Chapter 2
Current Progress in Static and Dynamic Modeling of Biological Networks

Bernie J. Daigle, Jr., Balaji S. Srinivasan, Jason A. Flannick, Antal F. Novak, and Serafim Batzoglou

Abstract The relentless advance of biochemistry has enabled us to take apart biological systems with ever more fine-grained and precise instruments. The fruits of this dissection are millions of measurements of base pairs and biochemical concentrations. Yet to make sense of these numbers, we need to reverse our dissection by putting the system back together on the computer. This first step in this process is reconstructing molecular anatomy through *static modeling*, the determination of which pieces (DNA, RNA, protein, and metabolite) is present, and how they are related (e.g., regulator, target, inhibitor, cofactor). Given this broad outline of component connectivity, we may then attempt to reconstruct molecular physiology via *dynamic modeling*, computer simulations that model when cellular events occur (ODE), where they occur (PDE), and how frequently they recur (SDE). In this review we discuss techniques for both of these modeling paradigms, illustrating each by reference to important recent papers.

Keywords Biological networks · Computer simulation · Dynamic modeling · Static modeling

2.1 Introduction

The term "post-genomic era" became a cliche even before the human genome was sequenced, but it has a definite meaning. It refers to the refocusing of effort on tasks that were insurmountable without the genome as platform, such as the construction of hybridization probes for every human gene (Schena et al. 1996) or the phenotyping of knockout strains for every yeast ORF (Winzeler et al. 1999). Many different kinds of these genome-scale data sets are now available (Collins et al. 2007; Foster

B.S. Srinivasan (✉)
Departments of Computer Science and Statistics, Stanford University, Stanford, CA, USA
e-mail: balajis@stanford.edu

B.J. Daigle Jr., and B.S. Srinivasan have contributed equally to this work.

S. Choi (ed.), *Systems Biology for Signaling Networks*, Systems Biology 1,
DOI 10.1007/978-1-4419-5797-9_2,

et al. 2006; Gavin et al. 2006; Kim et al. 2005; Krogan et al. 2006; Lamb et al. 2006; Sachs et al. 2005), and each analysis tells the same story: the components of biological systems are not free-floating parts, but are organized into functional modules (Hartwell et al. 1999).

Systems biology is the science of quantitatively defining and analyzing these modules (Bornholdt 2005) and can be divided into two broad areas: static modeling of an organism's interactome (Section 2.2) and dynamic modeling of a biological system's kinetics, spatial structure, or stochastic variation (Section 2.3).

In general, static models tend to be broader and coarser in scope, often encompassing the entire interactome, while dynamic models usually focus on the details of a single subsystem, such as chemotaxis (Alon et al. 1999), lysogeny (Arkin et al. 1998), or morphogenesis (Igoshin et al. 2004a,b). Static modeling is less demanding from an experimental perspective, as just about any assay on a population of cells will prove informative. By contrast, deterministic dynamic models require temporally and sometimes spatially (Meinhardt and de Boer 2001) resolved data, and stochastic dynamic models require even more data in the form of population ensembles. In this review, we discuss both modeling strategies with an eye toward describing statistical considerations and summarizing recent successes.

2.2 Static Models of Biological Networks

Static modeling is best conceptualized as the computerized reconstruction of molecular anatomy. In much the same way that macroscopic anatomy tells us that the shinbone is connected to the kneebone, molecular anatomy tells us which molecules interact with each other.

Yet the situation at the molecular level is complicated by the fact that we cannot yet "see" the molecular components of a cell at the same resolution that a pathologist can observe the bones and muscles of a cadaver. Our approach is rather more like that of an archaeologist who discovers many piles of bones in different configurations and must statistically reason that shinbones and kneebones are likely to be functionally related, as they are (1) often found near each other, (2) usually present together in different species, and (3) more correlated in size than random pairs of bones.

This concept of statistically inferring static relationships via "gnilt by association" is one of the core ideas behind static modeling. We visually represent these inferred static relationships by a *network*. Nodes of this network correspond to components of the system and edges to relationships between components. Different kinds of static models are usefully distinguished by the number and type of nodes and edges which are present. As a general rule, larger networks for more complex organisms require more data to reconstruct.

In this section, we review methods for the representation and inference of static network models from multiple data sources. We describe a common Bayesian formulation which unifies the steps of network integration and experimental validation. By analogy to the concept of a reference genome assembly (Lander et al. 2001;

Venter et al. 2001), we then describe how recent large-scale efforts at network determination, such as the recent connectivity map (Lamb et al. 2006) and the proposed human interactome project (Ideker and Valencia 2006), have led naturally to the concept of ontologically labeled, richly typed reference networks. We conclude by discussing methods for network alignment, network visualization, and network-guided experimental prioritization.

2.2.1 Advantages of Static Models

Because static modeling is about determining which elements are present (nodes) and how they are interconnected (edges), it is a basic prerequisite for any kind of systems biological analysis. As one example, determining whether two bacteria can metabolize the same sets of compounds requires an enumeration of their functional modules, roughly corresponding to the evolutionarily conserved subgraphs in their respective static models. As another example, identifying proteins which are essential for cellular function can be greatly aided by knowledge of which proteins are central in static network models. In particular, static models are essential starting points for more complex dynamic modeling strategies.

2.2.2 Limitations of Static Models

Perhaps the most obvious limitation of a static model is that it is in fact static: it does not incorporate temporal, spatial, or conditional information except indirectly. In particular, less detailed static models may give little information about how different nodes talk to each other. For example, low-resolution models that predict solely whether two proteins "interact" with some probability are useful for generating hypotheses, but give little mechanistic insight as to whether they are related by physical contact, presence in the same pathway, or regulation of the same genes. These limitations can be partially overcome by including more types of edges, though there are fundamental limitations on the level of conditional detail (Fig. 2.1) possible in a static network.

2.2.3 Specific Tasks Associated with Static Modeling

We can order the process of static modeling into five sequential tasks:

1. **Determine desired network detail**. The first step in static modeling is to determine the scope and detail of the network reconstruction. Put simply, how many nodes and edges are desired, and what are their types? The goal here is to quantitatively parametrize the network by a response variable. For example, this could be an $N^2 \times 1$ vector of boolean connectivities on N^2 edges (as per L in Fig. 2.2)

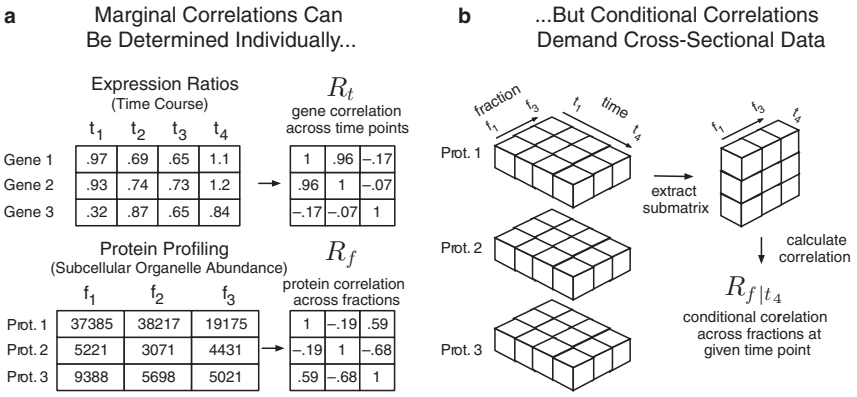


Fig. 2.1 Data availability constrains network detail. (**a**) Given a cell cycle time course of gene expression measurements, we can determine which genes are temporally coexpressed (Spellman et al. 1998). Similarly, from protein correlation profiling (Foster et al. 2006), we can determine which proteins are abundant in the same subcellular organelles, and thereby derive a rough measure of colocalization. (**b**) Suppose, however, that we wish to determine whether a given protein pair is colocalized at a particular time in the cell cycle. To calculate this conditional correlation we must (1) sharply increase the number of data points in our experiment and (2) collect both kinds of data on the same object at the same time. This may be difficult or impossible to do experimentally; for example, the methods for determining protein abundance across organelles are very different from those for determining an mRNA abundance time series. As more kinds of variables are incorporated (chemical stimuli, genetic background, etc.) the requisite number of data points increases exponentially. These constraints fundamentally limit the extent to which conditional interactions can be probed with independently collected data sets. Reproduced from (Srinivasan et al. 2007) with permission from Oxford University Press

or an $N \times 1$ vector of node properties (as per Y in Fig. 2.3). Ideally some high-resolution data on Y or L is already available, a so-called *training set*. This data could come from well-established individual publications and/or from low throughput, expensive experiments.

2. **Enumerate input data sources**. The next step is to work backward to determine which input variable could potentially predict the desired network properties. Input variables can usually be either $N \times 1$ vectors that give data on each node or $N^2 \times 1$ vectors that predict edge data (X and E, respectively, in Fig. 2.3).
3. **Network reconstruction**. Given predictors (X, E) and partial training data on (Y, L), we can use machine learning to predict the remaining elements of (Y, L) (Figs. 2.2 and 2.3). The result is a static network model $(\hat{Y}, \hat{L})$ with information on the nodes and edges of the biological system under consideration. Here the "hat" denotes the fact that these values are estimates and of lower reliability than the "gold-standard" training set.
4. **Experimental confirmation**. In the ideal scenario, the properties of the predicted network are then experimentally tested. The gold standard is to make new high-resolution measurements on (Y, L) using the same methods used to assemble the training set and to compare these experimental measurements to

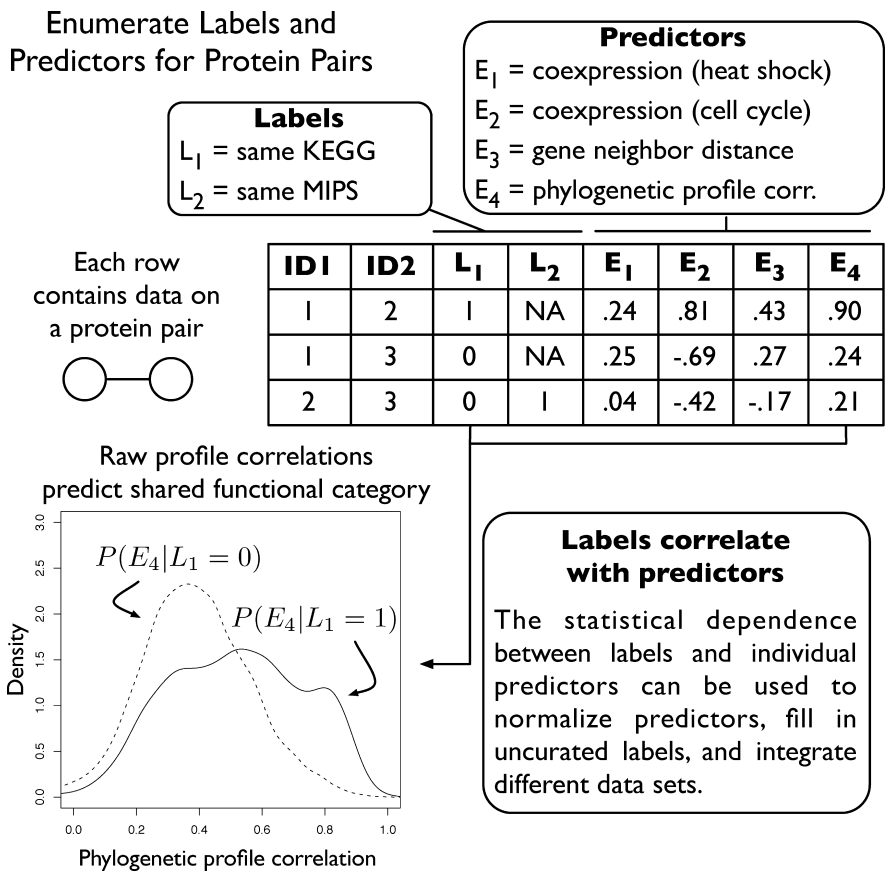


Fig. 2.2 Enumerating labels and predictors for data integration. For each protein pair, we can compute labels and predictors. At the *top* of the figure, two kinds of labels and four predictors have been tabulated for each pair of proteins; given N proteins, this table will have $N(N-1)/2$ rows. Labels are directly useful to humans while predictors represent raw experimental data. Importantly, many labels correlate with predictors. For example, calculating conditional density estimates (*lower left*) for the phylogenetic profile correlation over all pairs in *Mycoplasma genitalium* shows that highly coinherited pairs are likely to functionally interact in the same KEGG category (Srinivasan et al. 2006). This statistical dependence can be used to put predictors on the same scale, by normalizing them in terms of their ability to recapitulate functional interactions. It can also be used to fill in uncurated labels and integrate different data types (Fig. 2.3). Reproduced from Srinivasan et al. (2007) with permission from Oxford University Press

the predictions $(\hat{Y}, \hat{L})$. If the predictions match experiment sufficiently well, we can replace potentially expensive high-resolution measurements on (Y, L) with cheaper bulk measurements of the relevant (X, E) variables.

5. **Network applications**. Given an experimentally reliable static network, we can then proceed to further applications, such as comparisons of networks across species and conditions (network alignment) and network-guided experimental prioritization.

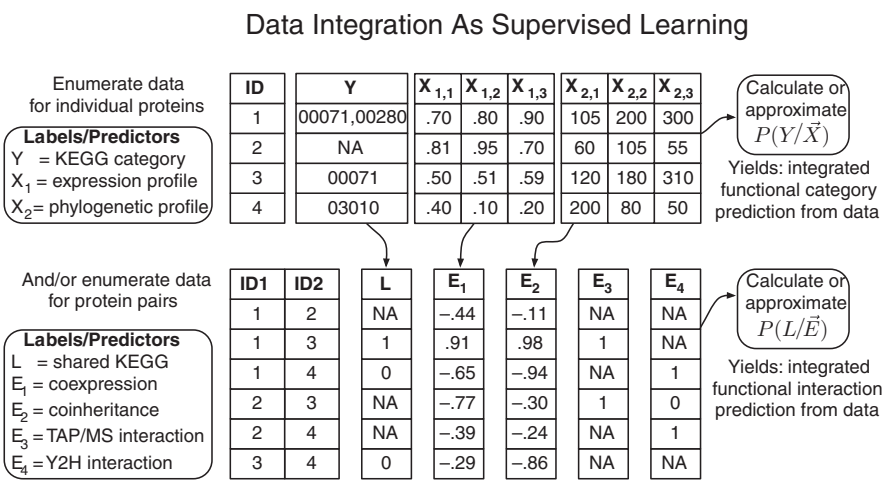


Fig. 2.3 Data integration as supervised learning. For each biological object, we tabulate labels and predictors as in Fig. 2.2. Rather than comparing predictors in terms of their correlation with the label, we use all the predictors at the same time to estimate the label. For example, if we do this for the specific biological object of individual proteins, we can obtain an integrative prediction of protein function. If instead we do this for pairs of proteins, we can obtain an integrative prediction of protein interaction. Note that some of the columns in the pair table are only defined for pairs (in this case, the TAP/MS and Y2H data) while other quantities can be computed from the protein table. Note also that for statistical reasons, the interaction prediction problem can be easier than the function prediction problem. In the former case, we have a multiclass classification problem with only a few thousand data points, while in the latter case we have a binary classification problem with millions of data points (Hastie et al. 2001). Importantly, the supervised learning framework can be applied to many other kinds of biological objects besides proteins and protein pairs. Reproduced from (Srinivasan et al. 2007) with permission from Oxford University Press

The remainder of the chapter is split into four parts: a summary of data sources used for static modeling, an overview of algorithms for network reconstruction, a discussion of network representations, and a survey of applications of static networks.

2.2.4 Data for Static Modeling

2.2.4.1 Data Types and Sources

The goal of static modeling is to infer the properties of nodes and edges for a given biological system, which is often the entire interactome of a single organism. Subgraphs of tightly interconnected objects in these networks represent functional modules (Barabasi and Oltvai 2004). Some of these networks are obtained from edge predictors *E* in Fig. 2.3, in that they come from direct measurements of pairwise interactions (Zhu et al. 2007), including physical (Gavin et al. 2006; Krogan

et al. 2006), signaling (Pokholok et al. 2006; Ptacek and Snyder 2006), transcriptional (Davidson et al. 2002; Wei et al. 2006), metabolic (Covert et al. 2004), and epistatic (Collins et al. 2007; Schuldiner et al. 2005; Tong et al. 2001) networks. Other networks have their connectivity inferred indirectly through measurements on node predictors *X*, such as coexpression under the same conditions (Lamb et al. 2006), in the same tissues (Chen et al. 2006), or at the same time points (Laub et al. 2000; Spellman et al. 1998); coinheritance in the same species (Pellegrini et al. 1999; Srinivasan et al. 2005); collocation on chromosomes (Overbeek et al. 1999); coevolution of residues (Pazos et al. 2005); or shared mutant phenotype (Dudley et al. 2005). These indirect networks are constructed by using variation along one dimension (time, space, environmental perturbation, etc.) to inform the construction of the global network. For example, proteins that are abundant in the same subcellular organelles (Foster et al. 2006) are likely to functionally interact, as are genes that are expressed at the same time (Spellman et al. 1998); such interacting sets represent subgraphs in the global interaction network.

Given that hundreds of these large-scale data sets are now available, it has become essential to consult meta-databases. Among the most useful are Pathguide (Bader et al. 2006), BiowareDB (Matthiessen 2003), BioGRID (Stark et al. 2006), the yearly Nucleic Acids Research Database (Galperin and Cochrane 2009) and Web Server (Benson 2009) issues, and a recent compilation of more than 150 publicly available functional genomic resources (Ng et al. 2006). As a general rule, it is probably best to limit one's use of raw data to data sets curated by the major databases (NCBI, EBI, DDBJ, UCSC, etc.). Otherwise a great deal of time will be spent mapping identifiers and parsing various data formats.

2.2.4.2 Data Limits Static Model Complexity

As we shall see, the advantage of static modeling is that it can incorporate data sets compiled by a number of investigators at different times and under different conditions. However, this very advantage also imposes fundamental limitations on static model complexity.

For example, consider the problem of determining a conditional network of interactions or correlations in each subcellular organelle. As Fig. 2.1 shows, this seemingly simple request dramatically increases the amount of data that must be simultaneously collected. Moreover, in many cases the extra resolution is simply unavailable with current experimental techniques. Microfluidic automation of basic laboratory procedures (Demello 2006; Hansen and Quake 2003) may make such cross-sectional measurements feasible in the future, but with few exceptions, such as the high-throughput construction and characterization of deletion strains (Giaever et al. 2002), fine-grained conditional data is usually unavailable. Even in large-scale studies, data is usually collected on only one variable at a time.

Thus, the limitations of the available data tend to force us toward a static lowest common denominator map of interactions for most organisms, averaged over time, space, perturbation, and other variables. All is not lost, however, as this static network is still a significant conceptual leap beyond the raw genome sequence of an

organism. Moreover, variation of different kinds (e.g., upregulation of genes or spatial localization of proteins) can be visualized by superimposing tracks and layouts upon such static networks (Hu et al. 2007; Shannon et al. 2003) in the same way we view gene and motif tracks upon a genome assembly (Kuhn et al. 2007).

2.2.5 Network Reconstruction

2.2.5.1 Labels vs. Predictors

For the purposes of data integration, a useful data set is one that provides measurements on at least one type of biological object, such as genes, proteins, or protein pairs (Fig. 2.2). Such data sets can be divided into two broad categories: labels and predictors. Predictors, such as expression ratio measurements on a gene (Schena et al. 1996) or phylogenetic profiles of a protein (Pellegrini et al. 1999), are often "dense" in that they are available for most instances of a biological object and are acquired in a high-throughput way. For example, because most genes are present on standard microarrays, expression profiles are available for most genes (modulo missing values). In contrast, labels such as GO consortium gene annotations (Ashburner et al. 2000) or phosphorylation interactions culled from the literature (Saric et al. 2006) tend to be sparse and of high quality. One of the most important discoveries (Jansen et al. 2003) in functional genomics is that these curated labels, which represent directly useful information, can be statistically predicted from combinations of uncurated predictors.

2.2.5.2 Early Methods for Clustering and Integration

The road to this discovery began with early attempts at unsupervised integration and clustering. When the first microarray data sets became available, dozens of different algorithms for unsupervised clustering of these data sets were published (Altman and Raychaudhuri 2001; Sherlock 2000). These techniques were also applied to other data sets, such as phylogenetic profiling (Pellegrini et al. 1999). While individual clusters of genes were sometimes experimentally validated (Srinivasan et al. 2005; Stuart et al. 2003), it was difficult to assess the extent to which any given clustering reflected the true modules of the organism. Given the fuzziness of the module concept, the fact that genes (and other biological objects) can belong to more than one module, and the often conditional nature of intra-module interactions, it was not clear whether the concept of a true set of modules was even a useful one.

This problem became more pronounced when investigators began to combine interaction networks inferred from different assays, which in turn had apparently different modular structures. The first attempts (Tong et al. 2002) applied arbitrary thresholds to the interactions derived from different assays and used the union or intersection of these sets as an integrated network. In some cases, such as large-scale yeast two hybrid data, the intersection was essentially the null set (Ito et al. 2001). While the goal of combining different assays to reduce noise was a step in the

right direction, the problem was that no clear method for weighting the confidence of different assays was available. As with unsupervised clustering, the underlying issue here was the lack of a true set of curated modules to benchmark different assays against.

