungal virulence and for the prediction of alternative or even novel therapeutic strategies.

Acknowledgments Work on systems biology in the KK group has been supported by a grant from the Christian Doppler Research Society, and by grants of the Austrian Science Foundation (FWF-DACH Proj.: I-746-B11, FWF-P25333-B09), by the SysMo-MOSES project through the FFG, the WWTF, and by the IP-UNICELLSYS from the FP7 programme. Work on systems biology of fungal infection in the EK group has been supported by grants from the German Research Council (DFG: CRC618), the EU Marie-Curie Program (FinSysB, PITN-GA-2008-214004) and the EU FP7 programme (IP-UNICELLSYS and IP-SysteMtb). We would like to thank Sara Tierney for contributing to the artwork.

References

- Ackermann J, Koch I (2011) Quantitative analysis. In: Koch I, Reisig W, Schreiber F (eds) Modeling in systems biology: the Petri Net approach. Computational biology, vol 16. Springer, Berlin, pp 153–178
- Aebbersold R, Mann M (2003) Mass spectrometry-based proteomics. *Nature* 422:198–207
- Albino D, Scaruffi P, Moretti S, Coco S, Truini M, Di Cristofano C et al (2008) Identification of low intratumoral gene expression heterogeneity in neuroblastic tumors by genome-wide expression analysis and game theory. *Cancer* 113:1412–1422
- Arkin A, Ross J, McAdams HH (1998) Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected *Escherichia coli* cells. *Genetics* 149:1633–1648
- Ay A, Arnosti DN (2011) Mathematical modeling of gene expression: a guide for the perplexed biologist. *Crit Rev Biochem Mol Biol* 46:137–151
- Baker M (2012) Databases fight funding cuts. *Nature* 489:19
- Baker SM, Schallau K, Junker BH (2010) Comparison of different algorithms for simultaneous estimation of multiple parameters in kinetic metabolic models. *J Integr Bioinform* 7:3
- Bauer AL, Beauchemin CA, Perelson AS (2009) Agent-based modeling of host-pathogen systems: the successes and challenges. *Inform Sci* 179:1379–1389
- Becker SA, Feist AM, Mo ML, Hannum G, Palsson BO, Herrgard MJ (2007) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA toolbox. *Nat Protoc* 2:727–738
- Bino RJ, Hall RD, Fiehn O, Kopka J, Saito K, Draper J et al (2004) Potential of metabolomics as a functional genomics tool. *Trends Plant Sci* 9:418–425

- Blank LM, Kuepfer L, Sauer U (2005) Large-scale ^{13}C -flux analysis reveals mechanistic principles of metabolic network robustness to null mutations in yeast. *Genome Biol* 6:R49
- Botstein D, Fink GR (2011) Yeast: an experimental organism for 21st century biology. *Genetics* 189:695–704
- Braun BR, van Het Hoog M, d'Enfert C, Martchenko M, Dungan J, Kuo A et al (2005) A human-curated annotation of the *Candida albicans* genome. *PLoS Genet* 1:36–57
- Bray D (2003) Molecular networks: the top-down view. *Science* 301:1864–1865
- Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C et al (2001) Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet* 29:365–371
- Brenner S, Noble D, Sejnowski T, Fields RD, Laughlin S, Berridge M, Segel L, Prank K, Dolmetsch RE (2001) Understanding complex systems: top-down, bottom-up or middle-out? In: Bock G, Goode J (eds) Complexity in biological information processing, vol 239, Novartis foundation symposium. Wiley, Chichester, pp 150–159
- Bruggeman FJ, Westerhoff HV (2007) The nature of systems biology. *Trends Microbiol* 15:45–50
- Bruno VM, Wang Z, Marjani SL, Euskirchen GM, Martin J, Sherlock G et al (2010) Comprehensive annotation of the transcriptome of the human fungal pathogen *Candida albicans* using RNA-seq. *Genome Res* 20:1451–1458
- Butler G, Rasmussen MD, Lin MF, Santos MA, Sakthikumar S, Munro CA et al (2009) Evolution of pathogenicity and sexual reproduction in eight *Candida* genomes. *Nature* 459:657–662
- Chen KC, Calzone L, Csikasz-Nagy A, Cross FR, Novak B, Tyson JJ (2004) Integrative analysis of cell cycle control in budding yeast. *Mol Biol Cell* 15:3841–3862
