Published in Computer Science Review. 3:1-17 (2009)

Computational Challenges in Systems Biology

Allison P. Heath[a], Lydia E. Kavraki[a,b,c]

[a]*Department of Computer Science, Rice University, Houston, TX 77005, USA*
[b]*Department of Bioengineering, Rice University, Houston, TX 77005, USA*
[c]*Structural and Computational Biology and Molecular Biophysics, Baylor College of Medicine, Houston, TX, 77005, USA*

Abstract

Systems biology is a broad field that incorporates both computational and experimental approaches to provide a system level understanding of biological function. Initial forays into computational systems biology have focused on a variety of biological networks such as protein-protein interaction, signaling, transcription and metabolic networks. In this review we will provide an overview of available data relevant to systems biology, properties of biological networks, algorithms to compare and align networks and simulation and modeling techniques. Looking towards the future, we will discuss work on integrating additional functional information with biological networks, such as three dimensional structures and the complex environment of cell. Combining and understanding this information requires development of novel algorithms and data integration techniques and solving these difficult computational problems will advance both computational and biological research.

Key words: computational systems biology, biological networks, graphs, network structures, graph comparison, network motifs

1. Introduction

Systems biology is an interdisciplinary field concerned with understanding biological processes on the system level (1; 2; 3; 4). The field is founded upon taking an integrative approach to understand the functioning of complex biological systems as a whole instead of individual pieces. The recent and rapid growth of systems biology belies the deep roots of its underlying concepts. The idea of applying systems theory to biology can be dated back to at least the early 20th century (5). Over the decades since then, a number of researchers theorized about biological systems, but it took the success of modern experimental biology to provide the data that feeds systems biology. The wealth of experimental data has been a huge boon by validating computational approaches and enabling the study of biological systems. At the same time, this data has created a number of challenges in terms of processing and understanding it. The systems biology approach includes both experimental and computational components working in conjunction in order to achieve understanding of biological processes. In this review we will focus on the computational aspects of systems biology.

There is no doubt that computational tools are required to process and understand the avalanche of available data, but important questions remain unanswered. Just a few of these questions include: What computational tools can be used to combine data obtained from different resources and at different levels of resolution? What computational models could capture the dynamic behavior of networks? What computational methodologies can be developed to reduce the available information to coherent hypotheses? This review will provide the background for answering these questions as well as

current approaches.

The review is organized as follows: Section 2 focuses on the different types of biological network data used in systems biology and how and where this data is stored. The data is the foundation upon which computational approaches can be built in systems biology. Section 3 discusses research on the topology of these biological networks. Then algorithms for motif finding in biological networks are discussed in Section 4 followed by algorithms for network alignment and comparison in Section 5. The modeling and simulation of the dynamics of biological networks are covered in Section 6. The review concludes with a discussion of future challenges in computational systems biology.

2. Availability of Biological Network Data

The quantity and quality of network data has been rapidly increasing in the last few years. It is important to understand the state of existing data and methods in order to further what can be done with this data. There are a large number of databases in existence which contain information on biological networks. In 2008, Pathguide, a listing of biological pathway resources, contains about 240 biological pathway resources. These resources vary in their focus, size, curation levels, availability and popularity (6). Instead of providing a comprehensive list, this section highlights some of more widely used resources to provide a flavor of what data is stored and how it can be accessed. Many of the databases initially limited their scope to one particular type of network or organism. Therefore, databases are often categorized based on this criteria. However, as systems biology moves forward

it is becoming apparent that these categorizations are for convenience and often do not represent reality because in the cell all of these networks work together. Therefore, a number of databases have been expanding their focus to encompass a larger variety of data.

2.1. Metabolic Pathways

Metabolic pathways consist of chemical compounds and the enzymatic reactions which catalyze the chemical reactions that support life. Two of the major online databases focusing on metabolic pathway information are The Kyoto Encyclopedia of Genes and Genomes (KEGG) PATHWAY (7) database and MetaCyc (8), part of the larger BioCyc Database Collection (9). MetaCyc and KEGG both contain similar data about metabolic pathways and their components, but they vary in how this data is organized, curated and made available. Figure 1 contains a side by side comparison of how KEGG and MetaCyc display the citric acid cycle, a pathway important for aerobic cellular respiration. However, the citric acid cycle is just a small part of an organism's entire metabolism. In comparison, Figure 2 shows all of the metabolic reactions in *Escherichia coli* extracted from KEGG.