2.2.5.3 Data Integration by Supervised Learning

Supervised Normalization

The solution (Jansen et al. 2003; Lee et al. 2004; Lu et al. 2005a; Srinivasan et al. 2006; Tanay et al. 2004; Troyanskaya et al. 2003; Wong et al. 2004) was to build a gold standard to compare different kinds of predictors. In general, different kinds of gold standards can be built from different labels; while a colocalization gold standard can be built from MIPS (Mewes et al. 1999), a functional interaction gold standard can be generated from EcoCyc (Karp et al. 2002), Reactome (Vastrik et al. 2007), GO (Harris et al. 2004), or KEGG (Kanehisa et al. 2006). Negative examples can then be easily generated via random permutation of positive labels (Ben-Hur and Noble 2006). Though simple, the permutation-based approach to generating negative examples has been shown to be superior to selecting a statistically biased subset of negative examples, such as proteins known to be in different subcellular localizations (Ben-Hur and Noble 2006).

Given this gold standard, a useful predictor will separate positive from negative examples. This observed statistical separation can then be converted into a posterior probability by applying Bayes' Rule (Srinivasan et al. 2006), allowing different predictors (uncurated data) to be compared in terms of their ability to recapitulate known biological labels (curated data). In the specific case of protein interaction prediction, a good predictor will recapitulate known labels by separating interacting protein pairs from non-interacting pairs (Figs. 2.2 and 2.3).

Detection of Corrupted Data

One important application of this result is screening microarray experiments for corrupted data (Srinivasan et al. 2006). In addition to a battery of internal consistency checks (Irizarry et al. 2005; Woo et al. 2004), a series of expression measurements can also be used to calculate a correlation matrix, which can then be compared to a gold standard. If coexpression correlations separate positive and negative training examples as in the lower left panel of Fig. 2.2, the data set contains at least some signal; if no separation is observed, problems may have occurred with some hybridizations.

Another important application is identifying which kinds of data may be systematically unreliable; for example, interactions from large-scale yeast two hybrid (Y2H) studies appear to be uncorrelated with any of several gold standards (Qi et al. 2006). This matches other results that indicate that the properties of hubs (Bloom and Adami 2003) and degree distributions (Deeds et al. 2006) in Y2H networks may be artifactual and may explain the low overlap of independently collected Y2H data

sets with each other (Gandhi et al. 2006; Goll and Uetz 2006; Hart et al. 2006) and with literature-curated interactions (Rual et al. 2005). Moreover, the generally low correlation of Y2H interactions with curated data stands in contrast to TAP/MS-derived interactions and most other kinds of functional genomic data, including expression arrays (Qi et al. 2006). The ability to perform such comparisons is one of the primary advantages of a gold standard.

Supervised Integration

In addition to allowing comparison of different predictors, a gold standard also enables us to perform data integration. In the context of protein interaction prediction, an array of association predictors is the input to a binary classifier function, which returns the integrated probability that two proteins are linked in the sense stipulated by the gold standard (Fig. 2.3). When this binary classifier function is applied to predict interaction probabilities for all protein pairs in a genome, the result is an integrated probabilistic protein interaction network. Variants of this approach have been used to predict functional associations (Jansen et al. 2003; Lee et al. 2004; Srinivasan et al. 2006), physical contacts (Qi et al. 2006), synthetically lethal genetic interactions (Wong et al. 2004), and colocalizations (Qi et al. 2006; Jansen et al. 2002).

Importantly, this supervised learning framework for data integration is not limited to interaction prediction and has also been applied to direct prediction of protein function (Han et al. 2004; Lu et al. 2005b) and transcription factor/DNA binding (Beyer et al. 2006). In fact, the applications of supervised learning in functional genomics can be seen as a natural outgrowth of supervised learning methods in gene finding (Ratsch et al. 2007), protein sequence alignment (Do et al. 2006a), and RNA secondary structure prediction (Do et al. 2006b; Gruber et al. 2007).

2.2.6 Network Representation

2.2.6.1 From Reference Assemblies to Reference Networks

Along with algorithms for network reconstruction, a fundamental question in static modeling is the issue of data structures, of how an inferred network model is represented on the computer. The field is currently moving from a number of ad hoc network representations to a better defined concept of a *reference network*.

To motivate this, recall the concept of a "reference human genome assembly." This concept is a fiction, as the genome coils and uncoils (Champoux 2001), moves about the cell (Riddihough 2003), is methylated and demethylated (Weber and Schubeler 2007), varies substantially between individuals (Abecasis et al. 2007) and has nontrivial three-dimensional structure (SantaLucia and Hicks 2004).

Nevertheless, it is a useful fiction, as each of these phenomena can be visualized and analyzed by superimposing tracks upon the reference assembly, which represents a lowest common denominator of analysis. In particular, by separating the

raw data (the reference assembly) from the metadata (the species-specific tracks and annotations), cross-species comparisons and genome alignments are enabled (ENCODE Project Consortium 2007; Brudno et al. 2003).

Similarly, a feasible near-term goal for static modeling is the construction of reference networks for key model organisms with explicitly typed edges (Figs. 2.4 and 2.5). These reference networks may integrate multiple data types (Fig. 2.3) and incorporate explicit models of uncertainty. However, as they are meant to represent the average cell of a given organism near the median of the norm of reaction (Lynch and Walsh 1998), they *should not* directly incorporate interactions which only occur during certain perturbations, at specific times, or within particular cell types. As with reference assemblies, such conditional interactions should be modeled by superimposing tracks and layouts on the static reference network rather than incorporating conditional interactions directly into the reference network (Fig. 2.4).

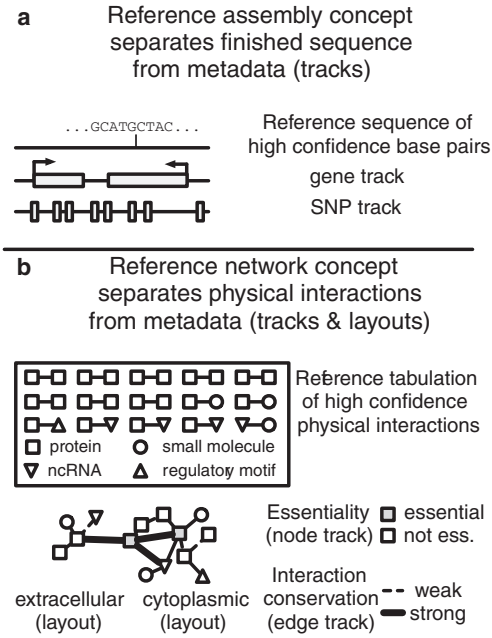


Fig. 2.4 Reference assemblies and reference networks. (**a**) The concept of a reference assembly allows us to enforce a divide between data and metadata. Everything other than finished sequence data is visualized and represented as a metadata track associated with the raw sequence (Kuhn et al. 2007). (**b**) Enforcing a similar kind of separation for a reference network will have key advantages. By enumerating a static list of highly probable physical interactions which occur for an average cell of a given species (averaged over condition, space, time, etc.), we can obtain a lowest common denominator of interaction information to compare between species. Given this physical backbone, metadata can then be visualized via tracks and layouts. For example, we can apply a node track to flag essential nodes, an edge track to highlight strongly and weakly conserved edges, and a layout to mirror the known physical separation of modules. Reproduced from Srinivasan et al. (2007) with permission from Oxford University Press

By keeping the building blocks of the reference network separate from the details of when or where they interact, a separation between data and metadata is enforced that permits powerful kinds of network visualizations and alignments (Fig. 2.4). This is particularly valuable because network metadata is likely to accumulate in bits and pieces due to the prohibitive cost of compiling cross-sectional data on different network states (Fig. 2.1). With respect to visualizing this metadata, the primary new feature in the network context is the availability of layouts in addition to tracks, which are particularly suitable for visualizing spatial or functional relationships (Fig. 2.4b).

2.2.6.2 Strongly Typed Static Network Models

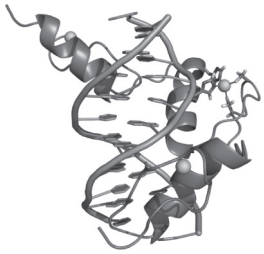
One of the most important lessons learned from genome sequencing was the value of the Gene Ontology's systematic, machine-readable approach to categorizing function (Ashburner et al. 2000). Before GO, it was impossible for a computer to discern that a protein annotated as an alcohol dehydrogenase was a kind of oxidoreductase.

We propose that a similar state of affairs is currently prevalent in systems biology, and believe that a Network Ontology for explicit ontological markup of reference networks will prove to be an essential tool (Fig. 2.5) specifically, note that the edges and nodes of the reference network in Fig. 2.4 have explicit ontological

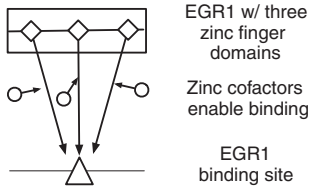
---

Fig. 2.5 (continued) Jansen and Gerstein 2004), or transcription factors and motifs (Beyer et al. 2006). In order to achieve the ambition of a refernce network, however, a notation must be devised for dealing with many kinds of typed interactions. (**a**) As a motivating example, consider the interaction of EGRI with a transcription factor-binding site, which involves three zinc finger domains and a zinc cofactor. (**b**) One possible schematic of this interaction is shown, where an individual protein with three domains (top layer) conditionally binds a DNA position (bottom layer) in the presence of zinc (middle layer). The problem is that it is not immedialtely obvious how to represent this in machine-readable terms. (**c**) One solution lies in representing a network as a list of triples encoded in a "Network Ontology". This proposed Network Ontology is a meta-ontology that draws on established ontologies and controlled vocabularies. By combining these source vocabularies, the small set of interactions described in panel (**b**) can be described in terms of a set of unordered triples. Each triple represents a fact about the network, expressed as (subject, predicate, object) tuple. In general, each member of the triple has its own canonical identifier. For example, the triple (`CID: 23994, MI: 0407, CDD: pfam 00096`) indicates that zinc (`CID: 23994` in PubChem) physically interacts (`MI: 0407`, in PSI-MI) with the zinc finger domain (`CDD: pfam 00096` in the CDD). For simplicity, we have represented the `is_a` and `part_of` predicates as literals, but in general these should also be specified by URIs. For example, the subtleties regarding the Sequence Ontology's `part_of` definition are treated during the discussion of extensional mereology operators in (Eilbeck et al. 2005). (**d**) The advantage of the triple-based representation of the network is that it corresponds to the RDF standard of the W3C consortium. While RDF can be expressed as an XML file, the N3/Turtle notation (Beckett and Berners-Lee 2007) is far more compact and human readable. Shown is an example of a Turtle format encoding of the triplestore described in panel (**c**). After the preliminary enumeration of namespaces, each non-comment line corresponds to a single triple. Reproduced from Srinivasan et al. (2007) with permission from Oxford University Press

a EGR1 interaction with DNA (three zinc fingers in complex)



b Schematic of physical interactions



Problem: implicit semantics prevent machine readability

c Represent network as a list of triples
(Network Ontology is a meta-ontology)

Subject	Predicate	Object	Note
CID:23994	is_a	MI:0682	Zinc is a cofactor
CID:23994	MI:0407	CDD:pfam00096	zinc directly interacts w/ zinc fnger domain
UniProt:P18146	is_a	GO:0003700	EGR1 is a transcription factor
craHsap:197014	is_a	SO:0000235	This motif is transcription factor-binding site
dom:P18146-d1	part_of	UniProt:P18146	First domain in protein
dom:P18146-d1	is_a	CDD:pfam00096	Type of domain
...	...	...	...

d Explicit RDF representation (Turtle/N3 format)

```
#Define Namespaces
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix CID: <http://pubchem.ncbi.nlm.nih.gov/> .
@prefix CDD: <http://ncbi.nlm.nih.gov/Structure/cdd/> .
@prefix UniProt: <http://www.ncbi.nlm.nih.gov/entrez/> .
@prefix craHsap: <http://www.cisred.org/> .
@prefix MI: <http://www.psidev.info/> .
@prefix GO: <http://www.geneontology.org/> .
@prefix SO: <http://song.cvs.sourceforge.net/> .

#Begin List of Triples
CID:23994 is_a MI:0682 .
CID:23994 MI:0407 CDD:pfam00096 .
UniProt:P18146 is_a GO:0003700 .
craHsap:197014 is_a SO:0000235 .
dom:P18146-d1 part_of UniProt:P18146 .
dom:P18146-d1 is_a CDD:pfam00096 .

#...more triples below...
```

Fig. 2.5 Network Ontology and RDF representation. Most current networks involve only one or two kinds of biological objects, such as proteins alone (Lee et al. 2004; Srinivasan et al. 2006;

labels. This Network Ontology is a kind of meta-ontology that derives largely from existing ontologies, something like a more focused analog of the Unified Medical Language System (Bodenreider 2004) for systems biology. Many of the terms can be derived from existing ontologies like the Gene and Sequence Ontology and from lists of canonical identifiers such as those available through Entrez Gene (Wheeler et al. 2007), UniProt (Mulder et al. 2007), CDD (Marchler-Bauer et al. 2003), and PubChem (Wheeler et al. 2007). There are also several available standards in the systems biology space (Stromback and Lambrix 2005) which can serve as building blocks for this project, including SBML (Hucka et al. 2003), CellML (Nielsen and Halstead 2004), BioPax (Luciano 2005), and PSI-MI (Orchard et al. 2005). Of these ontologies. SBML and CellML are invaluable tools for detailed, time-dependent modeling but may be too granular for genomic scale networks. BioPax and PSI-MI are more appropriate; BioPax was originally developed for exchanging pathway data between database such as KEGG and Ecocyc, and PSI-MI was built for describing the results of high-throughput experiments (Hermjakob et al. 2004).

By combining these source vocabulaties, a Network Ontology provides a unified framework for defining a reference network and its associated metadata in terms of lists of triples (Fig. 2.5). Each triple corresponds to a fact about the network, represented as a subject/predicate/object tuple of uniform resource identifiers (URIs). Each URI represents a canonical identifier drawn from one of the established databases or ontologies. In addition to the vast number of ontological terms compiled by the members of the OBO foundary (Rubin et al. 2006), good URIs currently exist for proteins via UniProt, domains via the CDD, genes via Entrez Gene, and small molecules via PubChem. Canonical names are also emerging for ncRNAs (Kin et al. 2007) and regulatory motifs (Robertson et al. 2006), though a consensus solution will remain elusive until NCBI or EBI launches a database.

Given a consensus set of URIs for biological objects, an explicitly typed reference network can then be naturally represented as a set of ontological triples, such as `A physically_interacts_with B, or X is_a Y`, in which canonical URIs are used for each member of the triple (Fig. 2.5). This triple-based representation of a network corresponds to the RDF format of the World Wide Web Consortium (Prudhommeaux and Seaborne 2007). Though originally developed for the Semantic Web (i.e. web page X links to web page Y), a list of triples (also known as a triplestore) is clearly also a natural representation for pathway and network information. Importantly, significant progress has already been made by the BioRDF working group (Stephens 2007) toward converting key biological databases into RDF format.

One of the principle advantages of representing network data as an RDF triplestore with canonical URIs for each member of the triple is that if everyone uses the same URIs, then facts produced by different providers can be integrated by forming the union of the two triple stores (though in practice statistical methods will be used to resolve any contradictory triples). Another advantage is that a network in RDF format with explicitly typed nodes and edges can be the subject of nontrivial queries based on the SPARQL query language (Prudhommeaux and Seaborne

2007), such as find all X's which are regulated by Y or find all singal transduction paths between A and B. A network with explicitly marked nodes and edges also suggests natural possibilities for data visualization and enables rich kinds of network alignment (Section 2.2.7.2).

Reference networks can be inferred by a direct extension of the supervised learning methods described in Section 2.2.4. As depicted in Fig. 2.3, the shared thread behind the supervised learning methods for network integration and protein function predication is to (1) select a biological object (protein pair, gene pair, protein, etc.), (2) calculate a list of desired labels and predictive features, and (3) use machine learning to compute a mapping between features and labels. Given sufficient labels and predictors data on any kind of biological object can be integrated.

This kind of approach has already been used to score interaction confidence during the process of data collection (Krogan et al. 2006); in the long run such techniques may become as common to network determination as PHRED and PHRAP (Ewing and Green 1998; Ewing et al. 1998) became in the early days of sequence determination.

2.2.7 Applications of Network Models

Now that static network modeling has become commonplace for several years, the trend is to make network analysis a starting point for applications, such as user-friendly network visualization, network-guided experimental validation, and network alignment.

2.2.7.1 Experimental Prioritization

Ultimately, an interaction network is a model of a system, and a model is only useful to the extent that it successfully predicts experiments. In particular, one of the most important ways to leverage network data is not simply to analyze it, but to use it to understand what data to gather next.

One way to formulate this problem is in terms of an experiment recommender, which uses network context to prioritize experiments. For example, network context can be used to identify genes that are likely to be in pathways of interest (Owen et al. 2003). Experiment recommenders of different kinds have also been used to determine rate constants (Flaherty et al. 2005), define metabolic topologies (Barrett and Palsson 2006), determine disease genes (Aerts et al. 2006), and discern causal structure in signaling pathways (Sachs et al. 2005).

It is important to note that many such recommendation problems can be viewed as updates of an uncertain state variable, such as the GO category of a protein or the value of a rate constant. On a formal basis, this is highly similar to the Bayesian supervised learning model for data integration described in Fig. 2.3, in which a prior gold standard is updated to produce a posterior distribution. There is thus a

significant opportunity to unify the problems of data integration and experiment recommendation in a common Bayesian framework, where experiments are recommended in order of their ability to reduce the uncertainty of state variables of interest.

2.2.7.2 Network Alignment

Once multiple genome sequences became available, research attention naturally turned to the question of comparative genomics (ENCODE Project Consortium 2007). Similarly, the availability of several different kinds of networks from different sources and species has ignited interest in comparative functional genomics. Many questions are still open in this area: for example, can we enumerate an organism's inventory of modules much as we can enumerate its inventory of genes? Is it feasible to transfer module annotations from well-studied organisms to newly sequenced ones? And can we identify conserved modules of unknown function?

One promising way of answering such questions is through network alignment, which is a systems biological analog of sequence alignment. Network alignment allows us to compare interaction networks between different species to find conserved modules. When comparing protein interaction networks, conserved modules are sets of proteins that have both conserved primary sequences and conserved pairwise interactions between species. For example, we can apply network alignment to find all species with nitrate reduction systems similar to that of *Escherichia coli*, or to examine the extent to which the cell division apparatus is conserved across a set of microbes. A sample alignment found with the Graemlin network aligner is shown in Fig. 2.6; the figure displays a putative DNA uptake and transformation module in which seven protein families across four species show a conserved pattern of functional association (Flannick et al. 2006).

Network alignment has attracted much interest in recent years, beginning with manual alignments of metabolic pathways (Dandekar et al. 1999; Forst and Schulten 2001), proceeding to precursors of network alignment guided by best bidirectional BLAST hits (Ogata et al. 2000; Stuart et al. 2003; Yu et al 2004), and culminating in more recent graph-based formulations (Kelley et al. 2003). Recent alignment algorithms have introduced the ability to compare three networks at once (Sharan et al. 2005) as well as simple models of network evolution (Koyuturk et al. 2006). We recently developed the Graemlin network aligner, which was the first program capable of identifying conserved functional modules across an arbitrary number of dense association networks. By using a number of BLAST-like optimizations Graemlin's running time scaled linearly rather than exponentially with the number of species (Flannick et al. 2006).

Just as sequence alignment rests upon substitution matrices (Henikoff and Henikoff 1993) and models of sequence evolution (Durbin et al. 1999), it will be crucial to provide a prinicipled foundation for network alignment by developing a detailed theory of network evolution (Berg and Lassig 2006; Weitz et al. 2007).