- Covert MW, Schilling CH, Palsson B (2001) Regulation of gene expression in flux balance models of metabolism. *J Theor Biol* 213:73–88
- Cui X, Churchill GA (2003) Statistical tests for differential expression in cDNA microarray experiments. *Genome Biol* 4:210
- Cuomo CA, Birren BW (2010) The fungal genome initiative and lessons learned from genome sequencing. *Methods Enzymol* 470:833–855
- de Godoy LM, Olsen JV, Cox J, Nielsen ML, Hubner NC, Frohlich F et al (2008) Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature* 455:1251–1254
- Dembele D, Kastner P (2003) Fuzzy C-means method for clustering microarray data. *Bioinformatics* 19:973–980
- Di Ventura B, Lemerle C, Michalodimitrakis K, Serrano L (2006) From in vivo to in silico biology and back. *Nature* 443:527–533
- Dujon B (1993) Mapping and sequencing the nuclear genome of the yeast *Saccharomyces cerevisiae*: strategies and results of the European enterprise. *Cold Spring Harb Symp Quant Biol* 58:357–366
- Ebanks RO, Chisholm K, McKinnon S, Whiteway M, Pinto DM (2006) Proteomic analysis of *Candida albicans* yeast and hyphal cell wall and associated proteins. *Proteomics* 6:2147–2156
- Edwards JS, Palsson BO (2000) The *Escherichia coli* MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. *Proc Natl Acad Sci USA* 97:5528–5533
- Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 95:14863–14868
- Feist AM, Palsson BO (2010) The biomass objective function. *Curr Opin Microbiol* 13:344–349
- Fernandez-Arenas E, Molero G, Nombela C, Diez-Orejas R, Gil C (2004) Contribution of the antibodies response induced by a low virulent *Candida albicans* strain in protection against systemic candidiasis. *Proteomics* 4:1204–1215
- Fernandez-Arenas E, Cabezon V, Bermejo C, Arroyo J, Nombela C, Diez-Orejas R et al (2007) Integrated proteomics and genomics strategies bring new insight into *Candida albicans* response upon macrophage interaction. *Mol Cell Proteomics* 6:460–478
- Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P et al (2008) The minimum information about a genome sequence (MIGS) specification. *Nat Biotechnol* 26:541–547
- Flöttmann M, Scharp T, Klipp E (2012) A stochastic model of epigenetic dynamics in somatic cell reprogramming. *Front Physiol* 3:216
- Folcik VA, Broderick G, Mohan S, Block B, Ekbote C, Doolittle J et al (2011) Using an agent-based model to analyze the dynamic communication network of the immune response. *Theor Biol Med Model* 8:1
- Forche A, Magee PT, Selmecki A, Berman J, May G (2009) Evolution in *Candida albicans* populations during a single passage through a mouse host. *Genetics* 182:799–811
- Forster J, Famili I, Fu P, Palsson BO, Nielsen J (2003) Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res* 13:244–253
- Fozard JA, Lees M, King JR, Logan BS (2012) Inhibition of quorum sensing in a computational biofilm simulation. *Bio Syst* 109:105–114
- Fradin C, Kretschmar M, Nichterlein T, Gaillardin C, d'Enfert C, Hube B (2003) Stage-specific gene expression of *Candida albicans* in human blood. *Mol Microbiol* 47:1523–1543
- Frey HC, Patil SR (2002) Identification and review of sensitivity analysis methods. *Risk Anal* 22:553–578, Risk analysis: an official publication of the Society for Risk Analysis
- Friesen ML, Saxer G, Travisano M, Doebeli M (2004) Experimental evidence for sympatric ecological

- diversification due to frequency-dependent competition in *Escherichia coli*. *Evolution; Int J Org Evol* 58:245–260
- Fu L, Medico E (2007) FLAME, a novel fuzzy clustering method for the analysis of DNA microarray data. *BMC Bioinformatics* 8:3
- Galagan JE, Calvo SE, Cuomo C, Ma LJ, Wortman JR, Batzoglou S et al (2005) Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. *Nature* 438:1105–1115
- Gamerman D, Lopes HF (2006) Markov chain Monte Carlo: stochastic simulation for Bayesian inference. Taylor & Francis, Boca Raton