The KEGG PATHWAY database was initially released in December 1995 and contained a selected number of metabolic pathways. KEGG PATHWAY is manually curated and drawn based on the literature. The database in 2008 contains 159 reference pathways under the metabolism category. Each reference pathway represents a metabolic function (e.g., glycolysis) and contains the enzymatic reactions and chemical compounds which make up the pathway. The reference pathways can be thought of as a union of the reactions and compounds found across organisms, one can then select a particular

organism of interest and the reactions and compounds found in that organism will be highlighted on the pathway. KEGG is continually expanding, in December 2008, KEGG contained 92,565 pathways generated from 344 reference pathways and information on 104 eukaryotes, 726 bacteria and 53 archaea.

KEGG also hosts databases which provide further information about the enzymes and compounds, along with cross linking to other databases via GenomeNet (10). All of the data found in KEGG is free for academic users and can be downloaded via FTP in different flat file formats or accessed through web services. The reference metabolic pathways are also provided in KEGG Markup Language (KGML), an XML-based format created by KEGG. More recently, KEGG PATHWAY has been expanding to include other biological pathways such as those involved in genetic information processing, environmental information processing, cellular processes, human diseases and drug development. However, the metabolic pathway maps still remain the most popular feature (7).

MetaCyc is another metabolic pathway database which contains manually curated information from scientific literature. MetaCyc was first released in 1999, in conjunction with EcoCyc, a database containing pathway and genome information about *Escherichia coli*. MetaCyc uses a more hierarchical categorization than KEGG, usually with smaller and more specialized individual pathways. MetaCyc contains more than 1,100 metabolic pathways, across more than 1,500 organisms. It also contains information about the compounds and enzymes in these pathways. The MetaCyc data is freely available to all users in different file formats, including BioPAX and SBML.

There is also a BioCyc software package, free for academic usage, called Pathway Tools which provides analysis and visualization capabilities.

2.2. Signaling Pathways

Signaling pathways are the physical and chemical processes by which the cell transmits signals. Often, the signal is initiated by an extra cellular stimulus which causes a cellular response. This cellular response is usually mediated by a number of protein-protein and protein-ligand interactions. Signaling pathways play a key role in the regulation of biochemical systems in the cell and therefore are an important research area for the study of disease and drug development.

For the most part, signaling pathways are much smaller and utilize a higher level of experimental verification than protein-protein interaction (PPI) networks and therefore there has not been as much of a push for database services as with PPI data. The smaller size of signaling pathways allows for them to be published in a single article and so signaling pathways are often found in the literature. This is demonstrated by the fact that both Science (12) and Nature, along with the National Cancer Institute (NCI) (13), have their own databases containing cell signaling pathways.

Several other biological network databases, such as KEGG, have expanded to incorporate signaling pathways, but it is not their main focus. There have been some newer efforts to provide a resource for signaling pathways such as SPIKE (14). Signaling pathways can also be found in databases which strive to contain a number of different types of pathways as discussed in Section 2.6.