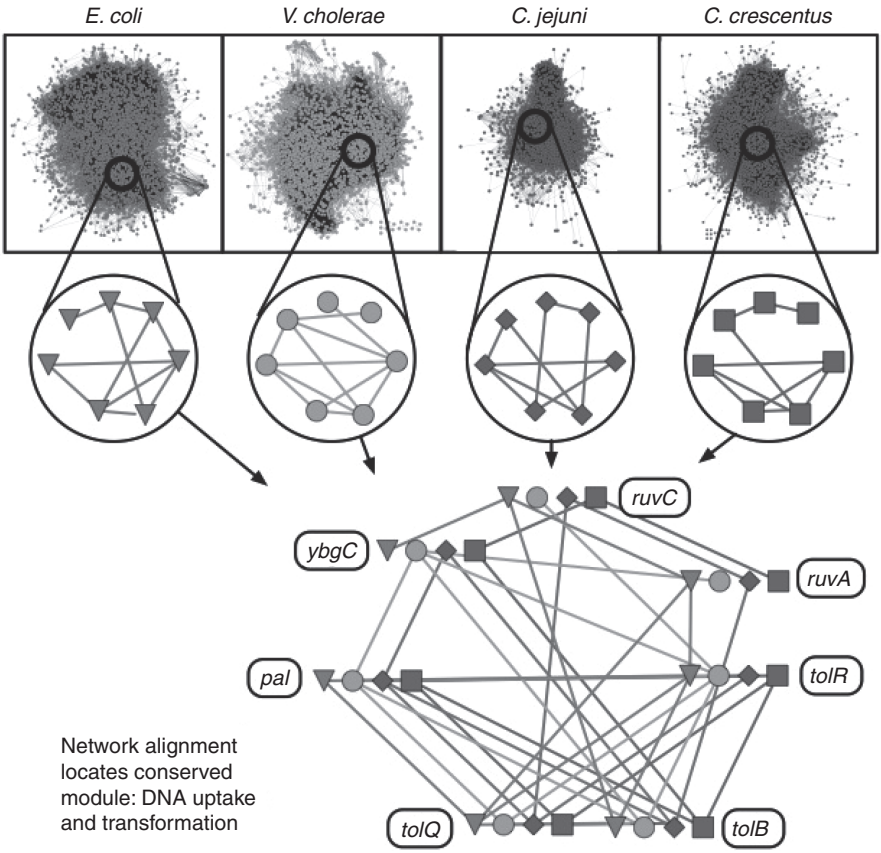


Fig. 2.6 Network alignment. A sample network alignment calculated with the Graemlin algorithm (Flannick et al. 2006). In the *top row*, integrated association networks for four microbes are depicted. In these large graphs, nodes represent proteins and edge weights are probabilities of association between proteins. Calculating a global network alignment finds several conserved modules, including one consisting of seven conserved protein families: *ruvC, ruvA, tolR, tolB, tolQ, pal*, and *ybgC*. Each family contains four homologous proteins, one in each species; node shape denotes the species of origin and proteins from a given family are grouped near each other. Moreover, the pattern of functional associations between protein families (as revealed by the edges) displays significant conservation. The alignment suggests a possible function for the module: exogenous DNA is allowed into the cell by the tol/exb membrane channel proteins and then incorporated into the chromosome by the ruv recombination proteins. The literature supports this hypothesis, as insertional disruption of tol/exb family proteins in *Pseudomonas stutzeri* reduces transformational efficiency to 20% of its previous level (Graupner and Wackernagel 2001). This strongly suggests that exogenous DNA travels through these channels before chromosomal incorporation. Reproduced from Srinivasan et al. (2007) with permission from Oxford University Press

Moreover, just as fast algorithms for sequence alignment such as BLAST became ever more essential as sequence data accumulated, it seems clear that the utility of network alignment will rise in direct proportion to the quality of inferred interaction networks in different organisms.

Indeed, the pace of research in this area is accelerating, with several papers published in the last few months (Zhenping et al. 2007; Liang et al. 2006; Singh et al. 2007; Stumpf et al. 2007). Part of the reason for this interest is that many of the signal successes of bioinformatics have been concentrated in the area of alignment (Batzoglou 2005). Even though the vast majority of objects in biology have not been directly characterized by experimentalists, information on objects which have good digital encodings, like sequences and structures, can easily propagated with an appropriate alignment tool. For example, we can characterize a protein in *Drosophila melanogaster* and immediately BLAST its digital representation to get some clue as to the function of that protein in other insects, or possibly even in humans or yeast.

Yet the lack of digital representation means that many other interesting objects (like tissues or developmental hierarchies) are not yet easily "aligned" between organisms. Currently, we resort to simple phylogenetic interpolation to reason that if organism X is phylogenetically equidistant between organism Y and organism Z, then its characteristics are intermediate between these two organisms. However, it is well known that gene trees are not the same as species trees (Degnan and Rosenberg 2006; Nichols 2001; Pamilo and Nei 1988), and that it is far more accurate to compare genes via sequence alignment. While the divergence of a network tree from the species tree is likely to be less than that of a gene tree (as a collection of genes will have lower sampling variance than an individual gene), nevertheless the same principle holds: the evolutionary history of a module is distinct from that of its host. The promise of network alignment, then, is that we may be to improve upon crude phylogenetic interpolation by directly comparing network models of higher order processes (such as organs and developmental hierarchies) between species and individuals.

2.2.7.3 Network Visualization

Large interaction data sets with thousands of nodes and edges are best visualized interactively rather than statically. Several tools for this purpose are now available and can be divided into standalone applications, programming libraries, and web applications.

Desktop Tools

Among standalone programs, several options are available, including Cytoscape (Shannon et al. 2003), Osprey (Breitkreutz et al. 2003), Medusa (Hooper and Bork 2005), and Pajek (de Nooy et al. 2005). Cytoscape is a popular choice with many features and plugins, but as it is written in Java it requires large amounts of memory to navigate dense networks. Osprey is similar in functionality and is somewhat more responsive, but has a smaller user community. Medusa has several novel features, including support for multigraphs with multiple edges between a given pair of nodes. Pajek has many features for mathematical graph analysis but a comparatively steep learning curve.

Programming Libraries

Data analysts often wish to dynamically generate network visualizations from within programs, and many libraries for this purpose are available. Cytoscape, mentioned has an API that can be called from within Java. The Boost Graph Library (Siek et al. 2007) and AT&T's Graphyviz library (Ellson and North 2007) are open source C+ libraries which have bindings for many different programming languages, including R, Python, and Perl.

Online Network Browsers

Several rich web applications for network visualization have been described in recent years, including STRING (von Mering et al. 2007), PubGene (Jenssen et al. 2001), iHOP (Fernandez et al. 2007), and PSTIING (Ng et al. 2006). STRING provides several different kinds of interaction predictions between genses for many sequenced genomes. STRING, PubGene, and iHOP all allow browsing of literature co-occurrence networks. PSTIING is a powerful data browser which is particularly useful for analysts looking for new data sets to integrate.

2.2.8 Outstanding Challenges in Static Modeling

Now that hundreds of different functional genomic data sets are available through resources like NCBI's GEO, an important near-term goal is the generation of static reference networks for major model organisms. In order to make these networks relevant, every predicted node and edge should have an associated gold-standard empirical test for verification purposes. For example, a postulated network of physical protein protein interactions is in theory confirmable by exhaustive coimmunoprecipitation of protein pairs. Moreover, the parameters of static models should be designed to be flexible enough to be updated in the light of new information, e.g., by using Bayesian updates.

2.3 Dynamical Models of Biological Networks

A dynamical model is a reconstruction of molecular physiology, a description of how the state of a system evolves over time. This description usually consists of equations that describe the time dependence of each of the state variables of the system. To describe a biological networks as a dynamical system requires identification of the variables (protein species, signaling molecules, and their associated amounts), how they interact (network connectivities), and how both the values of the state variables and their interactions change over time. A variety of approaches have been used to model signaling networks as dynamical systems; one useful means of organization is by whether the model uses discretely or continuously varying states.

2.3.1 Discrete Models

Discrete models require the states of the system variables (genes, proteins, signaling molecules) to take on integer values. Although at a molecular level this requirement is the most realistic, it is often used at a higher level to simplify the resulting models. A boolean model provides one such simplification: it consists of binary-valued variables whose interrelationships are captured by boolean functions. In systems biology, this expresses the state of a gene ("on" or "off") as a boolean function of the states of other genes. As an example, a Boolean model was constructed for the mammalian cell cycle (Faure et al 2006), and it was shown to reproduce known wild-type and mutant behavior. Boolean models can either be deterministic or stochastic; the latter as referred to as probabilistic boolean networks (PBNs) (Shmulevich et al. 2002).

In cases where a boolean model is too coarse grained for a particular system, a more elaborate dynamic Bayesian network can be used. These models can be either discrete or continuous, and they allow dynamical systems to be described probabilistically. An example of a recent discrete DBN applied to yeast cell cycle time series data is found in Zou and Conzen (2005).

Although DBNs are more realistic than boolean models, they are still more descriptive than mechanistic. Short of molecular dynamics simulations that track the simultaneous position and velocity of every molecule in the system, the most realistic (and mechanistic) signaling network models fall under the stochastic chemical kinetics framework (Gillespie 2007). These models represent biological systems as well-stirred collections of finite numbers of chemical species; reactions are simulated probabilistically according to known reaction propensities. We shall return to these models in Section 2.3.8.

2.3.2 Continuous Models

Continuous models permit system variables to take on non-negative real-valued states. We focus on so-called chemical kinetic (mechanistic) models where states represent concentrations of molecules. These models, though approximate, are sufficiently accurate when the molecular populations of all species are orders of magnitude larger than one (Gillespie 2007). The oldest and most common modeling formalism uses *ordinary differential equations* (ODEs) and known chemical kinetic/physico-chemical principles (Cornish-Bowden 1979) to deterministically model molecular concentrations as a function of time. Though these equations are not usually analytically solvable, there exist a wide variety of numerical tools that can efficiently model relatively complex systems (Rangamani and Jyengar 2007). We shall cover ODEs in more detail in Section 2.3.6.

Partial differential equation (PDE) models of signaling networks describe the evolution of molecular concentrations as functions of both space and time. These models are more physically realistic than ODEs, but they are also significantly more

difficult to solve and typically require custom-made numerical solution methods (Eungdamrong and Iyengar 2004). We discuss PDEs in detail in Section 2.3.7.

The addition of a noise term to a deterministic differential equation yields a *stochastic differential equation* (SDE), which in chemical kinetic systems often takes the form of a chemical Langevin equation (CLE) (Gillespie 2000). The CLE follows from approximations to discrete stochastic chemical kinetics, and its solution can be computed much more efficiently than solutions for the corresponding discrete models (Wilkinson 2009). We discuss SDE modeling of signaling networks in Section 2.3.8.

Discrete dynamical models represent an active area of research in systems biology, and they have recently been discussed elsewhere (Uhrmacher et al. 2005). In the remainder of this chapter we restrict our focus to the three classes of continuous dynamical models listed. As these models are mechanistic in nature, their means of specification and analysis are the most dissimilar to the descriptive models of static interaction networks (Section 2.2) and most of the discrete dynamical models mentioned above. We begin by describing some general advantages and limitations of representing biological networks with differential equations, followed by common tasks carried out when applying such models. These points will motivate the remaining discussion and the particular examples used for illustration.

2.3.3 Advantages of Continuous Dynamical Models

The cellular environment is constantly changing as a result of deterministic chemical reactions and stochastic fluctuations. Thus, dynamical systems are more realistic depictions of biology than static models, and they can be used to answer detailed questions unanswerable by the latter (Mogilner et al. 2006). In particular, the modeler can test hypotheses that would be hard to query experimentally (Angeli et al. 2004). Through simulation, dynamical models enable characterization of nonlinear, emergent behavior that evolves over time. Such behavior is often only visible at a systems level and would be missed by reductionist methods (Bhalla and Iyengar 1999).

The outputs of differential equation models relate more closely to experimentally observed phenotypes than coarser-grained alternatives (Sauer 2004). As a result, though these models often require extensive parameterization, the parameter space can be constrained such that the model reproduces experimental data. This significantly reduces the complexity of model calibration and also enables easier model validation (Rangamani and Iyengar 2007). In addition, the models we shall discuss are mechanistic, and first principles of chemical kinetics (Cornish-Bowden 1979) and physics can reduce parametric uncertainty. These same principles are often unapplicable in more approximate models (Price and Shmulevich 2007). In general, the process of parameter learning sheds light on correctness of initial hypotheses: if no parameter values exist which reproduce observed behavior, initial assumptions must be revisited (Tomlin and Axelrod 2007; You 2004; Ideker et al. 2001).

Finally, though these models require large amounts of high-resolution data, experimental systems are in place to make many of the needed measurements (Albeck et al. 2006).

2.3.4 Limitations of Dynamical Models

The level of detail present in differential equation models can also impose limitations. Implementation of these systems often requires detailed prior biochemical/network knowledge which is not always readily available or uniformly reliable (Herrgard et al. 2003; Mogilner et al. 2006). Given the extensive parameterization needed, it can be hard to validate the entire model and multiple solutions (network structures/parameter values) often exist. With limited amounts of data, models are also prone to overfitting (Amonlirdviman et al. 2005; You 2004).

Though dynamical systems are able to reproduce experimental observations, their calibration is not always compatible with high-throughput data (Price and Shmulevich 2007). Instead, these models require costlier quantitative data to define concentrations of signaling components, kinetic/diffusion parameters, and initial/boundary conditions (Weng et al. 1999; Schnell and Turner 2004).

Finally, due to their complexity, simulation of these models is computationally intensive, and models are often limited by size (Rangamani and Iyengar 2007; Tomlin and Axelrod 2007).

In light of the above points, it is not surprising that successful examples of dynamical biological modeling are in systems that benefit from the advantages while minimizing the effects of the limitations. We will cover some of these examples in detail in the remainder of the chapter.

2.3.5 Specific Tasks Associated with Dynamical Modeling

We categorize the undertakings and objectives of dynamical modeling into five common tasks (Aldridge et al. 2006a), many of which follow from the characteristics of dynamical models listed above:

1. **Model construction and calibration**. The first step is to specify the structure and parameterization of a model from prior knowledge and experimental data. As we discuss below, this often requires advanced computational and statistical methods to process noisy or incomplete data (Brewer et al 2008; Wilkinson 2009; van Riel and Sontag 2006; Jaqaman and Danuser 2006).
2. **Model validation and testing.** After calibration, it is important to compare model output with existing experimental data (Eungdamrong and Iyengar 2004; Ideker et al. 2001). This procedure is necessary (though not sufficient) to determine whether a model is specified correctly.
3. **Parameter sensitivity analysis.** Sensitivity analysis involves determining which molecular concentrations or kinetic parameters have the greatest influence on

model behavior. This is valuable when prioritizing parameters for subsequent experimental measurement or perturbation (Rangamani and Iyengar 2007).

4. **Analysis of emergent behavior.** As mentioned emergent behavior arises from systems level properties that are not apparent from studying individual components. Many of these phenomena, which can include robustness to noise, feedback, bistability, and oscillation, are best characterized through simulation of the model (Gilbert et al 2006; Angeli et al 2004).
5. **Predictive modeling and discovery**. One of the most exciting areas of systems biology is prospective modeling to test hypotheses that are too difficult or expensive to query in vivo. Here, a prerequisite for making accurate predictions is a sufficiently detailed and accurate model (You 2004).

The remainder of the chapter is split into three parts: one each covering ODE, PDE, and SDE modeling of biological systems. Each section begins with an introduction to the corresponding modeling framework, followed by a brief review of early successes from the literature. We then focus in depth on current (within the last 5 years) examples, which we discuss in terms of the five tasks listed above. We conclude each section with outstanding research challenges.

2.3.6 ODE Systems

Ordinary differential equation models are by far the most common dynamical model used in biology (Andrews and Arkin 2006). They represent behavior at the level of chemical kinetics, whereby the concentration of each system component $y_i(t)$ as a function of time is represented in the following manner:

$$\frac{\mathrm{d}y_i(t)}{\mathrm{d}t} = f_i(\mathbf{y}(t)),\ 1 \le i \le n, \tag{2.1}$$

where $\mathbf{y}(t) = \left[y_i(t), \ldots, y_n(t)\right]$ and f_i is a function which describes the rate of change of $y_i(t)$. This function can be constant (uninhibited synthesis), linear (first-order reaction such as degradation), or nonlinear (second-order reaction like Michaelis–Menten kinetics), and its precise form follows from qualitative prior experimental knowledge. These coupled expressions are often collectively referred to as *reaction rate equations* (RREs). The RREs of most biologically realistic systems cannot be solved analytically, but numerous well-developed and efficient numerical methods for solving these systems are available.

2.3.6.1 Assumptions of ODE Biological Network Models

The relative ease with which ODE models of biological systems can be constructed and solved is a consequence of the simplifying assumptions made about the system. These assumptions include as follows:

- Reactions occur in a homogeneous, well-stirred volume (corollary: molecular concentrations are functions of time and not space)
- Reactions occur in a deterministic manner
- Discrete effects on molecular concentrations can be ignored (corollary: molecular populations of all species are orders of magnitude larger than one)

The solution to the RREs describes the deterministic time evolution of the system's component concentrations; this solution often represents the average (mean) result of a population of many individual reaction trajectories in the presence of noise (Gillespie 2007). However, if any of the above assumptions are violated, ODE models of the system may be invalid and even exact solutions of such models can differ substantially from population averages. Even when the assumptions are met, it can be shown that the solution to the RREs is not equivalent to the population ensemble mean (Samoilov and Arkin 2006). Nevertheless, these models have proven useful in describing the dynamic behavior of biological networks, and they have been in use for several decades.

2.3.6.2 Early Examples of ODE Models Describing Biological Systems

One of the earliest uses of an ODE model to describe a biological network comes from Goodwin, who constructed equations describing the change in concentration of an mRNA species and its corresponding protein product (Goodwin 1963). This work simulated feedback loops, which were shown to give rise to nonlinear oscillations. Walter built upon this work, where he identified a finite range of parameter values in a feedback system that led to oscillatory behavior (Walter 1970). Tyson and Othmer furthered our understanding of feedback control in biological networks, and they characterized emergent properties such as stability, bifurcation, periodicity, and hysteresis that were exhibited by these networks (Othmer 1976; Tyson 1975; Tyson and Othmer 1978).

Since that time, ODE models have been applied to biological networks governing a wide range of functions, including viral infection (Shea and Ackers 1985), chemotaxis (Spiro et al 1997), cell cycle regulation (Novak et al. 1998), and developmental patterning (von Dassow et al. 2000); see (You 2004) for additional references. Many of these biological ODE models focus on small, well-characterized biological systems, and for good reason: they can be easily parameterized from existing knowledge and they are computationally inexpensive to characterize and solve.

2.3.6.3 Modern Applications of ODE Models to Biological Networks

We now turn to more recent ODE modeling applications in systems biology. Several of the studies below perform many or all of the common tasks listed in Section 2.3.5, but we discuss only one per study. As the use of ODEs to model biological networks is quite widespread, we have tried to choose particularly novel or innovative examples.

Bayesian Calibration of a GPCR ODE Model Using Noisy Data

G-protein-coupled receptors (GPCRs) are a large family of transmembrane receptors that facilitate the transduction of a wide range of cellular signals. Cells exposed to multiple GPCR-binding ligands often respond as if the signals are additive, though occasionally the response can be synergistic. The precise mechanism of such synergy is unknown, which motivated the authors of Flaherty et al. (2008) to model the calcium release in mouse macrophage cells exposed to the signaling molecules complement factor 5a (C5a) and uridine diphosphate (UDP). Their mathematical model consisted of 53 ODEs (constructed using prior knowledge) with 84 parameters and 24 non-zero initial conditions. Parameters were estimated from a combination of preexisting data and knowledge and newly performed experiments.

For the latter, the authors made time-resolved intracellular calcium measurements of mouse RAW264.7 cells in response to varying doses of C5a and UDP. They also collected similar measurements using five knockdown cell lines, illustrating the effects of decreasing quantities of five key signaling proteins (GRK2, Gα i2, Gαq, PLCβ3, and PLCβ4). These data were used to learn 20 of the 84 parameters most relevant to the five knockdown targets.

Unlike most optimization procedures that choose point estimates of parameters maximizing the fit to the observed data, this study adopted a Bayesian procedure to estimate a full posterior distribution of parameters given the data. Bayesian methods are well suited for incorporating prior knowledge of parameter values with observed data to arrive at updated posterior parameter estimates. These posterior estimates are calculated as the mode of a posterior distribution, specified by Bayes' rule:

$$\Pr(\theta|\mathbf{y}) = \frac{p(\mathbf{y}|\theta)\Pr(\theta)}{\Pr(\mathbf{y})}, \tag{2.2}$$

where θ represents the parameters, $\mathbf{y}$ the observed data, $\Pr(\cdot)$ a probability measure, and $p(\cdot)$ a likelihood function. The posterior distribution often cannot be expressed in closed form; in these cases Markov chain Monte Carlo (MCMC) methods are used to generate samples from the distribution. The probabilistic nature of the Bayesian framework is appropriate for dealing with the presence of uncertainty; the authors note that measurement uncertainty and knockdown efficiency uncertainty are two such sources present in their data.

Informative prior distributions were placed on the 20 parameters of interest that excluded negative values and centered on previous estimates of these parameters from biochemical experiments. A Metropolis–Hastings algorithm was then used to empirically estimate the posterior density of the parameters using the data measurements in conjunction with a Gaussian likelihood function. This procedure resulted in posterior parameter estimates that were sometimes quite different from (but still influenced by) their prior values. Figure 2.7 shows two examples. Each parameter's posterior distribution provides an automatic measure of precision: the tighter the distribution, the more precise the estimate. As the authors note, parameters with low precision are good candidates for further biochemical experimentation.

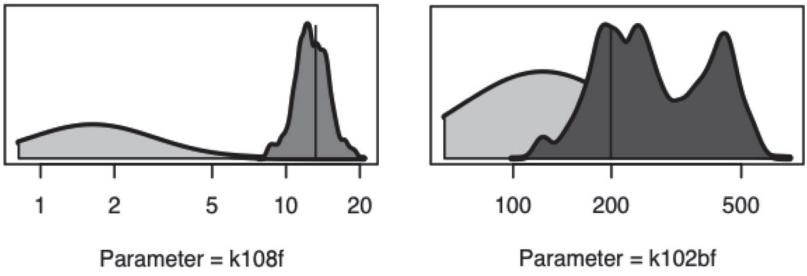


Fig. 2.7 Prior (*light gray*) and posterior (*dark gray*) density estimates for two parameters from the GPCR ODE model in Flaherty et al. (2008). Distributions consist of ∼30,000 MCMC samples; *vertical line* denotes parameter value chosen for the model. Densities are plotted as a function of parameter values on a log scale. Reproduced from Flaherty et al. (2008) with permission from PLoS

Simulation and analysis of the calibrated model led to new insight into synergistic GPCR-mediated calcium release in macrophage cells. Specifically, the authors discuss the mechanistic causes, robustness, and specificity of synergy. The authors also discuss two reasons why a Bayesian formulation was effective for calibration of their model: the abundance and quality of collected data and the speed and robustness of algorithmic methods for sampling from posterior distributions.