- Garber M, Grabherr MG, Guttman M, Trapnell C (2011) Computational methods for transcriptome annotation and quantification using RNA-seq. *Nat Methods* 8:469–477
- Garg A, Mohanram K, Di Cara A, De Micheli G, Xenarios I (2009) Modeling stochasticity and robustness in gene regulatory networks. *Bioinformatics* 25: i101–i109
- Gatenby RA, Vincent TL (2003) An evolutionary model of carcinogenesis. *Cancer Res* 63:6212–6220
- Ghaemmaghami S, Huh WK, Bower K, Howson RW, Belle A, Dephoure N et al (2003) Global analysis of protein expression in yeast. *Nature* 425:737–741
- Giacomantonio CE, Goodhill GJ (2010) A Boolean model of the gene regulatory network underlying mammalian cortical area development. *PLoS Comput Biol* 6
- Giaever G, Chu AM, Ni L, Connelly C, Riles L, Veronneau S et al (2002) Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* 418:387–391
- Glass L, Kauffman SA (1973) The logical analysis of continuous, non-linear biochemical control networks. *J Theor Biol* 39:103–129
- Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H et al (1996) Life with 6000 genes. *Science* 274(546):63–67
- Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabasi AL (2007) The human disease network. *Proc Natl Acad Sci USA* 104:8685–8690
- Gorochowski TE, Matyjaszkiewicz A, Todd T, Oak N, Kowalska K, Reid S et al (2012) BSim: an agent-based tool for modeling bacterial populations in systems and synthetic biology. *PLoS One* 7: e42790
- Grahl N, Puttikamonkul S, Macdonald JM, Gamcsik MP, Ngo LY, Hohl TM et al (2011) In vivo hypoxia and a fungal alcohol dehydrogenase influence the pathogenesis of invasive pulmonary aspergillosis. *PLoS Pathog* 7:e1002145
- Hall BG (2004) Predicting the evolution of antibiotic resistance genes. *Nat Rev Microbiol* 2:430–435
- Hastings WK (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57:97–109
- Haynes BC, Skowrya ML, Spencer SJ, Gish SR, Williams M, Held EP et al (2011) Toward an integrated model of capsule regulation in *Cryptococcus neoformans*. *PLoS Pathog* 7:e1002411
- Heinrich R, Rapoport TA (1974) A linear steady-state treatment of enzymatic chains. General properties, control and effector strength. *Eur J Biochem/FEBS* 42:89–95
- Herman PK (2002) Stationary phase in yeast. *Curr Opin Microbiol* 5:602–607
- Hickman GJ, Hodgman TC (2009) Inference of gene regulatory networks using Boolean-network inference methods. *J Bioinform Comput Biol* 7:1013–1029
- Himmelreich U, Malik R, Kuhn T, Daniel HM, Somorjai RL, Dolenko B et al (2009) Rapid etiological classification of meningitis by NMR spectroscopy based on metabolite profiles and host response. *PLoS One* 4:e5328
- Hnisz D, Tscherner M, Kuchler K (2011) Targeting chromatin in fungal pathogens as a novel therapeutic strategy: histone modification gets infectious. *Epigenomics* 3:129–132
- Hnisz D, Bardet AF, Nobile CJ, Petryshyn A, Glaser W, Schock U et al (2012) A histone deacetylase adjusts transcription kinetics at coding sequences during *Candida albicans* morphogenesis. *PLoS Genet* 8: e1003118
- Homann OR, Dea J, Noble SM, Johnson AD (2009) A phenotypic profile of the *Candida albicans* regulatory network. *PLoS Genet* 5:e1000783
- Hoops S, Sahle S, Gauges R, Lee C, Pahle J, Simus N et al (2006) COPASI—a complex pathway simulator. *Bioinformatics* 22:3067–3074
- Hope WW, Warn PA, Sharp A, Howard S, Kasai M, Louie A et al (2006) Derivation of an in vivo drug exposure breakpoint for flucytosine against *Candida albicans* and Impact of the MIC, growth rate, and resistance genotype on the antifungal effect. *Antimicrob Agents Chemother* 50:3680–3688
- Hoyer LL (2001) The ALS gene family of *Candida albicans*. *Trends Microbiol* 9:176–180
- Hoyer LL, Green CB, Oh SH, Zhao X (2008) Discovering the secrets of the *Candida albicans* agglutinin-like sequence (ALS) gene family—a sticky pursuit. *Med Mycol* 46:1–15
- Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, Kitano H et al (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* 19:524–531
- Huh WK, Falvo JV, Gerke LC, Carroll AS, Howson RW, Weissman JS et al (2003) Global analysis of protein localization in budding yeast. *Nature* 425: 686–691
- Hummert S, Hummert C, Schroter A, Hube B, Schuster S (2010) Game theoretical modelling of survival strategies of *Candida albicans* inside macrophages. *J Theor Biol* 264:312–318
- Inada M, Pleiss JA (2010) Genome-wide approaches to monitor pre-mRNA splicing. *Methods Enzymol* 470:51–75

- Jackson AP, Gamble JA, Yeomans T, Moran GP, Saunders D, Harris D et al (2009) Comparative genomics of the fungal pathogens *Candida dubliniensis* and *Candida albicans*. *Genome Res* 19:2231–2244
- Jaqaman K, Danuser G (2006) Linking data to models: data regression. *Nat Rev Mol Cell Biol* 7:813–819
- Jewett MC, Hofmann G, Nielsen J (2006) Fungal metabolite analysis in genomics and phenomics. *Curr Opin Biotechnol* 17:191–197
- Jones T, Federspiel NA, Chibana H, Dungan J, Kalman S, Magee BB et al (2004) The diploid genome sequence of *Candida albicans*. *Proc Natl Acad Sci USA* 101:7329–7334
- Jones GM, Stalker J, Humphray S, West A, Cox T, Rogers J et al (2008) A systematic library for comprehensive overexpression screens in *Saccharomyces cerevisiae*. *Nat Methods* 5:239–241
- Kacser H, Burns JA (1973) The control of flux. *Symp Soc Exp Biol* 27:65–104
- Karlebach G, Shamir R (2008) Modelling and analysis of gene regulatory networks. *Nat Rev Mol Cell Biol* 9:770–780
- Kaufman A, Kupiec M, Ruppin E (2004) Multi-knockout genetic network analysis: the Rad6 example. *Proc IEEE Comput Syst Bioinform Conf* 2004:332–340
- Kerr G, Ruskin HJ, Crane M, Doolan P (2008) Techniques for clustering gene expression data. *Comput Biol Med* 38:283–293
- Kersey P, Apweiler R (2006) Linking publication, gene and protein data. *Nat Cell Biol* 8:1183–1189
- Kirschner MW (2005) The meaning of systems biology. *Cell* 121:503–504
- Klipp E, Nordlander B, Kruger R, Gennemark P, Hohmann S (2005) Integrative model of the response of yeast to osmotic shock. *Nat Biotechnol* 23:975–982
- Klipp E, Wierling WLC, Kowald A, Lehrach H, Herwig R (2009) Systems biology: a textbook. Wiley-Blackwell, Weinheim
- Kniemeyer O, Schmidt AD, Vodisch M, Wartenberg D, Brakhage AA (2011) Identification of virulence determinants of the human pathogenic fungi *Aspergillus fumigatus* and *Candida albicans* by proteomics. *Int J Med Microbiol* 301:368–377
- Kodzius R, Kojima M, Nishiyori H, Nakamura M, Fukuda S, Tagami M et al (2006) CAGE: cap analysis of gene expression. *Nat Methods* 3:211–222
- Kohl P, Crampin EJ, Quinn TA, Noble D (2010) Systems biology: an approach. *Clin Pharmacol Ther* 88:25–33
- Krause F, Uhlendorf J, Lubitz T, Schulz M, Klipp E, Liebermeister W (2010) Annotation and merging of SBML models with semanticSBML. *Bioinformatics* 26:421–422
- Kuhn C, Wierling C, Kuhn A, Klipp E, Panopoulou G, Lehrach H et al (2009) Monte Carlo analysis of an ODE model of the sea urchin endomesoderm network. *BMC Syst Biol* 3:83
- Kusch H, Engelmann S, Bode R, Albrecht D, Morschhauser J, Hecker M (2008) A proteomic view of *Candida albicans* yeast cell metabolism in exponential and stationary growth phases. *Int J Med Microbiol* 298:291–318
- Lashkari DA, DeRisi JL, McCusker JH, Namath AF, Gentile C, Hwang SY et al (1997) Yeast microarrays for genome wide parallel genetic and gene expression analysis. *Proc Natl Acad Sci USA* 94:13057–13062
- Le Novere N, Finney A, Hucka M, Bhalla US, Campagne F, Collado-Vides J et al (2005) Minimum information requested in the annotation of biochemical models (MIRIAM). *Nat Biotechnol* 23:1509–1515
- Leach MD, Tyc KM, Brown AJ, Klipp E (2012) Modelling the regulation of thermal adaptation in *Candida albicans*, a major fungal pathogen of humans. *PLoS One* 7:e32467