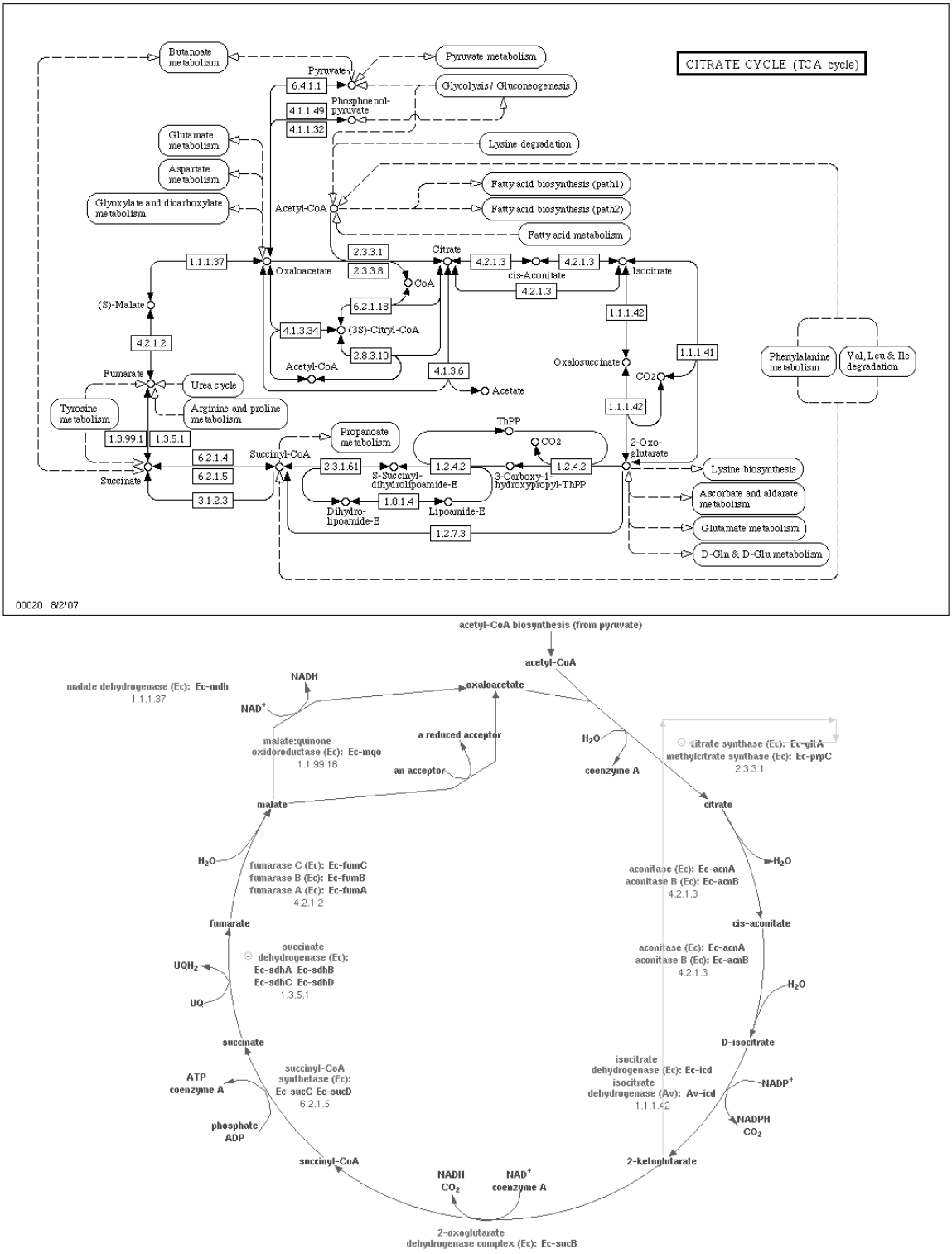

Figure 1: Two representations of the citric acid cycle. The top figure is taken from KEGG's website. The small circles represent individual chemical compounds and the rectangles represent enzymes and are labeled with Enzyme Commission (EC) numbers. The solid directed edges connect the substrate and products to their enzymatic reactions. The rounded ovals are other metabolic pathways and the dotted edges connect chemical compounds to other metabolic pathways. The bottom figure is taken from MetaCyc's website. MetaCyc offers several levels of detail to display, this one was chosen as being most similar to the level of detail displayed by KEGG. The chemical compounds are connected by directed edges which are labeled with the enzyme's name and EC number which catalyzes the reactions. On both the KEGG and MetaCyc website each individual element is clickable and takes the user to a page with more information about that element.