Validation and Testing of a Mathematical Model of Cell Death

The proper regulation of cellular apoptosis is essential for multicellular development, and its misregalation has been implicated in cancer, HIV progression, and viral infection, among other disorders. One of the mysteries of the apoptosis mechanism stems from the observation that cells receiving a tumor necrosis factor (TNF) or TNF-related apoptosis-inducing ligand (TRAIL) signal undergo a variable length delay followed by immediate cell breakdown. This breakdown is due to effector caspase activity on cellular substrates. To better understand the overall process, termed "variable-delay, snap-action" switching, the authors of Albeck et al. 2008b built an ODE model including reactions both upstream and downstream of a pivotal apoptotic process: mitochondrial outer membrane permeabilization (MOMP). Their model, referred to as EARM v1.0 (extrinsic apoptosis reaction model), consists of 58 coupled ODEs describing 18 gene products and their modifications across two cellular compartments. The model requires values for 70 rate constants, which were manually adjusted to minimize the difference between simulated and experimental data measuring caspase activity, timing of MOMP, and effects of protein depletion and overproduction.

Once calibrated, an essential requirement of any mechanistic model is that it accurately reproduce experimental data. The authors simulated TRAIL treatment over a range of concentrations and measured the switching time between initial and complete effector caspase substrate cleavage (T_s), the fraction of cellular substrate

cleaved by caspases upon cell death (f), and dose-dependent variation of the variable length delay period (T_d). These matched previously experimentally observed values of ~30 min and 1.0 for the first two, and a negatively sloped curve ranging from 3–10 h for the latter. Simulated time courses of processes involving three gene products (Bid cleavage, Smac translocation, and cPARP levels) also closely matched experimentally observed trends. Figure 2.8 displays these results.

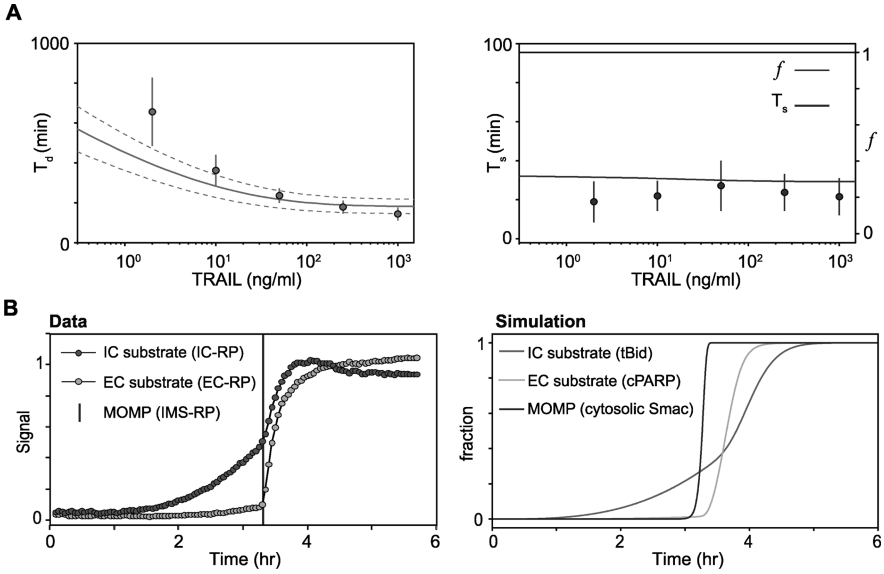


Fig. 2.8 Training data derived from live-cell microscopy used in Albeck et al. (2008b). **a** Simulation of T_d (*left*) or T_s and f (*right*) as a function of TRAIL dose (*lines*) alongside corresponding experimental values (points with *error bars* indicating standard deviations). For predicted values of T_d, an envelope of constant coefficient of variation (CV) is shown, as estimated from experimental data (CV ≈ 20%); the source of variation is not known. **b** Composite plot of IC-RP and EC-RP cleavage (measuring initiator and effector caspase activity, respectively) for >50 cells treated with 50 ng/ml TRAIL and aligned by the average time of MOMP (*left*) and model-based simulation of the corresponding species (*right*). Data in the left panel were originally reported in Albeck et al. (2008a). Reproduced from Albeck et al. (2008b) with permission from PLoS

Upon proper experimental validation, the model was then used to make six predictions concerning the molecular mechanisms of variable-delay, snap-action switching. These predictions were all supported experimentally, leading to a deeper understanding of TNF/TRAIL-regulated apoptosis. The authors also demonstrated that the level of mechanistic detail of their model (including compartmentalization) was necessary to faithfully reproduce experimental results, as a series of simpler models did not adequately fit the data. Though it is noted that the parameters of EARM v1.0 are mathematically non-identifiable, the empirical approach used to select parameter values that best matched observed data led to an accurate model of a relatively complex biological system.

Multivariate, Transient Response Sensitivity Analysis of Model Initial Conditions

Traditional sensitivity analysis measures the effects of single parameter changes on time-evolving model behavior. This univariate approach is useful for identifying reactions and species of importance to the overall reaction scheme, but it cannot characterize multiparameter effects on behavior. Naïve approaches that measure effects of changing multiple parameters simultaneously are often computationally intractable. In contrast steady-state sensitivity analysis can identify and describe equilibrium system states as a function of multiple parameter values, but this approach necessarily ignores transient effects on system dynamics. As signal transduction networks often utilize short-lived signals to enact downstream function, methods that characterize transient parameter sensitivities would be beneficial.

To satisfy both of the above requirements, the authors of Aldridge et al. (2006b) have applied direct finite-time Lyapunov exponent (DLE) analysis to a biological ODE system to determine sensitivities to model initial conditions (hereafter referred to as "parameters"). DLE analysis captures transient behavior as a function of all parameters simultaneously. The method can be used to identify *separatrices* or regions in multivariate initial condition space that separate qualitatively different downstream responses. A DLE takes on the following form:

$$\mathrm{DLE}\,(t, \mathbf{x}_0) = \log\left[\lambda_{\max}\left(\left(\frac{\partial \mathbf{x}\,(t)}{\partial \mathbf{x}_0}\right)^T\left(\frac{\partial \mathbf{x}\,(t)}{\partial \mathbf{x}_0}\right)\right)\right], \tag{2.3}$$

where $\mathbf{x}_0$ is a vector of initial conditions, $\mathbf{x}(t)$ is a vector of species concentrations as a function of time, and $\lambda_{\max}$ is the square of the spectral norm of the deformation gradient $\partial\mathbf{x}(t)/\partial\mathbf{x}_0$. Thus, a DLE measures the local sensitivity to changes in parameters evaluated at a finite time, with large DLE values corresponding to large sensitivity of the system trajectory to parameter changes. Practically speaking, DLEs are calculated numerically across a multidimensional grid of parameter values; the presence of separatrices can be visualized in plots of DLE versus a two or three-dimensional subset of parameters.

In Aldridge et al. (2006b), DLE analysis is applied to a subset of the apoptosis model in Albeck et al. (2008b) containing eight ODEs. This portion of the apoptosis pathway contains the activation of caspase-3 by caspase-8, leading to cell death, and the influence of X-linked inhibitor of apoptosis (XIAP), which negatively regulates caspase-3 activity. The authors note that this system is expected to have a separatrix due to the cell's binary decision of life or death. Systems with more graded responses would have uniform DLEs and thus be unlikely to have discernible separatrices. DLE analysis on the apoptosis system identified a pronounced nonlinear separatrix between cell survival and death. The separatrix tends toward increased XIAP concentrations as the amount of active caspase-8 increases, highlighting the antagonistic effects of these species. For comparison, the authors also applied steady-state sensitivity analysis to the model and demonstrate that cell fate is indistinguishable based on steady-state locations.

More generally, results of a DLE analysis allow the prediction of cell fate at a given time based only on initial species concentrations. This ability will likely be useful for a number of applications, including the characterization of cellular disease states. As ODE models of biological networks continue to increase in complexity and scale, it is likely that sophisticated dynamical systems tools like DLE analysis will be more frequently used for the identification and characterization of systems-level properties.

Dose-to-Duration Encoding as a Means to Transmit Quantitative Information

Signaling networks are responsible for transmitting extracellular signals to intracellular components to generate appropriate cellular responses. This transmission is not solely passive; many signaling networks involve signal modulation leading to phenomena like cross inhibition and negative feedback. Often, cellular response to a signal depends on the dose of that signal, so the signaling system is capable of transmitting quantitative information about the dose to downstream effectors. One example of such a system lies in the pheromone response pathway of *Saccharomyces cerevisiae*, where the dose of the pheromone leads to qualitatively different yeast phenotypes. At low dose, cells engage in vegetative growth; at intermediate pheromone levels the cells adopt an elongated shape, and at high dose the cells undergo growth arrest and extension of mating projections.

The authors of Behar et al. (2008) propose a *dose-to-duration* mechanism as a means for encoding pheromone dose into varying downstream behaviors. Unlike linear response pathways, whose dynamic range is limited by saturation levels of network components (i.e., receptors), a dose-to-duration mechanism can increase the dynamic range of the system in such a way that dose-dependent responses can continue even after saturation of pathway components. This ability is due to the nonlinearity of the signaling pathway and can lead to a more robust transmission mechanism when acting between heterogeneous components.

The authors begin with observations from previous work (Hao et al. 2008), which suggest that increasing doses of pheromone signal lead to increased dose and duration or only duration of two intracellular MAP kinases (Fus3 and Kss1). A hypothetical pathway architecture is constructed consisting of four components that make up a negative feedback loop (Fig. 2.9). Through calibration and simulation of a simple ODE model, it is shown that dose-to-duration encoding is a valid response of even this simple system. The authors then construct a similar signaling network using components of the yeast pheromone response pathway and fit the parameters of a six ODE model to observed data. They demonstrate that this simple dynamic model results in dose-to-duration encoding that matches experimental data closely. Though this agreement does not prove the correctness of the model, it does suggest a biologically plausible mechanism for the observed emergent behavior. The authors emphasize that information transfer via a dose-to-duration mechanism occurs through transient activation of pathway components (enacted by signals of varying durations). This underscores the necessity of using dynamical (i.e, ODE) models to understand such behavior, as static or steady-state models would preclude such transient phenomena.

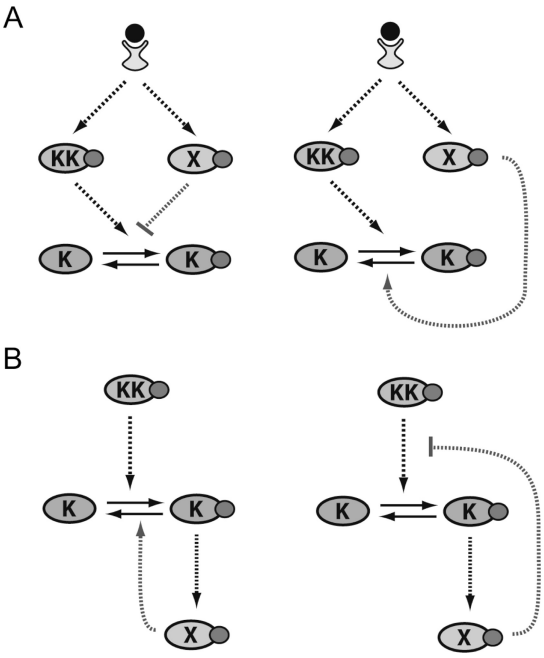


Fig. 2.9 Pathway architectures that convert stimulus dose to signal duration. **a** Feed-forward and **b** negative feedback encoding modules (KK: Kinase–Kinase, K: Kinase, X: Phosphatase). Shown are cases of negative regulation operating by inhibiting activation (*left*) or promoting deactivation (*right*). Reproduced from Behar et al (2008) with permission from PLoS

Further implications of the dose-to-duration mechanism are discussed, including its potential relevance to multicellular organisms. In particular, photoreceptors in rod cells encode intensity of light as the duration of downstream G-protein-mediated activity. Such behavior may be due to a biochemical mechanism similar to that observed in yeast. As typical signal transduction pathways exhibit more elaborate architecture than that modeled in this study, more complex variations of dose-to-duration encoding likely exist and the methods of analysis featured in this work will be useful to decipher such behavior.

Predictive Modeling with a Large-Scale ErbB Signaling ODE System

ErbB signaling, which encompasses the pathways activated by the ErbB1-4 receptor tyrosine kinases, is one of the best-studied components of multicellular eukaryotic signal transduction. Abnormal ErbB signaling has been implicated in many human cancers, and members of these pathways are common drug targets. The four ErbB receptors orchestrate a complex array of cellular signals, as they are known to bind 13 distinct ligands, form hetero-and homo-oligomers once bound, and activate multiple downstream pathways including the MAPK/ERK and PI3K/Akt cascades. It is

not surprising that the precise mechanisms for how different ligands induce differing downstream responses are poorly understood.

To improve our understanding of ErbB signaling, Chen et al.(2009) developed a large-scale ODE model including all four ErbB receptors and the ERK and Akt signaling pathways. In the interest of computational tractability, the authors made several simplifications in the number and type of receptor dimers, phosphorylation states, and structure of degradation pathways when constructing the model. Nevertheless, 828 reactions remained, which were described by 499 ODEs with 229 parameters. To calibrate the model, the authors set parameters to literature-derived values when possible, and a subset of the rest were learned from experimental data (chosen according to their impact on an objective function describing model fit).

Experimental data consisted of ErbB1, Akt, and ERK activity levels across a 2-hour time course following stimulation with two different ligands. Given the complexity of the model, the parameters were expected to be non-identifiable (multiple combinations of parameter values fit the data equally well), and a simulated annealing optimization scheme was used repeatedly to identify these best-fit parameter value combinations.

Once the model was (partially) constrained, the authors used simulation results to make predictions and test them experimentally. The first validated prediction involved differential sensitivity of ERK and Akt activity to treatment with the anti-ErbB drugs gefitinib and lapatinib. The ODE model predicted that Akt activity would be more sensitive to both drugs and experimental results corroborated this result. Next, several predictions concerning the dose–response of the ErbB network to ligand were made and subsequently tested. One of these predictions was for a Hill coefficient (H_{app}) describing the steepness of the pathway response to increasing ligand concentration. This coefficient is used in the Hill equation.

$$\text{signal}\,(x) = \frac{x^{H_{\text{app}}}}{x^{H_{\text{app}}} + k^{H_{\text{app}}}}, \tag{2.4}$$

where "signal" is a measure of the pathway response, x represents the concentration of ligand, and k is the concentration of ligand that gives half-maximal response. Previous work in *Xenopus* oocytes predicted a switch-like ERK response to progesterone (acts as a proxy for EGF) (Huang and Ferrell 1996). This corresponded to a Hill coefficient of 4.9. In contrast, the ODE model of Chen et al. predicted a much more gradual response to EGF treatment ($H_{\text{app}} \sim 0.30$), and experimental data confirmed this result. The reason for the discrepancy was identified when the authors created a sub-model of the ERK response pathway. Simulation of this model when treated with EGF reproduced switch-like activity, suggesting that modeling of the larger signaling context was necessary to faithfully reproduce the experimental observations.

This study demonstrates that a large-scale, partially constrained ODE model consisting of many elementary reactions is capable of accurately predicting observed data.

2.3.6.4 Outstanding Challenges in ODE Modeling

As mentioned above, use of ODE models is widespread in systems biology, and recent applications have begun to model larger and larger signaling networks. We expect this trend to continue; thus, an obvious challenge is the proper calibration of these large, complex models. This will require advances in the quality and quantity of time-resolved data generation and collection. Several recent developments in this area are discussed in Albeck et al. (2006). Additionally, as models grow in size, parameter learning can become prohibitively difficult, and certain parameters will be non-identifiable given limited experimental data. Computational methods to identify and correctly deal with these parameters will be needed; work in (Chen et al. 2009) provides a nice example of how to address this task.

An additional challenge arising with larger models is defining the structure of the individual reactions and their constituent species. In the past, this was mostly performed manually based on prior knowledge, but this process is time-consuming and often not feasible for less well-studied systems. Thus, automatic generation of model structure from high-throughput data is an area of active research; recent examples can be found in Carrera et al. (2009) and Bonneau (2008).

2.3.7 PDE Systems

Biological systems are known to exhibit spatial inhomogeneity, and some tasks require explicit modeling of the spatial dimension. This is especially true when the biological system in question extends across several cellular organelles, each potentially containing different components, or when the diffusion of individual components across the modeled space cannot be treated as an instantaneous process. Compartmental ODE models have been successfully used to model the former case, where components are assumed to be well mixed within compartments and transport between compartments occurs at a much slower measurable rate (Aldridge et al. 2006a). As these models are modified versions of the ODE models described above, we will not discuss them further.

In the latter case, i.e., when explicitly modeling the diffusion of certain components, partial differential equation models are necessary. Here, the spatial dimension is modeled as a continuous quantity, and the concentration of each component becomes a function of both space and time. The PDEs most commonly used to describe such systems are *reaction–diffusion* equations, where the concentration of each component $y_i(t)$ of the system can be represented as follows (derived using Fick's second law of diffusion):

$$\frac{\partial y_i(t)}{\partial t} = f_i(\mathbf{y}(t)) + D_i \sum_{j=1}^{m} \frac{\partial^2 y_i(t)}{\partial x_j^2}, \; 1 \le i \le n, 1 \le m \le 3, \tag{2.5}$$

where $\mathbf{y}(t)$ is as above, D_i is a diffusion coefficient, x_j represents a spatial dimension, and m is the number of spatial dimensions modeled. The first term on the RHS, f_i, describes the contributions of chemical reactions to the time derivative, and the second term describes the contributions of diffusion.

Compared to ODE models, PDE systems are much more challenging to solve, in part because they require many more parameters (Eungdamrong and Iyengar 2004). Aside from the kinetic parameters needed to specify f_i, the reaction–diffusion system requires a diffusion coefficient for each species (which are difficult to measure experimentally (Rangamani and Iyengar 2007)), and fluxes and/or concentrations of each component must be specified at the boundary of the physical space being modeled. This latter constraint becomes even more prohibitive when considering complex physical geometries. Solutions to nonlinear PDE systems are almost exclusively numerical, and the added realism of the model comes at a computational cost due to the increased dimensionality of the system.

2.3.7.1 Early Examples of PDE Models Describing Biological Systems

Models of biological systems governed by PDEs employ two of the three simplifying assumptions of ODE models, with spatial homogeneity being the exception. Nonetheless, when mathematically and computationally tractable, these models can accurately reproduce spatially verying molecular behavior. One of the first examples of a PDE model describing a biological system modeled the behavior of two (generic) morphogens reacting and diffusing through simple geometries of cells (Turing 1952). These equations were constructed in a way that provided for analytical solutions, and these solutions (expressed as functions of the morphogen concentrations) gave rise to spatial patterns reminiscent of those seen in organismal development.

Subsequent work elaborated upon this simple model of morphogen-controlled patterning. One study simulated a four morphogen reaction–diffusion system to mimic pattern formation in Drosophila embryogenesis (Lacalli 1990). By adjusting model parameters, the authors could produce striped patterns of morphogen concentration compatible with observed wild-type and mutant phenotypes. Another application modified the simple morphogen model to allow diffusion coefficients to depend on the spatial variable (Maini et al. 1992). Model-derived patterns were shown to produce behavior more compatible with known mechanisms of vertebrate limb development. A model of pattern formation in the context of *E. coli* cell division was created using six coupled PDEs modeling three proteins in both membrane-bound and cytoplasmic states (Meinhardt and de Boer 2001). Results from the model confirmed observed spatial oscillatory behavior of two of the proteins and suggested a molecular mechanism for the centralized localization of the third.

Additional reaction–diffusion models have been constructed to describe diverse biological processes such as striping patterns in fish (Kondo and Asai 1995; Asai et al. 1999), cell migration in butterfly wings (Sekimura et al. 1999), and avian embryogenesis (Painter et al. 2000). See Baker et al. (2008) for many additional references.

2.3.7.2 Modern Applications of PDE Models to Biological Networks

Recent PDE models capture more biological detail (and are thus more realistic) than their earlier counterparts. As with the ODE section above, we focus on work published within the last 5 years and discuss one particularly innovative example per dynamical modeling task. We note that there are considerably fewer numbers of published studies using PDEs to model signaling networks (when compared to ODEs); this is due to the increased computational complexity and demand for experimental data imposed by these models.

Calibration of a Planar Cell Polarity Model Using Qualitative Phenotypes

The process by which planar cell polarity (PCP) signaling generates distally oriented hairs in cells of the D. *melanogaster* wing is not fully understood. Aside from wild-type function, certain single gene mutants in cell clones result in an aberrant hair phenotype in adjacent cells, a process called *domineering non-autonomy*. The primary molecular players in this process include the transmembrance receptors *Van Gogh/strabismus* (Vang) and *frizzled* (Fz), and the cytoplasmic proteins *Dishevelled* (Dsh) and *Prickle-spiny-legs* (Pk). Experimental evidence indicates that these proteins selectively accumulate on the distal or proximal sides of wing cells during wild-type function. In addition, unknown diffusible factors X and Z have been proposed to explain domineering non-autonomy, although no such factors have yet been experimentally identified.