- Lelandais G, Tanty V, Geneix C, Etchebest C, Jacq C, Devaux F (2008) Genome adaptation to chemical stress: clues from comparative transcriptomics in *Saccharomyces cerevisiae* and *Candida glabrata*. *Genome Biol* 9:R164
- Li F, Long T, Lu Y, Ouyang Q, Tang C (2004) The yeast cell-cycle network is robustly designed. *Proc Natl Acad Sci USA* 101:4781–4786
- Li Z, Vizeacoumar FJ, Bahr S, Li J, Warringer J, Vizeacoumar FS et al (2011) Systematic exploration of essential yeast gene function with temperature-sensitive mutants. *Nat Biotechnol* 29:361–367
- Liang S, Fuhrman S, Somogyi R (1998) Reveal, a general reverse engineering algorithm for inference of genetic network architectures. In: Pacific symposium on biocomputing. Pacific symposium on biocomputing, pp 18–29
- Liu OW, Chun CD, Chow ED, Chen C, Madhani HD, Noble SM (2008) Systematic genetic analysis of virulence in the human fungal pathogen *Cryptococcus neoformans*. *Cell* 135:174–188
- Loftus BJ, Fung E, Roncaglia P, Rowley D, Amedeo P, Bruno D et al (2005) The genome of the basidiomycetous yeast and human pathogen *Cryptococcus neoformans*. *Science* 307:1321–1324
- Lorenz MC, Bender JA, Fink GR (2004) Transcriptional response of *Candida albicans* upon internalization by macrophages. *Eukaryot Cell* 3:1076–1087
- Madeira SC, Oliveira AL (2004) Biclustering algorithms for biological data analysis: a survey. *IEEE/ACM Trans Comput Biol Bioinf* 1:24–45
- Marcet-Houben M, Gabaldon T (2009) The tree versus the forest: the fungal tree of life and the topological diversity within the yeast phylome. *PLoS One* 4:e4357
- Margolin AA, Wang K, Lim WK, Kustagi M, Nemenman I, Califano A (2006) Reverse engineering cellular networks. *Nat Protoc* 1:662–671
- Marinach C, Alanio A, Palous M, Kwasek S, Fekkar A, Brossas JY et al (2009) MALDI-TOF MS-based drug susceptibility testing of pathogens: the

- example of *Candida albicans* and fluconazole. *Proteomics* 9:4627–4631
- Marino S, Hogue IB, Ray CJ, Kirschner DE (2008) A methodology for performing global uncertainty and sensitivity analysis in systems biology. *J Theor Biol* 254:178–196
- Maskos U, Southern EM (1992) Oligonucleotide hybridizations on glass supports: a novel linker for oligonucleotide synthesis and hybridization properties of oligonucleotides synthesised in situ. *Nucleic Acids Res* 20:1679–1684
- Metzker ML (2010) Sequencing technologies - the next generation. *Nat Rev Genet* 11:31–46
- Mitri S, Xavier JB, Foster KR (2011) Social evolution in multispecies biofilms. *Proc Natl Acad Sci USA* 108 (Suppl 2):10839–10846
- Moles CG, Mendes P, Banga JR (2003) Parameter estimation in biochemical pathways: a comparison of global optimization methods. *Genome Res* 13:2467–2474
- Moretti S, Fioravante P, Bonassi S (2007) The class of microarray games and the relevance index for genes. *TOP* 15:256–280
- Moretti S, van Leeuwen D, Gmuender H, Bonassi S, van Delft J, Kleinjans J et al (2008) Combining Shapley value and statistics to the analysis of gene expression data in children exposed to air pollution. *BMC Bioinformatics* 9:361
- Mura I, Csikasz-Nagy A (2008) Stochastic Petri Net extension of a yeast cell cycle model. *J Theor Biol* 254:850–860
- Mustacchi R, Hohmann S, Nielsen J (2006) Yeast systems biology to unravel the network of life. *Yeast* 23:227–238
- Netotea S, Bertani I, Steindler L, Kerenyi A, Venturi V, Pongor S (2009) A simple model for the early events of quorum sensing in *Pseudomonas aeruginosa*: modeling bacterial swarming as the movement of an “activation zone”. *Biol Direct* 4:6
- Nielsen J (2003) It is all about metabolic fluxes. *J Bacteriol* 185:7031–7035
- Nierman WC, Pain A, Anderson MJ, Wortman JR, Kim HS, Arroyo J et al (2005) Genomic sequence of the pathogenic and allergenic filamentous fungus *Aspergillus fumigatus*. *Nature* 438:1151–1156
- Niittylae T, Chaudhuri B, Sauer U, Frommer WB (2009) Comparison of quantitative metabolite imaging tools and carbon-13 techniques for fluxomics. *Methods Mol Biol* 553:355–372