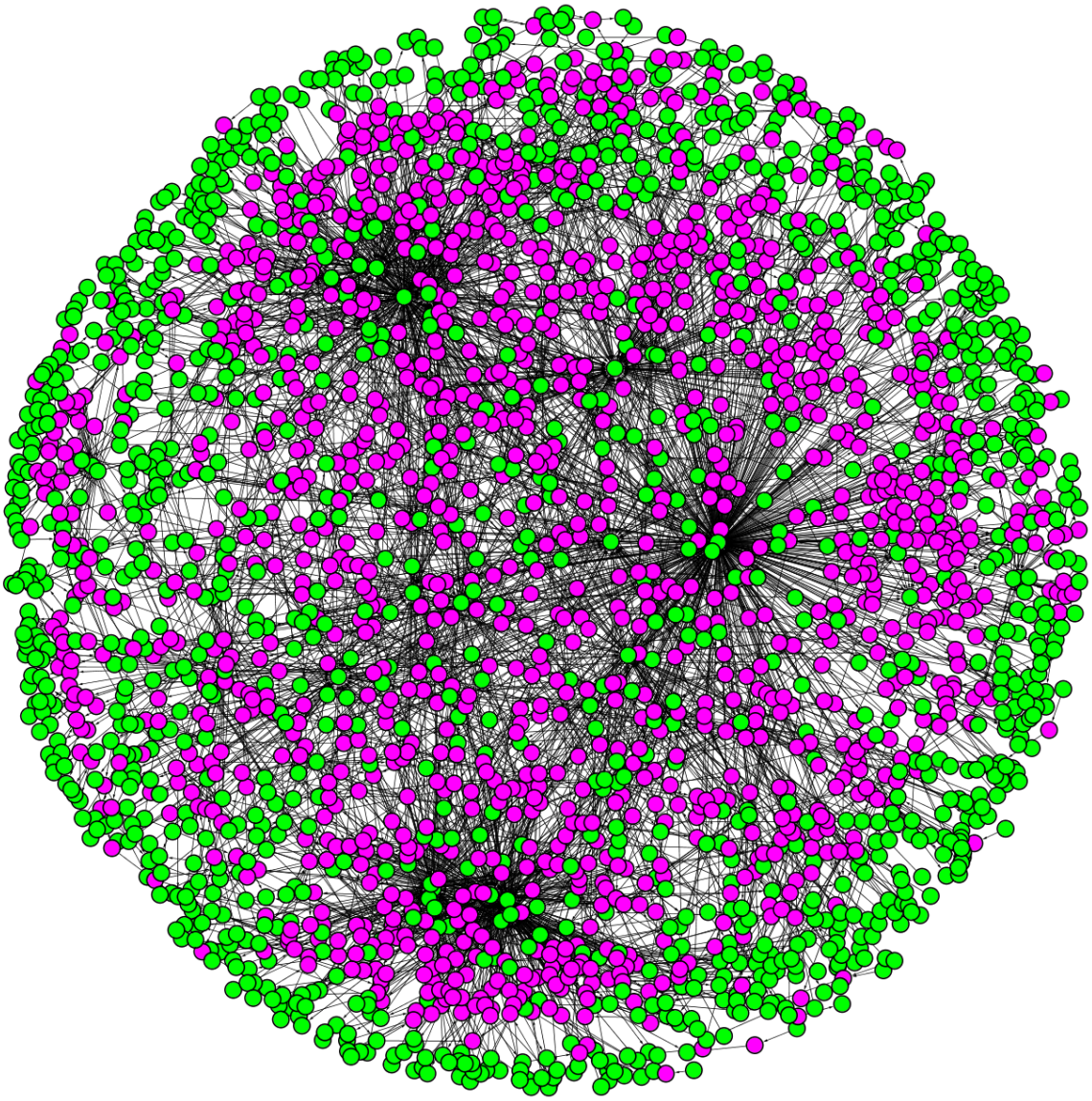

Figure 2: Visualization of the metabolic network of *Escherichia coli* obtained from KEGG. The magenta nodes represent enzymatic reactions and the green nodes are chemical compounds. The network contains 2570 nodes and 5238 edges. The figure was made using Cytoscape (11).

2.3. Transcription Regulatory Networks

Transcription regulatory networks consist of the set of transcription factors and the genes that they regulate. Similar to signaling pathways, the main source of transcription regulatory networks data is the literature. Research groups will perform a number of experiments to identify transcription factors and their targets and publish this data in scientific journals, often creating a website from which this data can be downloaded. For example one of the first large transcription regulatory networks was for *Saccharomyces cerevisiae* (15) and the associated data can be found on the group's website. Over time these data sets are built upon and analyzed and new data sets are published in the literature. However, few centralized sources for transcription regulatory networks exist. This creates extra difficulty for data cleaning, verification and integration. Transcription factor information can also be found in other places, such as databases which contain functional annotation of genes and proteins. However, extracting this data can be difficult and often requires expert knowledge.