To better understand the process of PCP and to test whether domineering non-autonomy is possible without implicating unknown diffusible factors, Amonlirdviman et al. developed a reaction–diffusion model of hexagonal cells arranged in a planar array (Amonlirdviman et al. 2005). They included the four identified proteins listed above and they simulated their known influences on each other via a feedback loop by allowing the formation of six protein complexes (DshFz, VangPk, FzVang, DshFzVang, FzVangPk, and DshFzVangPk). The FzVang interactions were designed to occur across adjacent cells, and all others are conducted intracellularly. The existence of most of these interactions is supported by experimental evidence.

The feedback loop created by these four proteins is thought to amplify an initial asymmetry cue, resulting in their polarized spatial accumulation. The authors implemented two different forms of such a signal and found that both resulted in similar behavior. The overall model contained a system 10 nonlinear PDEs, whose rate constants, diffusion coefficients, and initial protein concentrations were unknown. To calibrate these parameters, the authors created an objective function describing the error between the model output and 12 qualitative experimentally observed phenotypes (both wild-type and mutant). Table 2.1 lists the phenotypes and corresponding genotypes. Numerical optimization methods were used to identify parameter values that satisfied all constraints, and sensitivity analysis demonstrated that some parameters were more tightly constrained than others.

Table 2.1 Characteristic PCP phenotypes (and associated references) used in model objective function. [Table reproduced from Table S1 of Amonlirdviman et al. (2005)]

Genotype	Phenotype
Wild-type	Asymmetric accumulation of Dsh and Fz on the distal cell membrane. Asymmetric accumulation of Pk and Vang on the proximal cell membrane (Tree et al. 2002; Bastock et al. 2003; Axelrod 2001; Strutt 2001).
dsh	Polarity disruption inside of the mutant clone. Autonomous phenotype (Kligensmith et al. 1994; Theisen et al. 1994).
fz	Distal domineering non-autonomy (Gubb and García-Bellido 1982).
Vang	Proximal domineering non-autonomy (Taylor et al. 1998).
pk	No polarity reversal (Amonlirdviman et al. 2005).
>>*dsh*[a]	Proximal domineering non-autonomy (unpublished).
>>*fz*	Proximal domineering non-autonomy (Strutt 2001).
>>*Vang*	Distal domineering non-autonomy (unpublished).
>>*pk*	Distal domineering non-autonomy.
$fz^{autonomous}$	Polarity disruption inside of the mutant clone. Autonomous phenotype (Jones et al. 1996).
>>$fz^{autonomous}$	Proximal domineering non-autonomy (Strutt 2001).
EnGAL4, UASpk	Overexpression of Pk results in protein accumulation to a degree greater than or equal to that for wild-type results (Tree et al. 2002).

a>> denotes overexpression

After calibration, numerical simulation of the model was used to investigate potential mechanisms for domineering non-autonomy. In particular, two mutant alleles of frizzled (fz^{F31}, fz^{R52}) that cause autonomous and non-autonomous phenotypes, respectively, were hypothesized to differ in their interactions with Vang. By making the necessary nodifications to the model for each allele, the desired mutant behavior was reproduced and experimental evidence confirmed the hypothesized differences in Fz-Vang interactions.

The authors note that though simulation of their PDE system is able to reproduce all known phenotypes, this does not prove the correctness of the underlying biological model. Nevertheless, model results demonstrate the feasibility of the proposed mechanism for domineering non-autonomy and suggest that unknown diffusible factors are not needed to explain the behavior of this system.

Validation of a Simple diffusion Model of Bicoid in the Drosophila Embryo

Another well-studied signal transduction system in *D. melanogaster* controls anteroposterior patterning in the developing embryo. Here, gradients in the concentration of maternal proteins establish gene expression domains that lead to eventual body segmentation. The *Bicoid* (Bcd) transcription factor is one of the best-studied maternal morphogens. Bcd RNA is deposited during oogenesis at the anterior pole of the egg, resulting in an anteroposterior protein gradient. Bcd has been shown to regulate the *hunchback, krüppel*, and *even-skipped* genes which collectively generate striped gene expression patterns.

Though it was hypothesized that gradients of Bcd arise through simple diffusion, this claim had never been rigorously verified. To address this, Gregor et al. injected dextran particles of similar size to Bcd into a Drosophila egg and made concentration measurements across a time course at 18 spatial positions (Gregor et al. 2005). They compared these results to those predicted numerically by a simple three-dimensional diffusion model in an embryo-shaped volume, governed by the following equation:

$$\frac{\partial c\,(\mathbf{r}, t)}{\partial t} = D\nabla^2 c\,(\mathbf{r}, t), \tag{2.6}$$

where $c(\mathbf{r},\ t)$ represents particle concentration at position $\mathbf{r}$ and time t, D is the diffusion coefficient, and ∇^2 is the Laplace operator (sum of the unmixed second partial derivatives with respect to each spatial dimension). A nonlinear fitting routine was used to select the value of D that minimized the difference between the experimental and predicted concentrations. When simulation output was compared to experimental results, it was found that the simple diffusion model fit the data very closely.

Gregor et al. then constructed a reaction–diffusion model for Bcd protein, parameterized by the Bcd diffusion coefficient and decay lifetime (τ). The corresponding PDE takes the following form:

$$\frac{\partial c\,(\mathbf{r}, t)}{\partial t} = D\nabla^2 c\,(\mathbf{r}, t) - \frac{1}{\tau} c\,(\mathbf{r}, t), \tag{2.7}$$

where the second term on the RHS of Eq. (7) represents the degradation rate of Bcd. Immunofluorescence data of Bcd protein in Drosophila embryos were used to fit the parameter τ, resulting in an estimate of $\sim$6 min. The authors then asked how Bicoid behavior would scale with increasing embryo size, such as is observed in *Calliphora vicina*, a fly with an egg length $\sim$3 times greater than Drosophila. Dextran injection experiments were repeated for three additional species of fly with varying embryo sizes, and estimated diffusion coefficients were shown to be similar to that of Drosophila. The values of τ for these other species were estimated as before, yielding values that ranged from 3 for the smallest embryo size to 32 min for the largest. These values are plausible, yet near the upper limit given the species' respective developmental time courses. Such a limit further supports the hypothesis of simple diffusion for Bcd behavior, as active cellular mechanisms restricting Bcd diffusion would require decay lifetimes for proper gradient formation that exceed developmental time scales.

The collective findings from this study argue that Bcd gradient formation is controlled by simple diffusion and that the protein's decay lifetime increases with increasing egg size in different fly species. Interestingly, as fly embryos of different species develop along similar time scales, the above conclusion implies that pattern formation based on diffusible Bicoid would become physically impossible in fly embryos much larger than *C. vicina*.

Sensitivity Analysis of a Sonic hedgehog Signaling PDE Model

Traditional studies of morphogen gradients in development have focused on steady-state signal levels and their effects on target genes. Recently, it has become clear that gradient dynamics are important to tissue patterning as well, as evidenced by the *Sonic hedgehog* (Shh) signaling pathway. Shh forms a concentration gradient during vertebrate development, and it is involved in limb bud, midbrain, and spinal cord patterning. It has been shown that both time of exposure to Shh and the timing of Shh secretion are determinants of tissue patterning. Besides passive diffusion, mechanisms like active transport and interactions with cell surface and extracellular matrix components are known to affect the temporal dynamics of signaling.

To achieve a better understanding of Shh signaling dynamics, Saha and Schaffer have constructed a multicellular PDE model of spinal cord patterning in the chick embryo (Saha and Schaffer 2006). The model comprises a transverse section of the developing neural tube during the time when Shh secretion from the floorplate induces dorsally oriented cells to switch from an interneuron to motoneuron fate (∼33–116 hours after egg laying). System behavior is governed by eight coupled reaction–diffusion equations describing Shh diffusion and its interaction with receptors, membrane proteins, and downstream transcription factors. Model parameters were chosen from known values in the literature or estimated from similar biological systems, and parameters were adjusted to match experimental observations. Simulation was carried out using a finite element method numerical solver.

The authors conducted sensitivity analysis on model parameters to determine how varying each one affected the response of a target transcription factor (Gli 1) to changing Shh concentration. Previous work identified two steady states corresponding to a Gli 1 switch being ‘on’ and ‘off’ (concentration above and below a threshold, respectively). Three behavioral regimes were described: one where the ‘on’ state is stable, another where the ‘off’ state is stable, and a bistable regime. By varying the values of each parameter separately across four orders of magnitude, it was determined that some parameters did not affect the behavioral regime (rate constant of Shh receptor outflux), while others altered it given large enough changes (rate constant of Shh receptor influx). Choice of regime was most sensitive to the value of the maximum rate of Glil synthesis, where even small changes altered behavior.

The authors then employed the model to reproduce known tissue patterning results and characterize novel behavior. They confirmed that Shh interactions and active transport played an important role in Shh-induced downstream changes, due in part to modulation of Shh signal dynamics. Characterization of this behavior would not have been possible using a steady-state model, underscoring the importance of dynamical models for understanding signal transduction systems.

Wave Propagation in Astrocyte Signaling Networks

The nervous system is composed of two cell types: neurons and glial cells. For years glial cells were regarded as nothing more than support cells for neurons, until

it was shown in 1990 that glutamate can induce Ca^{2+} waves in astrocyte (a type of glial cell) cultures (Cornell-Bell et al. 1990). Proposed mechanisms for astrocytic Ca^{2+} wave production fall into two categories: simple intercellular diffusion of the IP_3 signaling molecule and active regeneration of signal by released ATP. Recent experimental evidence supports the latter theory, although the mechanism of ATP release is still largely unknown. A better understanding of this process would be clinically useful, as abnormal astrocytic wave propagation has been linked to disorders including migraine and epilepsy.

Stamatakis and Mantzaris (2006) have attempted to clarify the mechanism of astrocytic wave propagation through mathematical modeling. They constructed both a single-cell ODE and a multiple cell PDE model. The latter comprises a coupled set of four reaction–diffusion equations taking the following form:

$$\frac{\partial \mathbf{u}}{\partial t} = \mathbf{D}\nabla^2\mathbf{u} + \mathbf{f}(\mathbf{u}), \tag{2.8}$$

where $\mathbf{u} = \left[[ATP]\,[IP_3]\left[Ca^{2+}\right]h\right]^T$, $\mathbf{f}(\mathbf{u})$, is a vector of reaction terms for each species derived from the single cell model, and $\mathbf{D}$ contains diffusion coefficients. h is a dimensionless variable containing information about the fraction of open channels. Values for the diffusion coefficients were estimated from the literature, and the PDE model was numerically simulated in both one-and two-dimensional domains.

ATP-mediated wave propagation was characterized in PDE models employing one of two hypothetical mechanisms: Ca^{2+}-dependent, excitable (by positive feedback) ATP release and IP_3-dependent, non-excitable ATP release. The authors found that the first mechanism led to frequency-encoded oscillations in single cells and propagation of one-dimensional waves of infinite range in multiple cells. In two dimensions, a point stimulus of ATP led to spiral waves of ATP and Ca^{2+}. In contrast, the second mechanism did not lead to single-cell oscillatory behavior, and multiple cells exhibited propagation of waves with finite range. This behavior was due to the extracellular ATP concentration falling below a threshold at a certain distance from the original stimulus.

Experimental data have been observed that support both ATP-mediated mechanisms. On the one hand, spiral waves have been detected in cell culture, and previous explanations invoked spatial inhomogeneities between cells as the cause. Model results from this study argue that a simple mechanism of Ca^{2+}-dependent ATP release is sufficient to explain such patterns. On the other hand, experimentally observed astrocytic wave propagation is finite in range. This is suggestive of an IP_3-dependent ATP release mechanism. Though the results of this study do not definitively favor either of the two mechanisms of astrocytic wave propagation (or argue strongly for both), the proposed spatial model does provide a testing ground for further hypotheses to ultimately elucidate the true mechanism.

Predictive Modeling of Molecular Mechanisms for Hair Follicle Spacing

As mentioned above, one of the earliest PDE models of a biological signaling system implicated only two morphogens reacting and diffusing through cellular space.

This model, proposed by Alan Turing, treated one morphogen as an activator and the other as an inhibitor, and its simulation produced patterns reminiscent of pigmentation in the animal kingdom (Turing 1952). Until recently, a bona fide real-world example of the model had never been identified; thus, doubts existed regarding its authenticity (Maini et al. 2006).

Sick et al. (2006) model and experimentally characterize such an example: the Wnt signaling pathway controlling hair follicle spacing in mice. A collection of proteins from the Wnt family and their known inhibitors Dkk1 and Dkk4 are present in and around the hair follicle during development, and available data suggest that the Wnt and Dkk proteins are the primary determinants of follicle spacing patterns. To test this hypothesis in the framework of the Turing reaction–diffusion model, the authors constructed the following PDE model:

$$\begin{aligned}\frac{\partial a}{\partial t} &= D_a \nabla^2 a + \rho_a \frac{a^2}{(K_h + h)\left(1 + \kappa a^2\right)} - \mu_a a, \\ \frac{\partial h}{\partial t} &= D_h \nabla^2 h + \rho_h \frac{a^2}{(K_h + h)\left(1 + \kappa a^2\right)} - \mu_h h,\end{aligned} \tag{2.9}$$

where a and h are the concentrations of generic Wnt and Dkk proteins, respectively; D_a and D_h are diffusion cofficients; ρ_a and ρ_h are reaction constants scaling the speed of the interaction between Wnt and Dkk; and μ_a and μ_h are decay constants (K_h and κ are additional reaction constants). Parameters were set arbitrarily, though it is noted that variations in their values do not qualitatively change behavior. The system was numerically simulated to create consecutive waves of hair follicle formation in a square grid of mouse skin.

The author then made experimental predictions based on modifications of their PDE model. Results from model simulation suggested that moderate overexpression of Wnt proteins increases follicular density, while strong overexpression completely disrupts patterning. In contrast, overexpression of Dkk in the model led to increased interfollicular spacing and clustering of new follicles in subsequent waves of formation (due to higher levels of Wnt around preexisting follicles). Transgenic expression of Dkk2 (another Wnt inhibitor) in mouse skin was used to test the latter two predictions. As expected, with increasing levels of inhibitor, hair follicle formation was impeded, leading to lower follicular density. A closer examination of mutant mouse skin demonstrated that follicle clusters were present, and ringlike patterns of Wnt signal-receiving cells were present around preexisting follicles. These results confirmed both model predictions concerning increased Wnt inhibitor concentration.

The computational and experimental results of this study provide compelling evidence for a reaction–diffusion mechanism for the Wnt signaling pathway. Though additional signaling pathways are known to be involved in follicular patterning, acting mostly downstream of formation, a model consisting of just two morphogens provides an accurate representation of experimentally observed behavior.

2.3.7.3 Outstanding Challenges in PDE Modeling

Simulation of PDE network models is substantially more computationally intensive than with their ODE counterparts, so a primary challenge is the development of mathematical approaches enabling the characterization of larger systems. Coarsegrained spatial methods like compartmental modeling can lessen the computational load in systems that are tolerant of reduced spatial resolution (de Jong 2002; Rangamani and Iyengar 2007). As with ODE systems, improvements in high-quality data collection will further enable the simulation of larger PDE models; Rangamani and Iyengar (2007) discusses techniques for experimentally estimating protein diffusion constants.

Another challenge is the efficient spatial modeling of complex cellular geometries. Most existing work treats cells and subcellular components as simple geometric shapes; though this simplifies computation, it may not be sufficiently accurate for certain model systems. Advances in finite volume modeling enable representation of and computation on arbitrary cell morphologies, which can lead to more realistic (and accurate) models.

2.3.8 SDE Systems

Both ODE and PDE models of biological systems assume that reactions occur in a deterministic manner. This assumption seems to imply that biological reactions exhibit little to no heterogeneity or stochasticity ("intrinsic noise"), which is known to be false (McAdams and Arkin 1997). Rather, the main reason for the success of deterministic biological models is that stochastic effects are often rendered negligible by averaging across large numbers of molecules or cells. This phenomenon also underlies the success of continuous mechanistic models, where discrete numbers of molecules can be approximated with continuous concentrations.

There are, however, a number of well-characterized biological systems where the modeling assumption of deterministic reactions leads to qualitatively incorrect depictions of behavior. A deterministic model of the circadian rhythm oscillator parameterized with particular degradation rates fails to oscillate; in contrast, the noise present in the corresponding stochastic model gives rise to more robust oscillatory behavior (Vilar et al. 2002). In a common class of biochemical reaction mechanisms, enzymatic futile cycles, extrinsic noise (i.e., noise due to components/processes outside the system) in a stochastic model was shown to induce bistable oscillatory behavior that was absent in a similar deterministic model (Samoilov et al. 2005).

These (and other) important exceptions to deterministic reaction mechanisms have led to the application of stochastic models to biological systems. In this section, we review continuous stochastic chemical kinetic models in the from of stochastic differential equations (SDEs). We focus mostly on a class of SDEs that can be derived from principles of discrete stochastic chemical kinetics; for illustrative purposes, our treatment starts with this derivation in brief (for more details see Gillespie 2000; El Samad et al. 2005; Gillespie 2007; Resat et al. 2009).

We begin by representing the state of the system as a function of time with $\mathbf{Z}(t) = [Z_i(t), \ldots, Z_n(t)]$, where $Z_i(t)$ represents the number of molecules of species i. Capital letters are used to emphasize the stochastic nature of the model; the Z_i's are random variables. A specific instantiation of the system is represented by lowercase letters; i. e., $\mathbf{z} = [z_i, \ldots, z_n]$. The system state can be altered by the firing of any of p reactions; each reaction changes the state by $\mathrm{v_k} = [v_{1\,k}, \ldots, v_{nk}]$, $1 \leq k \leq p$, where v_{ik} represents the change in the number of molecules of species i after the completion of reaction k. Each reaction can be characterized by its propensity function $a_k(\mathbf{z})$, defined so that $a_k(\mathbf{z})\,\mathrm{d}t$ is equivalent to the probability that reaction k will occur once in the system in the infinitesimal time interval $[t, t + \mathrm{d}t]$ given $\mathbf{Z}(t)$=$\mathbf{z}$. Given that any instantiation of the system is random, it would be useful to have a probabilistic expression for the time evolution of the system $P(\mathbf{z}, t|\mathbf{z}_0, t_0)$ (probability that the system is in state $\mathbf{z}$ at time t, given that it is in state $\mathbf{z}_0$ at time t_0). Using the above quantities and the laws of probability, this can be derived as follows:

$$\begin{aligned} P(\mathbf{z}, t + \mathrm{d}t|\mathbf{z}_0, t_0) = P(\mathbf{z}, t|\mathbf{z}_0 s\, t_0) \times \left(1 - \sum_{k=1}^{p} a_k(\mathbf{z})\,\mathrm{d}t\right) \\ + \sum_{k=1}^{p} P(\mathbf{z} - \mathbf{v}_k, t|\mathbf{z}_0, t_0) \times a_k(\mathbf{z} - \mathbf{v}_k)\,\mathrm{d}t. \end{aligned} \tag{2.10}$$

After rearranging and taking the limit as $\mathrm{d}t \to 0$

$$\frac{\partial P(\mathbf{z}, t|\mathbf{z}_0, t_0)}{\partial t} = \sum_{k=1}^{p} \left(a_k(\mathbf{z} - \mathbf{v}_k) P(\mathbf{z} - \mathbf{v}_k, t|\mathbf{z}_0, t_0) - a_k(\mathbf{z}) P(\mathbf{z}, t|\mathbf{z}_0, t_0)\right). \tag{2.11}$$

Equation (2.11) is called the *chemical master equation* (CME). Since the possible values of $\mathbf{z}$ are discretely varying, the CME is actually a set of coupled ODEs that is nearly as large as the number of possible combinations of molecules in the system. Consequently, except for very simple systems, these equations are not solvable analytically and numerical solutions are usually intractable. Progress has been made in developing approximation schemes for numerically solving the CME (Munsky and Khammash 2006; Deuflhard et al. 2007; Jahnke and Huisinga 2008), but most applications turn to Monte Carlo methods to sample from the distribution $P(\mathbf{z}, t|\mathbf{z}_0, t_0)$. The *stochastic simulation algorithm* (SSA), also known as the Gillespie algorithm, simulates each reaction sequentially as they occur in time (Gillespie 1977). This approach has been widely used in stochastic modeling of biological networks, in part because it produces a draw from the exact probability distribution that solves the CME. A few example applications of the SSA include McAdams and Arkin (1997), Samoilov et al. (2005), Arkin et al. (1998), Weinberger et al. (2005), El-Samad et al. El-Samad and Khammash (2006), and wang et al. (2006).

For systems with large numbers of molecules, use of the SSA becomes very computationally intensive. An efficient approximation known as *tau-leaping* was developed to instantiate multiple reactions that occur during the elapse of a

preselected time τ (Gillespie 2001). This gives the following (approximate) update equation for the system state, given that $\mathbf{Z}(t) = \mathbf{z}$ at time t:

$$\mathbf{Z}(t+\tau) \approx \mathbf{z} + \sum_{k=1}^{p} \mathbf{v}_k \times \mathrm{Pois}_k(a_k(\mathbf{z})\tau), \tag{2.12}$$

where the $\mathrm{Pois}_k(\lambda)$ are i.i.d. Poisson random variables with mean (and variance) λ. Subsequent work in approximate methods has led to further speed up of the SSA; some example include Rao and Arkin (2003), Rathinam and El Samad (2007), Cao et al. (2005, 2007), and Cao and Petzold (2008). Discrete stochastic kinetic models are the most common stochastic approach for modeling biological networks, and they have previously been reviewed extensively (Li et al. 2008; Higham 2008; Resat et al. 2009 Wilkinson 2006, 2009). As our focus in this chapter is continuous differential equation models, we will not discuss them further.