- Nobile CJ, Fox EP, Nett JE, Sorrells TR, Mitrovich QM, Hernday AD et al (2012) A recently evolved transcriptional network controls biofilm development in *Candida albicans*. *Cell* 148:126–138
- Noble SM, French S, Kohn LA, Chen V, Johnson AD (2010) Systematic screens of a *Candida albicans* homozygous deletion library decouple morphogenetic switching and pathogenicity. *Nat Genet* 42:590–598
- Novak B, Tyson J (2002) Cell cycle controls. Springer, Berlin, pp 261–284
- Nowak MA, Sasaki A, Taylor C, Fudenberg D (2004) Emergence of cooperation and evolutionary stability in finite populations. *Nature* 428:646–650
- O’Farrell PH (1975) High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* 250:4007–4021
- O’Malley MA, Dupre J (2005) Fundamental issues in systems biology. *Bioessays* 27:1270–1276, *BioEssays: news and reviews in molecular, cellular and developmental biology*
- Oosthuizen JL, Gomez P, Ruan J, Hackett TL, Moore MM, Knight DA et al (2011) Dual organism transcriptomics of airway epithelial cells interacting with conidia of *Aspergillus fumigatus*. *PLoS One* 6:e20527
- Opel M, Lando D, Bonilla C, Treweek SC, Boukaba A, Walfridsson J et al (2007) Genome-wide studies of histone demethylation catalysed by the fission yeast homologues of mammalian LSD1. *PLoS One* 2:e386
- Orth JD, Thiele I, Palsson BO (2010) What is flux balance analysis? *Nat Biotechnol* 28:245–248
- Oshlack A, Robinson MD, Young MD (2010) From RNA-seq reads to differential expression results. *Genome Biol* 11:220
- Pappalardo F, Martinez Forero I, Pennisi M, Palazon A, Melero I, Motta S (2011) SimB16: modeling induced immune system response against B16-melanoma. *PLoS One* 6:e26523
- Peng M, Taouatas N, Cappadona S, van Breukelen B, Mohammed S, Scholten A et al (2012) Protease bias in absolute protein quantitation. *Nat Methods* 9:524–525
- Petranovic D, Nielsen J (2008) Can yeast systems biology contribute to the understanding of human disease? *Trends Biotechnol* 26:584–590
- Pfeiffer T, Schuster S (2005) Game-theoretical approaches to studying the evolution of biochemical systems. *Trends Biochem Sci* 30:20–25
- Picotti P, Clement-Ziza M, Lam H, Campbell DS, Schmidt A, Deutsch EW et al (2013) A complete mass-spectrometric map of the yeast proteome applied to quantitative trait analysis. *Nature* 494:266–270
- Pogson M, Smallwood R, Qwarnstrom E, Holcombe M (2006) Formal agent-based modelling of intracellular chemical interactions. *Bio Syst* 85:37–45
- Pogson M, Holcombe M, Smallwood R, Qwarnstrom E (2008) Introducing spatial information into predictive NF-kappaB modelling—an agent-based approach. *PLoS One* 3:e2367
- Price ND, Papin JA, Schilling CH, Palsson BO (2003) Genome-scale microbial in silico models: the constraints-based approach. *Trends Biotechnol* 21:162–169
- Reinartz J, Bruyns E, Lin JZ, Burcham T, Brenner S, Bowen B et al (2002) Massively parallel signature

- sequencing (MPSS) as a tool for in-depth quantitative gene expression profiling in all organisms. *Brief Funct Genomic Proteomic* 1:95–104
- Reiss DJ, Baliga NS, Bonneau R (2006) Integrated biclustering of heterogeneous genome-wide datasets for the inference of global regulatory networks. *BMC Bioinformatics* 7:280
- Ringner M (2008) What is principal component analysis? *Nat Biotechnol* 26:303–304
- Roguev A, Bandyopadhyay S, Zofall M, Zhang K, Fischer T, Collins SR et al (2008) Conservation and rewiring of functional modules revealed by an epistasis map in fission yeast. *Science* 322:405–410
- Royce TE, Rozowsky JS, Gerstein MB (2007) Toward a universal microarray: prediction of gene expression through nearest-neighbor probe sequence identification. *Nucleic Acids Res* 35:e99
- Rubin-Bejerano I, Fraser I, Grisafi P, Fink GR (2003) Phagocytosis by neutrophils induces an amino acid deprivation response in *Saccharomyces cerevisiae* and *Candida albicans*. *Proc Natl Acad Sci USA* 100:11007–11012