Under certain conditions (see Section 3.8.1), we can approximate a Poisson random variable with one that is normally distributed (with the same mean and variance), yielding

$$\begin{aligned}
\mathbf{Z}(t+\tau) &\approx \mathbf{z} + \sum_{k=1}^{p} \mathbf{v}_k \times N_k(a_k(\mathbf{z})\tau, a_k(\mathbf{z})\tau) \\
&= \mathbf{Z} + \sum_{k=1}^{p} \mathbf{v}_k \left(a_k(\mathbf{z})\tau + \sqrt{a_k(\mathbf{z})} \times N_k(0,\tau)\right) \\
&= \mathbf{Z} + \sum_{k=1}^{p} \mathbf{v}_k a_k(\mathbf{z})\tau + \sum_{k=1}^{p} \mathbf{v}_k \sqrt{a_k(\mathbf{z})} \times N_k(0,\tau),
\end{aligned} \tag{2.13}$$

where $N_k\left(\mu, \sigma^2\right)$ is a normally distributed random variable with mean μ and variance σ^2. Equation (2.13) is a form of the *Langevin leaping formula*, where the discretely valued $\mathbf{Z}(t)$ has become continuously valued due to the normal approximation. To convert (2.13) to a differential equation, we note that the increment $B_k(t+\tau) - B_k(t)$, where B_k is a standard Brownian motion, is normally distributed with mean 0 and variance τ (Karlin and Taylor 1975). Thus, by rewriting τ as dt and rearranging we have the approximate relation (given that $\mathbf{Z}(t) = \mathbf{z}$):

$$\begin{aligned}
\frac{\mathrm{d}\mathbf{Z}(t)}{\mathrm{d}t} &\approx \sum_{k=1}^{p} \mathbf{v}_k a_k(\mathbf{Z}(t)) + \sum_{k=1}^{p} \mathbf{v}_k \sqrt{a_k(\mathbf{Z}(t))} \times \frac{\mathrm{dB}_{k(t)}}{\mathrm{d}t} \\
&= \sum_{k=1}^{p} \mathbf{v}_k a_k(\mathbf{Z}(t)) + \sum_{k=1}^{p} \mathbf{v}_k \sqrt{a_k(\mathbf{Z}(t))} \times \mathrm{W}_k(t),
\end{aligned} \tag{2.14}$$

where $\mathrm{W}_k(t)$ is a *white noise process*, which, though not well-defined in an ordinary calculus sense, can be characterized using stochastic calculus and acts as a useful approximation for naturally occurring noise (Karlin and Taylor 1981). Equation (2.14) is an SDE known as the *chemical Langevin equation* (CLE).

It is useful to make one final approximation to establish a connection between the CLE and the RRE (Section 2.3.6). If the volume of the system and the number of molecules of each species $Z_i(t)$ approach infinity in such a way that the concentration of each species remains constant (*thermodynamic limit*), the reaction propensities grow linearly with the system size (Gillespie 2007; El Samad et al. 2005). Referring to (2.14), we see that this will result in the second (noise) term becoming negligibly small with respect to the first (drift) term. If we set the noise term to zero and divide each state variable by the system volume, random molecule counts ($Z_i(t)$'s) become deterministic concentrations ($y_i(t)$'s) and we have

$$\frac{\mathrm{d}\mathbf{y}(t)}{\mathrm{d}t} \approx \sum_{k=1}^{p} \mathbf{v}_k a_k(\mathbf{y}(t)). \tag{2.15}$$

This is a form of the RRE (1) which we have derived using principles of stochastic chemical kinetics.

2.3.8.1 Assumptions of SDE Biological Network Models

We now review the assumptions needed to derive and subsequently apply SDEs to modeling biological systems. They include

- As with ODE models, reactions occur in a homogeneous, well-stirred volume.
- There exist small-enough values for leap times τ such that, as the system evolves with each time leap, no propensity function changes by an appreciable amount as a result of executed reactions (permits approximation in Eq. (2.12)).
- The τ times from above are also large enough that the expected number of occurrences of each reaction during each leap is much greater than one (permits approximation in Eq. (2.13)).

The latter two assumptions also permit the replacement of τ with dt (*macroscopic infinitesimal*) in Eq. (2.14) (Gillespie 2002). These assumptions can usually be simultaneously satisfied if the numbers of molecules of each reacting species are sufficiently large (Gillespie 2002; 2007; El Samad et al. 2005). The major advantage of working with the CLE is that numerical solutions are much more efficient to generate than when using the Gillespie algorithm (Adalsteinsson et al. 2004), provided the above assumptions are satisfied, the solutions should not differ appreciably.

2.3.8.2 Modern Application of SDE Models to Biological Networks

Use of SDE models in system biology is even less common than PDEs, with most examples emerging in just the last few years. We thus focus immediately on recent applications to modeling biological networks, highlighting one example per dynamical modeling task.

Bayesian Calibration of an SDE Model from Noisy Time Course Measurements

One challenge in calibrating a dynamical model with time course measurements is the often coarse time resolution of the data. This is particularly true for SDE models, where approximations that assist in learning parameter values from data require measurements collected at uniformly closely spaced time points. Heron et al. illustrate this difficulty with an SDE model of the Hesl autoregulatory network, which takes the following (differential) form (Heron et al. 2007):

$$\begin{aligned} \mathrm{d}M(t) &= \left(\frac{s_M v_1}{1+\left(D\left[P(t)\right]/s_P k_1\right)^n} - v_2 M(t)\right)\mathrm{d}t \\ &\quad + \left(\sqrt{s_M}\sqrt{\frac{s_M v_1}{1+\left(D\left[P(t)\right]/s p k_1\right)^n} + v_2 M(t)}\right)\mathrm{d}B_M(t), \\ \mathrm{d}P(t) &= \left(\frac{s_P}{s_M} v_3 M(t) - v_4 P(t)\right)\mathrm{d}t \\ &\quad + \left(\sqrt{s_P}\sqrt{\frac{s_P}{s_M} v_3 M(t) + v_4 P(t)}\right)\mathrm{d}B_P(t), \end{aligned} \tag{2.16}$$

where $M(t)$ and $P(t)$ represent the relative quantities of Hesl mRNA and protein, respectively; $D[P(t)]$ represents a delay term acting between transcription and translation; $\mathrm{d}B_M(t)$ and $\mathrm{d}B_P(t)$ are independent infinitesimal increments of one dimensional Brownian motion; $\{v_1, v_2, v_3, v_4, n, k_1\}$ are reaction parameters; and s_M and s_P are scaling factors.

In order to learn parameter values from the data, an objective function is required that describes how well a given set of parameter values fit the data. As SDE models are by nature probabilistic, a likelihood-based approach is a natural choice. With small enough time steps $\Delta t_i = t_i + 1 - t_i$, the increments $M(t_i + 1) - M(t_i)$ and $P(t_i + 1) - P(t_i)$ are normally distributed with means and variances derived from Eq. (2.16). This property allows the calculation of a likelihood function for parameter values given discrete data. When the time course data do not differ by small enough time steps (as is true for the Hesl data used in Heron et al. 2007), this likelihood function cannot be used directly.

To compensate, the authors employ a latent data-based Bayesian approach, whose application to systems biology was first described in Golightly and Wilkinson (2005, 2006). The method infers unobserved (latent) data at finer time points than those measured, the presence of which satisfies the assumptions of the likelihood function. MCMC methods are used to sample distributions of the latent data and parameters simultaneously, and parameter values can be chosen which maximize the Bayesian posterior probability function.

Heron et al. applied this method to sparse Hes1 data from Hirata et al. (2002) and obtained convergent posterior distributions for all parameters except the scaling factors. They simulated the model with a wide range of high likelihood parameter values and consistently recovered the cyclicity observed in the experimental data. Thus, behavior of the calibrated Hes 1 model appears to be robust.

In an ideal world, experimental data would always be available at sufficient resolution for straightforward model calibration. As modeling is used to study more and more complex signaling networks, this ideal becomes less realistic. In the increasingly likely scenario where experimental data are sparse, probabilistic methods like the one described in this study will be useful for robust model calibration.

Comparison of a Stochastic Cell Cycle Model with Experimental data

The eukaryotic cell cycle has been a popular choice for systems biology modeling due to its fundamental role in development and reproduction. During this ordered sequence of molecular events, a cell duplicates its components and divides them into two daughter cells. In the fission of yeast *Schizosaccharomyces pombe*, fundamental players in this process include the M (mitotic)-phase promoting factor (MPF), its negative regulator Weel, and a collection of cell division cycle (Cdc) proteins (i.e., Cdc25, which leads to increased levels of MPF).

Over the years, more and more elaborate (and realistic) models have been built describing this system, most of them deterministic. In Steuer (2004), the authors adapt a previously published yeast ODE model (Novak et al. 2001) to include stochastic influences. Written as Langevin-type SDEs with multiplicative noise, the equations take the following form:

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = f_i(\cdots) + \sqrt{2D_i x_i} W_i(t), \tag{2.17}$$

where x_i represents the concentration of a single species, $f_i(\ldots)$ comes from the original deterministic equation, D_i is a constant denoting the noise amplitude, and $W_i(t)$ is white noise. The second term on the RHS of Eq. (2.17) is not derived from elementary biochemical reactions; rather, it constitutes a general noise term including both intrinsic and extrinsic sources. The amplitude of the noise is controlled by the parameter D_i.

After parameterizing their model with the values used in the original work (and a small but constant value for all D_i), the authors run simulations of both the original ODE and their SDE system. They compare the resulting behavior to well-known experimental observation of cell cycle time and cell division size distributions. Both models reproduce the negative correlation seen in wild-type cells between cycle time and mass at birth (representing a form of cell size control). The authors then evaluate simulations of a $wee1^{-}cdc25\Delta$ double mutant, whose behavior has been shown experimentally to result in three to four clusters in the cycle time vs. mass at birth plot. The deterministic model does not generate clustered behavior, whereas the stochastic model consistently does. Mechanistically, this is due to the occasional inability of MPF levels to reach the threshold needed for entry into mitosis, as a result of Cdc25 absence and stochastic fluctuations in Pyp3 activity (a weaker MPF activator). These fluctuations can lead to a random number of G2 phase resets, which in turn lead to varying but quantized cycle times.

Thus, an SDE depiction of the cell cycle appears to be more realistic than a deterministic model. The authors go on to characterize the cell size control checkpoint in their stochastic model, identifying noise-induced oscillations which occur at sufficiently (but not too) large amplitudes of noise (so-called "coherence resonance"). These oscillations have not as of yet been detected experimentally, suggesting additional regulatory mechanisms that stabilize this behavior.

Sensitivity of Cell Cycle Behavior to Intrinsic vs. Extrinsic Noise

More recently, Yi et al. investigated the separate effects of intrinsic and extrinsic noise on the *S. pombe* cell cycle (Yi et al. 2008). Unlike in Steuer (2004), Yi et al. derived a CLE from elementary reactions in the spirit of Section 2.3.8. Thus, the noise modeled in their SDE system is exclusively intrinsic in origin. The authors parameterize their model in the same manner as above, and they numerically simulate cell cycle behavior of the $wee1^{-}cdc25\Delta$ double mutant. Using a small-enough system size to exhibit fluctuations, the results look quite similar to those resulting from the SDE in Eq. (2.17) (which combines intrinsic and extrinsic noise), with visible clustering present in the cycle time vs. mass at birth plot.

To test the sensitivity of system behavior to noise that is of extrinsic origin, the authors added noise to the parameter governing Pyp3 activation of MPF (see above), turning it into a random variable. They inserted this parameter into the ODE model from Novak et al. (2001) and again performed a simulation on the double mutant strain. Clustering was still observed in the cycle time vs. mass plot, although the number of cells with long cycle time is markedly reduced.

Finally, the authors incorporate the above parametric noise into their SDE model to explicitly test the effects of both intrinsic and extrinsic noise. Once again, clustering in the cycle time plot was present, with the number of cells having longer cycle times resembling its original quantity.

In this study, the behavior of the fission yeast cell cycle model demonstrates sensitivity to the source of noise, with intrinsic noise giving rise to more cells with long cell cycle times. The precise cause of this difference was not investigated, and further experiments would be expected to clarify the relative importance of the two noise sources in this system. It is noteworthy that both sources of noise led to behavior that is qualitatively different from a deterministic model (and similar to experimental observations), which highlights the importance of modeling stochastic effects. Though most biological networks are subject to both intrinsic and extrinsic noise, modeling approaches that separate the two can be useful for understanding the mechanisms and consequences of noise generation.

Oscillatory Behavior Due to Coherence Resonance in a Stochastic Model of Circadian Rhythm

As mentioned above, the phenomenon of *coherence resonance* (CR) describes emergent behavior (e.g., oscillations) due to an optimal amount of noise present in the

system. In CR, noise amplitudes lower or higher than the optimum diminish the oscillatory behavior. Yi et al. (2006) characterize this phenomenon as a function of both intrinsic and extrinsic noise in an SDE model of the Drosophila circadian oscillator.

Their model includes two proteins, PER and dCLOCK, which combine to from both a negative and positive feedback loops. Exposure to light causes the degradation of PER, which bestows on the system a circadian rhythm of 24 h. The authors derive a CLE from elementary chemical reactions outlined in Smolen et al. (2002), resulting in a stochastic model with intrinsic noise. The relevant SDEs are as follows:

$$\begin{aligned}\frac{\mathrm{d}P(t)}{\mathrm{d}t} &= \left[v_{sp}\frac{L_{\text{free}}(t-\tau_1)}{K_1+L_{\text{free}}(t-\tau_1)} - k_{dp}P(t)\right] \\ &\quad + \frac{1}{\sqrt{V}}\left[\sqrt{v_{sp}\frac{L_{\text{free}}(t-\tau_1)}{K_1+L_{\text{free}}(t-\tau_1)}}W_1(t) - \sqrt{k_{dp}P(t)}W_2(t)\right], \\ \frac{\mathrm{d}P(t)}{\mathrm{d}t} &= \left[v_{sc}\frac{K_2}{K_2+L_{\text{free}}(t-\tau_2)} - k_{dc}L(t)\right] \\ &\quad + \frac{1}{\sqrt{V}}\left[\sqrt{v_{sc}\frac{K_2}{K_2+L_{\text{free}}(t-\tau_2)}}W_3(t) - \sqrt{k_{dc}L(t)}W_4(t)\right],\end{aligned} \tag{2.18}$$

where $P(t)$ and $L(t)$ represent the concentrations of the PER and dCLOCK proteins, respectively; $L_{\text{free}} = \max(L-P,0)$; τ_1 and τ_2 are time delays; $\{v_{sp}, v_{sc}, K_1, K_2, k_{dp}, k_{dc}\}$ are kinetic parameters, V is the system volume, and $W_1(t)$ and $W_2(t)$ are independent white noise. The authors parameterized the model according to Smolen et al. (2002), and they carried out simulations with $k_{dp} = 2.85\,\mathrm{h}^{-1}$ (light-induced degradation rate of PER), which in the deterministic system induces a non-oscillatory steady state.

By varying the system volume V, Yi et al. could control the effects of intrinsic noise (see Eq. (2.18)) and as a result, the amplitude of noise-induced oscillations. For a moderate system size, oscillations (with period ~24 h) were most clearly observed. The authors created a metric β, which measures the strength of CR oscillation in terms of *signal-to-noise ratio* (SNR), and they discovered that optimal oscillatory behavior resulted when $V = 500$.

Yi et al. then added extrinsic noise to the system by making k_{dp} a random variable with a mean of 2.85 h^{-1}. They used a parameter D to control the strength of the noise, and they studied the behavior of the modified SDE as a function of both V and D. They discovered that the highest values of β were achieved when V was very large and $D \approx 0.04$, which corresponds to no intrinsic noise and an optimal level of extrinsic noise. This value of β was roughly 10-fold higher than the maximum achieved above with only intrinsic noise in the system. In general, increasing levels of intrinsic noise in a system already under the influence of extrinsic noise

reduced oscillatory behavior, whereas increasing extrinsic noise in a system already exhibiting intrinsic noise-induced CR greatly increased the oscillations.

This study provides an initial characterization of noise-induced oscillations that arise in a well-characterized biological system. Future work will include uncovering the mechanisms involved in differential sensitivity of CR to noise sources. In addition, as the intrinsic and extrinsic noise sources in the above model were simulated to be independent, the authors plan to study the potential coupling effects of the two types of noise. Such effects would be expected to occur in biological systems where the number of reactant molecules is very small.

Prediction of Noise-Induced Bistability in an Enzymatic Futile Cycle

Enzymatic futile cycles are a ubiquitous control mechanism in biological systems consisting of two reactions: conversion of a substrate to a product via a forward enzyme and conversion of the product back to the substrate via a reverse enzyme. This motif is present in diverse signaling processes including MAPK cascades, cell division cycles, and stress response pathways. Deterministic models of futile cycles have demonstrated that they can act as molecular switches that convert continuous input signals to binary outputs as well as signal amplifiers.

To characterize the effects of noise on enzymatic futile cycles, Samoilov et al. developed an SDE model that adds an extrinsic noise term to the deterministic ODE, yielding the following differential form (Samoilov et al. 2005):

$$dX^* = \left[\frac{k_+E_+X}{k_+ + X} - \frac{k_-E_-\,(X_0 - X)}{K_- + X_0 - X}\right]\mathrm{d}t + \frac{\sigma_+E_+^p k_+X}{K_+ + X}\,\mathrm{d}B(t)\,, \tag{2.19}$$

where X and X^* represent the concentrations of the substrate and product, respectively; $X_0 = X + X^*, E_+$, and E_- are the concentrations of the forward and reverse enzymes, respectively; k_+ and k_- are reaction rate constants; k_+ and k_- are Michaelis–Menten constants; σ_+ and P parameterize the extrinsic noise; and $\mathrm{d}B(t)$ is an infinitesimal increment of Brownian motion. The authors solve for the stationary-state response curve $R(.)$, which gives

$$\begin{aligned} R\,(X_{ss}, E_+; E_-) =& E_+ - \frac{k_-E_-\,(X_0 - X_{ss})\,(K_+ + X_{ss})}{k_-X_{ss}\,(K_- + X_0 - X_{ss})} + \frac{\sigma_+E_+^p k_+K_+}{(K_+ + X_{ss})^2}, \\ =& 0 \end{aligned} \tag{2.20}$$

where X_{ss} is the steady-state concentration of X. This equation is fourth order in X_{ss}, which allows for a bistable solution (unlike the deterministic equivalent which is missing the second term in Eq (2.20)). Thus, a prediction of the SDE model is that X_{ss} can be multivalued. Further analysis by the authors suggests that the onset of bistability occurs with $20 \leq El+ \leq 30, p > 0.75$, and $\sigma_+ \approx 20\%$.

Samoilov et al. tested their bistability predictions by constructing a discrete stochastic model consisting of elementary biochemical reactions. Stochasticity in this system only arises through fluctuations in the individual components. Using the

Gillespie algorithm, the authors simulated trajectories of the model and confirmed the above predictions on bistability, suggesting that the form of extrinsic noise in the SDE formalism is sufficiently realistic to predict system behavior.

Ideally, experimental evidence would corroborate the existence of noise-induced bistability in a real-world enzymatic futile cycle. Unfortunately, to our knowledge no such evidence has yet been discovered, perhaps due in part to the difficulty of separating noise-induced behavior from measurement error. Nonetheless, as the authors suggest, given the ubiquity of enzymatic futile cycles in nature "it is reasonable to assume that such a behavior is exploited in at least some cellular systems."

2.3.8.3 Outstanding Challenges in SDE Modeling

SDE modeling of a signaling network requires the existence of a timestep τ that acts as a macroscopic infinitesimal (Section 2.3.8.1). If this assumption is violated, more computationally intensive (but accurate) discrete stochastic models must be used. Thus, an active area of research is a modeling framework that combines features of both models, partitioning the system into discrete and continuous components as necessary (Bentele and Eils 2005; Salis and Kaznessis 2005).

Another challenge arises when using probabilistic or MCMC techniques for SDE model calibration, as in Heron et al. (2007). With large enough models, these methods become computationally intractable, suggesting the need for a more efficient strategy. One approach, borrowed from weather and climate modeling, involves creating an approximate surrogate of the model called an *emulator*. Parameter estimation can then be performed more cheaply on the emulator than on the full model (Wilkinson 2009). Though few applications of this method exist so far (Henderson et al. 2009 demonstrates one), calibration approaches like these will be necessary as biological models increase in complexity.

2.3.9 Relevant Software

Several excellent reviews have been written detailing available software packages for constructing and simulating ODE, PDE, and SDE models of biological networks. Rather than reproduce that information here, we refer the reader to these articles: Gilbert et al. (2006); Resat et al. (2009); You (2004). Of particular note is SBML (Hucka et al. 2003), a data format for representing systems biology models; data files for many of the models discussed in this review are available at `http://www.sbml.org`.