- Ruppin E, Papin JA, de Figueiredo LF, Schuster S (2010) Metabolic reconstruction, constraint-based analysis and game theory to probe genome-scale metabolic networks. *Curr Opin Biotechnol* 21:502–510
- Santamaria R, Rizzetto L, Bromley M, Zelante T, Lee W, Cavalieri D et al (2011) Systems biology of infectious diseases: a focus on fungal infections. *Immunobiology* 216:1212–1227
- Schuster S, Kreft JU, Brenner N, Wessely F, Theissen G, Ruppin E et al (2010) Cooperation and cheating in microbial coenzyme production—theoretical analysis for biotechnological applications. *Biotechnol J* 5:751–758
- Schwarz Müller T, Ma B, Hiller E, Istel F, Tscherner M, Brunke S, Ames L, Firon A, Green B, Cabral V, Marcet-Houben M, Jacobsen ID, Quintin J, Seider K, Frohner I, Glaser W, Jungwirth H, d'Enfert C, Ferrandon D, Gabaldón T, Hube B, Rupp S, Cormack B, Haynes K, Kuchler K (in preparation) Systematic phenotyping of a genome-scale *Candida glabrata* deletion collection reveals novel antifungal tolerance genes
- Seal JB, Alverdy JC, Zaborina O, An G (2011) Agent-based dynamic knowledge representation of *Pseudomonas aeruginosa* virulence activation in the stressed gut: towards characterizing host-pathogen interactions in gut-derived sepsis. *Theor Biol Med Model* 8:33
- Segovia-Juarez JL, Ganguli S, Kirschner D (2004) Identifying control mechanisms of granuloma formation during *M. tuberculosis* infection using an agent-based model. *J Theor Biol* 231:357–376
- Sellam A, Hogues H, Askew C, Tebbji F, van Het Hoog M, Lavoie H et al (2010) Experimental annotation of the human pathogen *Candida albicans* coding and noncoding transcribed regions using high-resolution tiling arrays. *Genome Biol* 11:R71
- Shmulevich I, Dougherty ER, Kim S, Zhang W (2002) Probabilistic Boolean networks: a rule-based uncertainty model for gene regulatory networks. *Bioinformatics* 18:261–274
- Smedsgaard J, Nielsen J (2005) Metabolite profiling of fungi and yeast: from phenotype to metabolome by MS and informatics. *J Exp Bot* 56:273–286
- Sopko R, Huang D, Preston N, Chua G, Papp B, Kafadar K et al (2006) Mapping pathways and phenotypes by systematic gene overexpression. *Mol Cell* 21:319–330
- Sorgo AG, Heilmann CJ, Dekker HL, Brul S, de Koster CG, Klis FM (2010) Mass spectrometric analysis of the secretome of *Candida albicans*. *Yeast* 27:661–672
- Sorgo AG, Heilmann CJ, Dekker HL, Bekker M, Brul S, de Koster CG et al (2011) Effects of fluconazole on the secretome, the wall proteome, and wall integrity of the clinical fungus *Candida albicans*. *Eukaryot Cell* 10:1071–1081
- Stastna M, Van Eyk JE (2012a) Improved and emerging gel-free separation and detection methods for proteomics. *Proteomics* 12:2902–2903
- Stastna M, Van Eyk JE (2012b) Analysis of protein isoforms: can we do it better? *Proteomics* 12:2937–2948
- Steinhoff C, Vingron M (2006) Normalization and quantification of differential expression in gene expression microarrays. *Brief Bioinform* 7:166–177
- Sun Z, Zhu Y (2012) Systematic comparison of RNA-seq normalization methods using measurement error models. *Bioinformatics* 28:2584–2591
- Tabb DL, Vega-Montoto L, Rudnick PA, Variyath AM, Ham AJ, Bunk DM et al (2010) Repeatability and reproducibility in proteomic identifications by liquid chromatography-tandem mass spectrometry. *J Proteome Res* 9:761–776
- Tamayo P, Slonim D, Mesirov J, Zhu Q, Kitareewan S, Dmitrovsky E et al (1999) Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. *Proc Natl Acad Sci USA* 96:2907–2912
- Taylor CF, Paton NW, Lilley KS, Binz PA, Julian RK Jr, Jones AR et al (2007) The minimum information about a proteomics experiment (MIAPE). *Nat Biotechnol* 25:887–893
- Taylor CF, Field D, Sansone SA, Aerts J, Apweiler R, Ashburner M et al (2008) Promoting coherent minimum reporting guidelines for biological and biomedical investigations: the MIBBI project. *Nat Biotechnol* 26:889–896