2.3.10 Hybrid Dynamical Models of Biological Systems

As model systems continue to grow in size and complexity, hybrid models will become more important. A hybrid approach is computationally advantageous in that it can limit expensive procedures (e.g., stochastic versus deterministic modeling) to

system subsets where those procedures are necessary. The remaining parts of the system can be evaluated using computationally cheaper methods without appreciable losses in overall accuracy. One example of a hybrid dynamical model was given in Section 2.3.8.3, combining a continuous SDE and discrete stochastic model.

Another example of a hybrid scheme combines spatial with stochastic modeling, creating a so-called "spatial Langevin" system (Andrews and Arkin 2006; Elf and Ehrenberg 2004). This provides arguably the most realistic modeling framework short of molecular dynamics simulations. Several of the software packages listed above allow implementation of multiple types of hybrid modeling schemes.

Hybrid modeling can also be used to incorporate coarser-grained, non-dynamical approaches like flux-balance analysis (FBA). FBA uses the steady-state assumption to model reaction fluxes as a system of linear equations. Because the system is assumed to be at steady state, these equations are algebraic and can be solved efficiently using linear programming methods. Covert et al. create such a hybrid model of *E. coli* which combines FBA with boolean and ODE systems (Covert et al. 2008). As the systems biology community moves toward whole cell models, which in eukaryotic organisms could contain $> 10^{12}$ reactions (Resat et al. 2009), hybrid models will be essential in their efficient simulation and characterization.

2.4 Conclusions

With thousands of sequenced genomes (Wheeler et al. 2007) and hundreds of functional genomic data sets (Barrett et al. 2005), the future of systems biology is bright.

In static modeling, the supervised learning approach, in which high-throughput data is compared against a small training set of curated knowledge, has proven to be the most fruitful data integration strategy to date. In particular, supervised predictions of function and interaction from multiple data sets are more robust than those derived from individual data sets and have provided a foundation for recent work on network alignment and systematic validation. The primary challenges for static modeling are to (1) decide on a set of reference networks and (2) tie every predicted node and edge in such networks to a gold-standard experimental test such as co-immunoprecipitation for confirmation of physical protein interactions. These steps will be crucial to bringing network predictions to the same level of confidence and widespread utilization as gene predictions.

For dynamic models, the core problem is that the area will remain data starved (Albeck et al. 2006) until high-throughput methods for the determination of rate constants (Famili et al. 2005) and spatial substructure (Foster et al. 2006; Schubert et al. 2006) become commonplace. Recent efforts at compiling and curating a number of biological constants (Milo et al. 2009) and developing a repository of systems biology models (Hucka et al. 2003) are an important step in the right direction toward establishing a repository of "consensus constants."

Ultimately the relevance of both kinds of models is directly proportional to their ability to predict experiments. In particular, the use of a framework (likely Bayesian) for smoothly incorporating new measurements into updated parameter estimates is likely to be of central importance. We also believe that the incorporation of tools from system identification (Nelles 2000) and parameter estimation will prove useful in the years to come. In short, now that the mathematical aspects of the field have matured, from this point forward we expect experimental testability to increasingly become the focus of the field, with models specifically formulated to be updated as new experimental data arrives.

Acknowledgements We thank Russ Altman for helpful discussions.

References

Abecasis G, Tam P, Bustamante C, et al (2007) Human genome variation 2006; emerging views on structural variation and large-scale SNP analysis. Nat Genet 39(2):153–155

Adalsteinsson D, McMillen D, Elston T (2004) Biochemical network stochastic simulator (BioNetS): software for stochastic modeling of biochemical networks. BMC Bioinformatics 5

Aerts S, Lambrechts D, Maity S, et al (2006) Gene prioritization through genomic data fusion. Nat Biotechnol 24(5):537–544

Albeck JG, MacBeath G, White FM, et al (2006) Collecting and organizing systematic sets of protein data. Nat Rev Mol Cell Biol 7(11):803–812

Albeck JG, Burke JM, Aldridge BB, et al (2008a) Quantitative analysis of pathways controlling extrinsic apoptosis in single cells. Mol Cell 30(1):11–25

Albeck JG, Burke JM, Spencer SL, et al (2008b) Modeling a snap-action, variable-delay switch controlling extrinsic cell death. PLoS Biol 6(12):2831–2852

Aldridge BB, Burke JM, Lauffenburger DA, et al (2006a) Physicochemical modelling of cell signalling pathways. Nat Cell Biol 8(11):1195–1203

Aldridge BB, Haller G, Sorger PK, et al (2006b) Direct Lyapunov exponent analysis enables parametric study of transient signalling governing cell behaviour. IEE Proceedings Syst Biol 153(6):425–432

Alon U, Surette MG, Barkai N, et al (1999) Robustness in bacterial chemotaxis. Nature 397(6715):168–171

Altman RB, Raychaudhuri S (2001) Whole-genome expression analysis: challenges beyond clustering. Curr Opin Struct Biol 11(3):340–347

Amonlirdviman K, Khare N, Tree D, et al (2005) Mathematical modeling of planar cell polarity to understand domineering nonautonomy. Science 307(5708):423–426

Andrews SS, Arkin AR (2006) Simulating cell biology. Curr Biol 16(14):R523–R527

Angeli D, Ferrell J, Sontag E (2004) Detection of multistability, bifurcations, and hysteresis in a large class of biological positive-feed back systems. Proc Natl Acad Sci USA 101(7): 1822–1827

Arkin A, Ross J, McAdams H (1998) Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected Escherichia coli cells. Genetics 149(4):1633–1648

Asai R, Taguchi E, Kame Y, et al (1999) Zebrafish Leopard gene as a component of the putative reaction-diffusion system. Mech Dev 89(1–2):87–92

Ashburner M, Ball CA, Blake JA, et al (2000) Gene Ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25(1):25–29

Axelrod JD (2001) Unipolar membrane association of Dishevelled mediates Frizzled planar cell polarity signaling. Genes Dev 15(10):1182–7

Bader GD, Cary MP, Sander C (2006) Pathguide: a pathway resource list. Nucleic Acids Res 34(Database issue)

Baker RE, Gaffney EA, Maini PK (2008) Partial differential equations for self-organization in cellular and developmental biology. Nonlinearity 21(11):R251–R290

Barabasi AL, Oltvai ZN (2004) Network biology: understanding the cell's functional organization. Nat Rev Genet 5(2):101–113

Barrett C, Palsson B (2006) Iterative reconstruction of transcriptional regulatory networks: an algorithmic approach. PLoS Comput Biol 2(5):e52

Barrett T, Suzek TO, Troup DB, et al (2005) NCBI GEO: mining millions of expression profiles–database and tools. Nucleic Acids Res 33(Database issue)

Bastock R, Strutt H, Strutt D (2003) Strabismus is asymmetrically localised and binds to Prickle and Dishevelled during Drosophila planar polarity patterning. Development 130(13): 3007–14

Batzoglou S (2005) The many faces of sequence alignment. Brief Bioinform 6(1):6–22

Beckett D, Berners-Lee T (2007) RDF Primer, Turtle Version, www.w3.org/TeamSubmission/turtle. Accessed 31 Aug 2009

Behar M, Hao N, Dohlman HG, et al (2008) Dose-to-duration encoding and signaling beyond saturation in intracellular signaling networks. PLoS Comput Biol 4(10)

Ben-Hur A, Noble WS (2006) Choosing negative examples for the prediction of protein-protein interactions. BMC Bioinform 7 Suppl 1:S2

Benson G (2009) Nucleic Acids Research annual Web Server Issue in 2009. Nucl Acids Res 37(suppl_2):W1–2

Bentele M, Eils R (2005) General stochastic hybrid method for the simulation of chemical reaction processes in cells. Comput Meth Syst Biol 3082:248–251

Berg J, Lassig M (2006) Cross-species analysis of biological networks by Bayesian alignment. Proc Natl Acad Sci USA 103(29):10,967–72

Beyer A, Workman C, Hollunder J, et al (2006) Integrated assessment and prediction of transcription factor binding. PLoS Computational Biol 2(6):e70

Bhalla US, Iyengar R (1999) Emergent properties of networks of biological signaling pathways. Science 283(5400):381–387

Bloom JD, Adami C (2003) Apparent dependence of protein evolutionary rate on number of interactions is linked to biases in protein-protein interactions data sets. BMC Evol Biol 3

Bodenreider O (2004) The Unified Medical Language System (UMLS): integrating biomedical terminology. Nucl Acids Res 32(suppl_1):D267–270

Bonneau R (2008) Learning biological networks: from modules to dynamics. Nat Chem Biol 4(11):658–64

Bornholdt S (2005) Systems Biology: less is more in modeling large genetic networks. Science 310(5747):449–451

Breitkreutz BJ, Stark C, Tyers M (2003) Osprey: a network visualization system. Genome Biol 4(3):R22

Brewer D, Barenco M, Callard R, et al (2008) Fitting ordinary differential equations to short time course data. Philos Trans Ro Soc A Math Phys Eng Sci 366(1865):519–544

Brudno M, Do CB, Cooper GM, et al (2003) LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. Genome Res 13(4):721–731

Cao Y, Petzold L (2008) Slow-scale tau-leaping method. Comput Meth Appl Mech Eng 197 (43–44):3472–3479

Cao Y, Gillespie D, Petzold L (2005) The slow-scale stochastic simulation algorithm. J Chem Phys 122(1)

Cao Y, Gillespie DT, Petzold LR (2007) Adaptive explicit-implicit tau-leaping method with automatic tau selection. J Chem Phy 126(22)

Carrera J, Rodrigo G, Jaramillo A (2009) Model-based redesign of global transcription regulation. Nucleic Acids Res 37(5):e38

Champoux JJ (2001) DNA topoisomerases: structure, function, and mechanism. Annu Rev Biochem 70:369–413

Chen WW, Schoeberl B, Jasper PJ, et al (2009) Input-output behavior of ErbB signaling pathways as revealed by a mass action model trained against dynamic data. Mol Syst Biol 5

Chen X, Wu JM, Homischer K, et al (2006) TiProD: the Tissue-specific Promoter Database. Nucleic Acids Res 34(Database issue)

Collins S, Miller K, Maas N, et al (2007) Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. Nature 446(7137):806–810

Cornell-Bell AH, Finkbeiner SM, Cooper MS, et al (1990) Glutamate induces calcium waves in cultured astrocytes: long-range glial signaling. Science 247(4941):470–3

Cornish-Bowden A (1979) Fundamentals of Enzyme Kinetics. Butterworths

Covert MW, Knight EM, Reed JL, et al (2004) Integrating high-throughput and computational data elucidates bacterial networks. Nature 429(6987):92–96

Covert MW, Xiao N, Chen TJ, et al (2008) Integrating metabolic, transcriptional regulatory and signal transduction models in Escherichia coli. Bioinformatics 24(18):2044–50

Dandekar T, Schuster S, Snel B, et al (1999) Pathway alignment: application to the comparative analysis of glycolytic enzymes. Biochem J 343 Pt 1:115–124

von Dassow G, Meir E, Munro EM, et al (2000) The segment polarity network is a robust developmental module. Nature 406(6792):188–192

Davidson EH, Rast JP, Oliveri P, et al (2002) A genomic regulatory network for development. Science 295(5560):1669–1678

Mrvar A, Batagelj V (2005) Exploratory social network analysis with Pajek. Cambridge University Press, Cambridge

Deeds EJ, Ashenberg O, Shakhnovich EI (2006) A simple physical model for scaling in protein-protein interaction networks. Proc Natl Acad Sci USA 103(2):311–316

Degnan JH, Rosenberg NA (2006) Discordance of species trees with their most likely gene trees. PLoS Genet 2(5)

Demello A (2006) Control and detection of chemical reactions in microfluidic systems. Nature 442(7101):394–402

Deuflhard P, Huisinga W, Jahnke T, et al (2007) Adaptive discrete Galerkin methods applied to the chemical master equation. SIAM J Sci Comp 30(6):2990–3011

Do C, Gross S, S B (2006a) CONTRAlign: discriminative training for protein sequence alignment. Proceedings of the tenth annual international conference on computational molecular biology, (RECOMB 2006) pp 160–164

Do CB, Woods DA, Batzoglou S (2006b) CONTRAfold: RNA secondary structure prediction without physics-based models. Bioinformatics 22(14)

Dudley A, Janse D, Tanay A, et al (2005) A global view of pleiotropy and phenotypically derived gene function in yeast. Mol Syst Biol 1(1):msb4100,004–E1–msb4100,004–E11

Durbin R, Eddy S, Krogh A, et al (1999) Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids. Cambridge University Press, Cambridge

Eilbeck K, Lewis SE, Mungall CJ, et al (2005) The Sequence Ontology: a tool for the unification of genome annotations. Genome Biol 6(5):R44

El-Samad H, Khammash M (2006) Regulated degradation is a mechanism for suppressing stochastic fluctuations in gene regulatory networks. Biophys J 90(10):3749–3761

El Samad H, Khammash M, Petzold L, et al (2005) Stochastic modelling of gene regulatory networks. Int J Robust Nonlinear Control 15(15):691–711

El-Samad H, Kurata H, Doyle J, et al (2005) Surviving heat shock: Control strategies for robustness and performance. Proc Natl Acad Sci USA 102(8):2736–2741

Elf J, Ehrenberg M (2004) Spontaneous separation of bi-stable biochemical systems into spatial domains of opposite phases. Syst Biol (Stevenage) 1(2):230–236

Ellson J, North S (2007) Graphviz: Graph Visualization Software. www.graphviz.org. Accessed 31 Aug 2009

ENCODE Project Consortium (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 447(7146):799–816

Eungdamrong N, Iyengar R (2004) Modeling cell signaling networks. Biol Cell 96(5): 355–362

Ewing B, Green P (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res 8(3):186–194

Ewing B, Hillier L, Wendl MC, et al (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res 8(3):175–185

Famili I, Mahadevan R, Palsson B (2005) k-Cone Analysis: Determining All Candidate Values for Kinetic Parameters on a Network Scale. Biophys J 88(3):1616–1625

Faure A, Naldi A, Chaouiya C, et al (2006) Dynamical analysis of a generic Boolean model for the control of the mammalian cell cycle. Bioinformatics 22(14):E124–E131

Fernandez J, Hoffmann R, Valencia A (2007) iHOP web services. Nucleic Acids Res

Flaherty P, Giaever G, Kumm J, et al (2005) A latent variable model for chemogenomic profiling. Bioinformatics 21(15):3286–3293

Flaherty P, Radhakrishnan ML, Dinh T, et al (2008) A dual receptor crosstalk model of G-protein-coupled signal transduction. PLoS Comput Biol 4(9)

Flannick J, Novak A, Srinivasan BS, et al (2006) Graemlin: general and robust alignment of multiple large interaction networks. Genome Res 16(9):1169–1181

Forst CV, Schulten K (2001) Phylogenetic analysis of metabolic pathways. J Mol Evol 52(6): 471–489

Foster L, de Hoog C, Zhang Y, et al (2006) A Mammalian Organelle Map by Protein Correlation Profiling. Cell 125(1):187–199

Galperin MY, Cochrane GR (2009) Nucleic acids research annual database issue and the NAR online Molecular Biology Database Collection in 2009. Nucl Acids Res 37(suppl_1):D1–4

Gandhi TKB, Zhong J, Mathivanan S, et al (2006) Analysis of the human protein interactome and comparison with yeast, worm and fly interaction datasets. Nat Genet 38(3):285–293

Gavin AC, Aloy P, Grandi P, et al (2006) Proteome survey reveals modularity of the yeast cell machinery. Nature 440(7084):631–636

Giaever G, Chu AM, Ni L, et al (2002) Functional profiling of the Saccharomyces cerevisiae genome. Nature 418(6896):387–391

Gilbert D, Fuss H, Gu X, et al (2006) Computational methodologies for modelling, analysis and simulation of signalling networks. Brief Bioinform 7(4):339–353

Gillespie D (1977) Exact stochastic simulation of coupled chemical-reactions. J Phys Chem 81(25):2340–2361

Gillespie D (2000) The chemical Langevin equation. J Chem Phys 113(1):297–306

Gillespie D (2001) Approximate accelerated stochastic simulation of chemically reacting systems. J Chem Phys 115(4):1716–1733

Gillespie D (2002) The chemical Langevin and Fokker-Planck equations for the reversible isomerization reaction. J Phys Chem A 106(20):5063–5071

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

Golightly A, Wilkinson D (2005) Bayesian inference for stochastic kinetic models using a diffusion approximation. Biometrics 61(3):781–788

Golightly A, Wilkinson D (2006) Bayesian sequential inference for stochastic kinetic biochemical network models. J Comput Biol 13(3):838–851

Goll J, Uetz P (2006) The elusive yeast interactome. Genome Biol 7(6):223

Goodwin BC (1963) Temporal organization in cells: a dynamic theory of cellular control processes. Academic Press, Newyork

Graupner S, Wackernagel W (2001) Identification and characterization of novel competence genes comA and exbB involved in natural genetic transformation of Pseudomonas stutzeri. Res Microbiol 152(5):451–460

Gregor T, Bialek W, van Steveninck R, et al (2005) Diffusion and scaling during early embryonic pattern formation. Proc Natl Acad Sci USA 102(51):18,403–18,407

Gruber AR, Neubeck R, Hofacker IL, et al (2007) The RNAz web server: prediction of thermodynamically stable and evolutionarily conserved RNA structures. Nucleic Acids Res 35: w335–338

Gubb D, García-Bellido A (1982) A genetic analysis of the determination of cuticular polarity during development in Drosophila melanogaster. J Embryol Exp Morphol 68:37–57

Han JD, Bertin N, Hao T, et al (2004) Evidence for dynamically organized modularity in the yeast protein-protein interaction network. Nature 430(6995):88–93

Hansen C, Quake SR (2003) Microfluidics in structural biology: smaller, faster em leader better. Curr Opin Struct Biol 13(5):538–544

Hao N, Nayak S, Behar M, et al (2008) Regulation of cell signaling dynamics by the protein kinase-scaffold Ste5. Mol Cell 30(5):649–56

Harris MA, Clark J, Ireland A, et al (2004) The Gene Ontology (Go) database and informatics resource. Nucleic Acids Res 32(Database issue)

Hart GT, Ramani AK, Marcotte EM (2006) How complete are current yeast and human protein-interaction networks? Genome Biol 7(11):120

Hartwell LH, Hopfield JJ, Leibler S, et al (1999) From molecular to modular cell biology. Nature 402(6761 Suppl)

Hastie T, Tibshirani R, Friedman JH (2001) The elements of statistical learning. Springer, New York, NY

Henderson DA, Boys RJ, Krishnan KJ, et al (2009) Bayesian emulation and calibration of a stochastic computer model of mitochondrial DNA deletions in substantia nigra neurons. J Am Stat Assoc 104(485):76–87

Henikoff S, Henikoff JG (1993) Performance evaluation of amino acid substitution matrices. Proteins 17(1):49–61

Hermjakob H, Montecchi-Palazzi L, Bader G, et al (2004) The HUPO PSI's molecular interaction format-a community standard for the representation of protein interaction data. Nat Biotechnol 22(2):177–183

Heron EA, Finkenstaedt B, Rand DA (2007) Bayesian inference for dynamic transcriptional regulation; the Hes l system as a case study. Bioinformatics 23(19):2596–2603

Herrgard M, Covert M, Palsson B (2003) Reconciling gene expression data with known genome-scale regulatory network structures. Genome Res 13(11):2423–2434

Higham DJ (2008) Modeling and simulating chemical reactions. SIAM Rev 50(2): 347–368

Hirata H, Yoshiura S, Ohtsuka T, et al (2002) Oscillatory expression of the bHLH factor Hes l regulated by a negative feedback loop. Scinence 298(5594):840–3

Hooper SD, Bork P (2005) Medusa: a simple tool for interaction graph analysis. Bioinformatics 21(24):4432–3

Hu Z, Mellor J, Wu J, et al (2007) Towards zoomable multidimensional maps of the cell. Nat Biotechnol 25(5):547–554

Huang CY, Ferrel JE Jr (1996) Ultrasensitivity in the mitogen-activated protein kinase cascade. Proc Natl Acad Sci USA 93(19):10,078–83

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

Ideker T, Valencia A (2006) Bioinformatics in the human interactome project. Bioinformatics 22(24):2973–2974

Ideker T, Galitski T, Hood L (2001) A new approach to decoding life: systems biology. Annu Rev Genomics Hum Genet 2:343–372

Igoshin O, Neu J, Oster G (2004a) Developmental waves in myxobacteria: A distinctive pattern formation mechanism. Phys Rev E 70(4)

Igoshin O, Welch R, Kaiser D, et al (2004b) Waves and aggregation patterns in myxobacteria. Proc Natl Acad Sci USA 101(12):4256–4261

Irizarry R, Warren D, Spencer F, et al (2005) Multiple-laboratory comparison of microarray platforms. Nat Meth 2(5):345–350

Ito T, Chiba T, Ozawa R, et al (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proc Natl Acad Sci USA 98(8):4569–4574

Jahnke T, Huisinga W (2008) A dynamical low-rank approach to the chemical master equation. Bull Math Biol 70(8):2283–2302

Jansen R, Gerstein M (2004) Analyzing protein function on a genomic scale: the importance of gold-standard positives and negatives for network prediction. Curr Opin Microbiol 7(5): 535–45

Jansen R, Lan N, Qian J, et al (2002) Integration of genomic datasets to predict protein complexes in yeast. J Struct Funct Genomics 2(2):71–81

Jansen R, Yu H, Greenbaum D, et al (2003) A Bayesian networks approach for predicting protein-protein interactions form genomic data. Science 302(5644):449–453