- Teutschbein J, Albrecht D, Potsch M, Guthke R, Aïmanianda V, Clavaud C et al (2010) Proteome profiling and functional classification of intracellular proteins from conidia of the human-pathogenic mold *Aspergillus fumigatus*. *J Proteome Res* 9:3427–3442

- Tierney L, Linde J, Muller S, Brunke S, Molina JC, Hube B et al (2012) An interspecies regulatory network inferred from simultaneous RNA-seq of *Candida albicans* invading innate immune cells. *Front Microbiol* 3:85
- Tokarski C, Hummert S, Mech F, Figge MT, Germerodt S, Schroeter A et al (2012) Agent-based modeling approach of immune defense against spores of opportunistic human pathogenic fungi. *Front Microbiol* 3:129
- Tomar N, Choudhury O, Chakrabarty A, De RK (2013) An integrated pathway system modeling of *Saccharomyces cerevisiae* HOG pathway: a Petri Net based approach. *Mol Biol Rep* 40(2):1103–1125
- Tyc KM, Klipp E (2011) Modeling dissemination of pathogenic fungi within a host: a cartoon for the interactions of two complex systems. *J Comput Sci Syst Biol* S1:001
- Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR et al (2000) A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* 403:623–627
- van het Hoog M, Rast TJ, Martchenko M, Grindle S, Dignard D, Hogues H et al (2007) Assembly of the *Candida albicans* genome into sixteen supercontigs aligned on the eight chromosomes. *Genome Biol* 8:R52
- Velculescu VE, Zhang L, Vogelstein B, Kinzler KW (1995) Serial analysis of gene expression. *Science* 270:484–487
- Villas-Boas SG, Moxley JF, Akesson M, Stephanopoulos G, Nielsen J (2005) High-throughput metabolic state analysis: the missing link in integrated functional genomics of yeasts. *Biochem J* 388:669–677
- von Bertalanffy L (1969) In: *General systems theory: foundations, development, applications*. Revised (ed) Vol. pp. George Braziller, New York
- Wang Z, Gerstein M, Snyder M (2009) RNA-seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 10:57–63
- Wang RS, Saadatpour A, Albert R (2012) Boolean modeling in systems biology: an overview of methodology and applications. *Phys Biol* 9:055001
- Weiner A, Chen HV, Liu CL, Rahat A, Klien A, Soares L et al (2012) Systematic dissection of roles for chromatin regulators in a yeast stress response. *PLoS Biol* 10:e1001369
- Weinstein JN (2001) Searching for pharmacogenomic markers: the synergy between omic and hypothesis-driven research. *Dis Markers* 17:77–88
- Wierling C, Kuhn A, Hache H, Daskalaki A, Maschke-Dutz E, Psycheva S et al (2012) Prediction in the face of uncertainty: a Monte Carlo-based approach for systems biology of cancer treatment. *Mutat Res* 746:163–170
- Wilkins MR, Pasquali C, Appel RD, Ou K, Golaz O, Sanchez JC et al (1996) From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Biotechnology* 14:61–65
- Winzler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B et al (1999) Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* 285:901–906
- Yan Z, Costanzo M, Heisler LE, Paw J, Kaper F, Andrews BJ et al (2008) Yeast barcoders: a chemogenomic application of a universal donor-strain collection carrying bar-code identifiers. *Nat Methods* 5:719–725
- Yates JR 3rd, Carmack E, Hays L, Link AJ, Eng JK (1999) Automated protein identification using microcolumn liquid chromatography-tandem mass spectrometry. *Methods Mol Biol* 112:553–569
- Yus E, Maier T, Michalodimitrakis K, van Noort V, Yamada T, Chen WH et al (2009) Impact of genome reduction on bacterial metabolism and its regulation. *Science* 326:1263–1268
- Zamboni N, Sauer U (2009) Novel biological insights through metabolomics and ¹³C-flux analysis. *Curr Opin Microbiol* 12:553–558
- Zhong Q, Simonis N, Li QR, Charloteaux B, Heuze F, Klitgord N et al (2009) Edgetic perturbation models of human inherited disorders. *Mol Syst Biol* 5:321