Jaqaman K, Danuser G (2006) Linking data to models: data regression. Nat Revi Mol Cell Biol 7(11):813–819

Jenssen TK, Laegreid A, Komorowski J, et al (2001) A literature network of human genes for high-throughput analysis of gene expression. Nat Genet 28(1):21–28

Jones KH, Liu J, Adler PN (1996) Molecular analysis of EMS-induced frizzled mutations in Drosophila melanogaster. Genetics 142(1):205–15

de Jong H (2002) Modeling and simulation of genetic regulatory systems: a literature review. J Comput Biol 9(1):67–103

Kanehisa M, Goto S, Hattori M, et al (2006) From genomics to chemical genomics: new developments in KEGG. Nucleic Acids Res 34(Database issue)

Karlin S, Taylor HM (1975) A first course in stochastic processes, 2nd edn. Academic Press, Newyork

Karlin S, Taylor HM (1981) A second course in stochastic processes. Academic Press, Newyork

Karp P, Riley M, Saier M, et al (2002) The EcoCyc Database. Nucl Acids Res 30(1):56–58

Kelley BP, Sharan R, Karp RM, et al (2003) Conserved pathways within bacteria and yeast as revealed by global protein network alignment. Proc Natl Acad Sci USA 100(20): 11,394–11,399

Kim JK, Gabel HW, Kamath RS, et al (2005) Functional genomic analysis of RNA interference in C. elegans. Science 308(5725):1164–1167

Kin T, Yamada K, Terai G, et al (2007) fRNAdb: a platform for mining/annotating functional RNA candidates from non-coding RNA sequences. Nucleic Acids Res 35(Database issue)

Klingensmith J, Nusse R, Perrimon N (1994) The Drosophila segment polarity gene dishevelled encodes a novel protein required for response to the wingless signal. Genes Dev 8(1): 118–30

Kondo S, Asai R (1995) A reaction-diffusion wave on the skin of the marine angelfish Pomacanthus. Nature 376(6543):765–768

Koyuturk M, Kim Y, Subramaniam S, et al (2006) Detecting conserved interaction patterns in biological networks. J Comput Biol 13(7):1299–1322

Krogan NJ, Cagney G, Yu H, et al (2006) Global landscape of protein complexes in the yeast saccharomyces cerevisiae. Nature 440(7084):637–643

Kuhn RM, Karolchik D, Zweig AS, et al (2007) The UCSC genome browser database: update 2007. Nucleic Acids Res 35(Database issue)

Lacalli TC (1990) Modeling the Drosophila pair-rule pattern by reaction-diffusion: gap input and pattern control in a 4-morphogen system. J Theor Biol 144 (2):171–194

Lamb J, Crawford ED, Peck D, et al (2006) The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science 313(5795):1929–1935

Lander ES, Linton LM, Birren B, et al (2001) Initial sequencing and analysis of the human genome. Nature 409(6822):860–921

Laub MT, McAdams HH, Feldblyum T, et al (2000) Global analysis of the genetic network controlling a bacterial cell cycle. Science 290(5499):2144–2148

Lee I, Data SV, Adai AT, et al (2004) A probabilistic functional network of yeast genes. Science 306(5701):1555–1558

Li H, Cao Y, Petzold LR, et al (2008) Algorithms and software for stochastic simulation of biochemical reacting systems. Biotechnol Prog 24(1):56–61

Liang Z, Xu M, Teng M, et al (2006) Comparison of protein interaction networks reveals species conservation and divergence. BMC Bioinformatics 7:457

Lu LJ, Xia Y, Paccanaro A, et al (2005a) Assessing the limits of genomic data integration for predicting protein networks. Genome Res 15(7):945–953

Lu P, Szafron D, Greiner R, et al (2005b) PA-GOSUB: a searchable database of model organism protein sequences with their predicted Gene Ontology molecular function and subcellular localization. Nucleic Acids Res 33(Database issue)

Luciano JS (2005) PAX of mind for pathway researchers. Drug Discov Today 10(13):937–942

Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits. Sunderland, MA: Sinauer Associates

Maini P, Benson D, Sherratt J (1992) Pattern-formation in reaction diffusion-models with spatially inhomogeneous diffusion-coefficients. Ima J Math Appl Med Biol 9(3):197–213

Maini PK, Baker RE, Chuong CM (2006) Developmental biology. The Turing model comes of molecular age. Science 314(5804):1397–8

Marchler-Bauer A, Anderson JB, DeWeese-Scott C, et al (2003) CDD: a curated Entrez database of conserved domain alignments. Nucleic Acids Res 31(1):383–387

Matthiessen MW (2003) BioWareDB: the biomedical software and database search engine. Bioinformatics 19(17):2319–2320

McAdams H, Arkin A (1997) Stochastic mechanisms in gene expression. Proc Natl Acad Sci USA 94(3):814–819

Meinhardt H, de Boer PA (2001) Pattern formation in Escherichia coli: a model for the pole-to-pole oscillations of Min proteins and the localization of the division site. Proc Natl Acad Sci USA 98(25):14,202–14,207

Mewes HW, Heumann K, Kaps A, et al (1999) MIPS: a database for genomes and protein sequences. Nucleic Acids Res 27(1):44–48

Milo R, Jorgensen P, Springer M (2009) Bionumbers: The Database of Useful Biological Numbers. bionumbers.hms.harvard.edu. Accessed 31 Aug 2009

Mogilner A, Wollman R, Marshall WF (2006) Quantitative modeling in cell biology: What is it good for? Dev Cell 11(3):279–287

Mulder NJ, Apweiler R, Attwood TK, et al (2007) New developments in the InterProdatabase. Nucleic Acids Res 35(Database issue)

Munsky B, Khammash M (2006) The finite state projection algorithm for the solution of the chemical master equation. J Chem Phys 124(4)

Nelles O (2000) Nonlinear system identification: from classical approaches to neural networks and fuzzy models, 1 st edn. Springer, New York, NY

Ng A, Bursteinas B, Gao Q, et al (2006) pSTING:a 'systems' approach towards integrating signalling pathways, interaction and transcriptional regulatory networks in inflammation and cancer. Nucleic Acids Res 34(Database issue)

Nichols R (2001) Gene trees and species trees are not the same. Trends Ecol Evol 16(7):358–364

Nielsen P, Halstead M (2004) The evolution of CellML. Conf Proc IEEE Eng Med Biol Soc 7:5411–5414

Novak B, Csikasz-Nagy A, Gyorffy B, et al (1998) Mathematical model of the fission yeast cell cycle with checkpoint controls at the G1/S, G2/M and metaphase/anaphase transitions. Biophys Chem 72(1–2):185–200

Novak B, Pataki Z, Ciliberto A, et al (2001) Mathematical model of the cell division cycle of fission yeast. Chaos 11(1):277–286

Ogata H, Fujibuchi W, Goto S, et al (2000) A heuristic graph comparison algorithm and its application to detect functionally related enzyme clusters. Nucleic Acids Res 28(20): 4021–4028

Orchard S, Hermjakob H, Taylor CF, et al (2005) Further steps in standardisation. Report of the second annual Proteomics Standards Initiative Spring Workshop (Siena, Italy 17–20th April 2005). Proteomics 5(14):3552–3555

Othmer H (1976) Qualitative dynamics of a class of biochemical control-circuits. J Math Bio 3(1):53–78

Overbeek R, Fonstein M, D'Souza M, et al (1999) The use of gene clusters to infer functional coupling. Proc Natl Acad Sci USA 96(6):2896–2901

Owen A, Stuart J, Mach K, et al (2003) A Gene Recommender Algorithm to Identify Coexpressed Genes in C. elegans. Genome Res 13(8):1828–1837

Painter KJ, Maini PK, Othmer HG (2000) A chemotactic model for the advance and retreat of the primitive streak in avian development. Bull Math Biol 62(3):501–525

Pamilo P, Nei M (1988) Relationships between gene trees and species trees. Mol Biol Evol 5(5):568–583

Pazos F, Ranea J, Juan D, et al (2005) Assessing Protein Co-evolution in the Context of the Tree of Life Assists in the Prediction of the Interactome. J Mol Biol 352(4):1002–1015

Pellegrini M, Marcotte EM, Thompson MJ, et al (1999) Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. Proc Natl Acad Sci USA 96(8):4285–4288

Pokholok DK, Zeitlinger J, Hannett NM, et al (2006) Activated signal transduction kinases frequently occupy target genes. Science 313(5786):533–6

Price ND, Shmulevich I (2007) Biochemical and statistical network models for systems biology. Curr Opin Biotechnol 18(4):365–370

Prudhommeaux E, Seaborne A (2007) SPARQL Query Language for RDF. www.w3.org/TR/rdf-sparql-query. Accessed 31 Aug 2009

Ptacek J, Snyder M (2006) Charging it up: global analysis of protein phosphorylation. Trends Genet 22(10):545–54

Qi Y, Bar-Joseph Z, Klein-Seetharaman J (2006) Evaluation of different biological data and computational classification methods for use in protein interaction prediction. Proteins: Struct Funct Bioinform 63(3):490–500

Rangamani P, Iyengar R (2007) Modelling spatio-temporal interactions within the cell. J Biosci 32(1):157–167

Rao C, Arkin A (2003) Stochastic chemical kinetics and the quasi-steady-state assumption: Application to the Gillespie algorithm. J Chem Phys 118(11):4999–5010

Rathinam M, El Samad H (2007) Reversible-equivalent-monomolecular tau: A leaping method for "small number and stiff" stochastic chemical systems. J Computational Phys 224(2):897–923

Ratsch G, Sonnenburg S, Srinivasan J, et al (2007) Improving the Caenorhabditis elegans genome annotation using machine learning. PLoS Comput Biol 3(2):e20

Resat H, Petzold L, Pettigrew MF (2009) Kinetic modeling of biological systems. Methods Mol Biol 541:311–335

Riddihough G (2003) Chromosomes through space and time. Science 301(5634):779

van Riel NAW, Sontag ED (2006) Parameter estimation in models combining signal transduction and metabolic pathways: the dependent input approach. IEE Proc Syst Biol 153(4):263–274

Robertson G, Bilenky M, Lin K, et al (2006) cisRED: a database system for genome-scale computational discovery of regulatory elements. Nucleic Acids Res 34(Database issue)

Rual JF, Venkatesan K, Hao T, et al (2005) Towards a proteome-scale map of the human protein-protein interaction network. Nature 437(7062):1173–8

Rubin DL, Lewis SE, Mungall CJ, et al (2006) National Center for Biomedical Ontology: advancing biomedicine through structured organization of scientific knowledge. OMICS 10(2):185–198

Sachs K, Perez O, Pe'er D, et al (2005) Causal Protein-Signaling Networks Derived from Multiparameter Single-Cell Data. Science 308(5721):523–529

Saha K, Schaffer D (2006) Signal dynamics in Sonic hedgehog tissue patterning. Development 133(5):889–900

Salis H, Kaznessis Y (2005) Accurate hybrid stochastic simulation of a system of coupled chemical or biochemical reactions. J Chem Phys 122(5)

Samoilov M, Plyasunov S, Arkin AP (2005) Stochastic amplification and signaling in enzymatic futile cycles through noise-induced bistability with oscillations. Proc Natl Acad Sci USA 102(7):2310–2315

Samoilov MS, Arkin AP (2006) Deviant effects in molecular reaction pathways. Nat Biotechnol 24(10) 1235–1240

SantaLucia J, Hicks D (2004) The thermodynamics of DNA structural motifs. Annu Rev Biophys Biomol Struct 33:415–440

Saric J, Jensen LJ, Ouzounova R, et al (2006) Extraction of regulatory gene/protein networks from medline. Bioinformatics 22(6):645–650
Sauer U (2004) High-throughput phenomics: experimental methods for mapping fluxomes. Curr Opin Biotechnol 15(1):58–63
Schena M, Shalon D, Heller R, et al (1996) Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. Proc Natl Acad Sci USA 93(20):10,614–10,619
Schnell S, Turner T (2004) Reaction kinetics in intracellular environments with macromolecular crowding: simulations and rate laws. Prog Biophys Mol Bio 85(2–3):235–260
Schubert W, Bonnekoh B, Pommer A, et al (2006) Analyzing proteome topology and function by automated multidimensional fluorescence microscopy. Nat Biotechnol 24(10):1270–1278
Schuldiner M, Collins S, Thompson N, et al (2005) Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. Cell 123(3):507–519
Sekimura T, Zhu M, Cook J, et al (1999) Pattern formation of scale cells in lepidoptera by differential origin-dependent cell adhesion. Bull Math Biol 61(5):807–827
Shannon P, Markiel A, Ozier O, et al (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13(11):2498–2504
Sharan R, Suthram S, Kelley RM, et al (2005) From the Cover: Conserved patterns of protein interaction in multiple species. Proc Natl Acad Sci USA 102(6):1974–1979
Shea M, Ackers G (1985) The "or" control-system of bacteriophage-lambda - a physical-chemical model for gene-regulation. J Mol Biol 181(2):211–230
Sherlock G (2000) Analysis of large-scale gene expression data. Curr Opin Immunol 12(2): 201–205
Shmulevich I, Dougherty E, Kim S, et al (2002) Probabilistic Boolean networks: a rule-based uncertainty model for gene regulatory networks. Bioinformatics 18(2):261–274
Sick S, Reinker S, Timmer J, et al (2006) WNT and DKK determine hair follicle spacing through a reaction-diffusion mechanism. Science 314(5804):1447–1450
Siek J, Lee L. Lumsdaine A (2007) The Boost Graph Library. www.boost.org/libs/graph/. Accessed 31 Aug 2009
Singh R, Xu J, Berger B (2007) Pairwise Global Alignment of Protein Interaction Networks by Matching Neighborhood Topology. Proceedings of the 11th Annual International Conference on Computational Molecular Biology (RECOMB 2007)
Smolen P, Baxter D, Byrne J (2002) A reduced model clarifies the role of feedback loops and time delays in the Drosophila circadian oscillator. Biophys J 83(5):2349–2359
Spellman PT, Sherlock G, Zhang MQ, et al (1998) Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. Mol Biol Cell 9(12):3273–3297
Spiro P, Parkinson J, Othmer H (1997) A model of excitation and adaptation in bacterial chemotaxis. Proc Natl Acad Sci USA 94(14):7263–7268
Srinivasan B, Caberoy N, Suen G, et al (2005) Functional genome annotation through phylogenomic mapping. Nat Biotechnol 23(6):691–698
Srinivasan BS, Novak AF, Flannick J, Batzoglou S, McAdams HH (2006) Integrated protein interaction networks for 11 microbes. In: RECOMB, pp 1–14
Srinivasan BS, Shah NH, Flannick JA, et al (2007) Current progress in network research: toward reference networks for key model organisms. Briefings In Bioinform 8(5):318–332
Stamatakis M, Mantzaris NV (2006) Modeling of ATP-mediated signal transduction and wave propagation in astrocytic cellular networks. J Theor Biol 241(3):649–668
Stark C, Breitkreutz BJ, Reguly T, et al (2006) BioGRID: a general repository for interaction datasets. Nucleic Acids Res 34(Database issue)
Stephens S (2007) HCLSIG BioRDF Subgroup. esw.w3.org/topic/HCLSIG_BioRDF_Subgroup. Accessed 31 Aug 2009
Steuer R (2004) Effects of stochasticity in models of the cell cycle: from quantized cycle times to noise-induced oscillations. J Theor Biol 228(3):293–301
Stromback L, Lambrix P (2005) Representations of molecular pathways: an evaluation of SBML, PSI MI and BioPAX. Bioinformatics 21(24):4401–4407

Strutt DI (2001) Asymmetric localization of frizzled and the establishment of cell polarity in the Drosophila wing. Mol Cell 7(2):367–75
Stuart J, Segal E, Koller D, et al (2003) A gene-coexpression network for global discovery of conserved genetic modules. Science 302(5643):249–255
Stumpf M, Kelly W, Thorne T, et al (2007) Evolution at the system level: the natural history of protein interaction networks. Trends Ecol Evol 22(7):366–373
Tanay A, Sharan R, Kupiec M, et al (2004) Revealing modularity and organization in the yeast molecular network by integrated analysis of highly heterogeneous genomewide data. Proc Natl Acad Sci USA 101(9):2981–2986
Taylor J, Abramova N, Charlton J, et al (1998) Van Gogh: a new Drosophila tissue polarity gene. Genetics 150(1):199–210
Theisen H, Purcell J, Bennett M, et al (1994) dishevelled is required during wingless signaling to establish both cell polarity and cell identity. Development 120(2):347–60
Tomlin CJ, Axelrod JD (2007) Biology by numbers: mathematical modelling in developmental biology. Nat Rev Genet 8(5):331–340
Tong AH, Evangelista M, Parsons AB, et al (2001) Systematic genetic analysis with ordered arrays of yeast deletion mutants. Science 294(5550):2364–2368
Tong AH, Drees B, Nardelli G, et al (2002) A combined experimental and computational strategy to define protein interaction networks for peptide recognition modules. Science 295(5553): 321–324
Tree DRP, Shulman JM, Rousset R, et al (2002) Prickle mediates feedback amplification to generate asymmetric planar cell polarity signaling. Cell 109(3):371–381
Troyanskaya OG, Dolinski K, Owen AB, et al (2003) A Bayesian framework for combining heterogeneous data sources for gene function prediction (in Saccharomyces cerevisiae). Proc Natl Acad Sci USA 100(14):8348–8353
Turing A (1952) The Chemical Basis of Morphogenesis. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences 237(641):37–72
Tyson J (1975) Existence of oscillatory solutions in negative feedback cellular control processes. J Math Biol 1(4):311–315
Tyson J, Othmer H (1978) The dynamics of feedback control circuits in biochemical pathways. Prog Theor Biol 5:1–60
Uhrmacher A, Degenring D, Zeigler B (2005) Discrete event multi-level models for systems biology. Transactions on computational systems biology I pp 66–89. Springer, Berlin
Vastrik I, D'Eustachio P, Schmidt E, et al (2007) Reactome: a knowledgebase of biological pathways and processes. Genome Biol 8:R39
Venter JC, Adams MD, Myers EW, et al (2001) The sequence of the human genome. Science 291(5507):1304–1351
Vilar JMG, Kueh HY, Barkai N, et al (2002) Mechanisms of noise-resistance in genetic oscillators. Proc Natl Acad Sci USA 99(9):5988–5992
von Mering C, Jensen LJ, Kuhn M, et al (2007) STRING 7–recent developments in the integration and prediction of protein interactions. Nucleic Acids Res 35(Database issue)
Walter CF (1970) The occurrence and the significance of limit cycle behavior in controlled biochemical systems. J Theor Biol 27(2):259–272
Wang X, Hao N, Dohlman H, et al (2006) Bistability, stochasticity, and oscillations in the mitogen-activated protein kinase cascade. Biophys J 90(6):1961–1978
Weber M, Schubeler D (2007) Genomic patterns of dna methylation: targets and function of an epigenetic mark. Curr Opin Cell Biol 19(3):273–280
Wei CL, Wu Q, Vega VB, et al (2006) A global map of p53 transcription-factor binding sites in the human genome. Cell 124(1):207–219
Weinberger LS, Burnett JC, Toettcher JE, et al (2005) Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity. Cell 122(2):169–182
Weitz J, Benfey P, Wingreen N (2007) Evolution, interactions, and biological networks. PLoS Biol 5(1):e11

Weng G, Bhalla U, Iyengar R (1999) Complexity in biological signaling systems. Science 284(5411):92–96
Wheeler DL, Barrett T, Benson DA, et al (2007) Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 35(Database issue)
Wilkinson DJ (2006) Stochastic modelling for systems biology. Chapman and Hall/CRC mathematical and computational biology series, Taylor and Francis
Wilkinson DJ (2009) Stochastic modelling for quantitative description of heterogeneous biological systems. Nat Rev Genet 10(2):122–133
Winzeler EA, Shoemaker DD, Astromoff A, et al (1999) Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. Science 285(5429):901–906
Wong SL, Zhang LV, Tong AH, et al (2004) Combining biological networks to predict genetic interactions. Proc Natl Acad Sci USA 101(44):15,682–15,687
Woo Y, Affourtit J, Daigle S, et al (2004) A Comparison of cDNA, Oligonucleotide, and Affymetrix GeneChip Gene Expression Microarray Platforms. J Biomol Tech 15(4):276–284
Yi M, Jia Y, Liu Q, et al (2006) Enhancement of internal-noise coherence resonance by modulation of external noise in a circadian oscillator. Phys Rev E 73(4)
Yi M, Jia Y, Tang J, et al (2008) Theoretical study of mesoscopic stochastic mechanism and effects of finite size on cell cycle of fission yeast. Phys A-Stat Mech Appl 387(1):323–334
You L (2004) Toward computational systems biology. Cell Biochem Biophys 40(2):167–184
Yu H, Luscombe NM, Lu HX, et al (2004) Annotation transfer between genomes: protein-protein interologs and protein-DNA regulogs. Genome Res 14(6):1107–1118
Zhenping L, Zhang S, Wang Y, et al (2007) Alignment of molecular networks by integer quadratic programming. Bioinformatics 23(13):1631–1639
Zhu X, Gerstein M, Snyder M (2007) Getting connected: analysis and principles of biological networks. Genes Dev 21(9):1010–1024
Zou M, Conzen S (2005) A new dynamic Bayesian network (DBN) approach for identifying gene regulatory networks from time course microarray data. Bioinformatics 21(1):71–79