The transcription regulatory network databases that exist are usually focused on one particular organism. One such source is RegulonDB (16), which focuses on the transcriptional regulatory network of *Escherichia coli* K12. RegulonDB provides curated data created from experimental data for download, as well as separate computational predicted interactions. It includes data on promoters such as sigma factors, transcription factor binding sites, small RNAs and other molecules and interactions involved in transcription. The data is available for download as plain text files. RegulonDB has also worked to provide cross linking and integration with other *Escherichia*

coli databases. Along with the data, RegulonDB also provides a number of tools for visualization and analysis of the data.

2.4. Protein-Protein Interaction Networks

Out of the many categories of biological networks, PPI network databases are perhaps the most varied and noisiest. This is partly because PPI is a very broad term. Interactions can come in the form of tightly bound protein complexes to fleeting bindings and modifications such as phosphorylation. There are also a wide variety of experimental techniques used in studying PPIs (17). Even so, it is often difficult to determine the level of error in experimentally determined PPI data and understanding the level of error is an active area of research (18; 19; 20). Despite these difficulties, a number of databases have arose which contain PPI networks for a number of organisms. In order to coordinate this effort the International Molecular Exchange Consortium (IMEx) was founded to help prevent duplicate efforts and provide standards for data storage, exchange and quality (21). The IMEx include a number of existing PPI databases such as DIP (22), IntAct (23), MINT (24), MPact (25) and BioGRID (26).

The Database of Interacting Proteins (DIP) was originally founded in 1999 to collect and provide a manually curated source of information on PPIs found in the literature (22). DIP continues to use literature as the main source of their data, but have moved towards including other high-quality sources such as protein complexes found in the Protein Data Bank (PDB) (27). As of 2008, DIP contains 19935 proteins and 56638 interactions across 204 organisms. DIP provides several different search functions based on keywords, protein sequences or specific articles. DIP also freely provides its

interaction data for download in several different formats. DIP also provides several different measures of quality and reliability of PPIs in the database. These measures are available as services on the DIP website so that users can calculate them for new interactions.

The Molecular INTeration (MINT) database is in many ways similar to DIP in that it is a hand-curated database of PPIs culled from the scientific literature (24). However, the actual interaction data contained in the two different databases are different. MINT has traditionally focused on interactions found in mammals, although in 2008 the organism with the most interactions in their database is *Saccharomyces cerevisiae*. Since 2005 MINT has focused on curating articles from *FEBS Letters*, *EMBO Journal* and *EMBO Reports*. By the numbers, MINT contains 28817 proteins and 105899 interactions. A study in 2006 of human PPIs found in several different databases, including DIP and MINT, demonstrated that there was low overlap between the interactions found in the two databases and in PPI databases in general (28). However, it is worth noting that since that article the number of human PPI in MINT has nearly doubled and the number in DIP has increased by about 20 percent. The data in MINT is freely available for research purposes and they also provide their data in a number of standard and non-standard formats.

While not a PPI database in the sense of DIP and MIP, the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) is worth mentioning as a site which makes PPI data readily available (29). The goal of STRING is to "collect, predict and unify most types of protein-protein associations" (30). STRING does not curate PPIs, but rather relies on other

databases such as DIP and MINT for experimentally determined interactions. It then integrates the data from the different sites to produce a data set using consistent identifiers. STRING also uses information across multiple PPI databases to compute confidence scores for each interaction. On top of this data, STRING has a number of search and visualization tools and cross links to other types of databases. STRING also provides several computationally determined or augmented data sets with predicted PPIs based on a number of criteria. The interaction data is free to download from STRING, but the entire database requires a license to prevent redistribution. STRING also provides an API for programmatic access to the database. Data integration sites such as STRING are an important part of the management of the increasing amount of pathway data.

2.5. Standardizing Data Representation

Bioinformatics and computational biology have been plagued by the creation of numerous, custom file formats for data representation and exchange. Too much effort has been put towards creating new parsers and converters for these different file formats. However, creating well designed and useful format standards for a wide range of researchers is a non-trivial task.

Given the wide range of data available, the systems biology community recognized that creating standards for data representation and exchange was important to enable collaboration and development of useful software tools (32). Several different efforts were started to create standard formats which could be widely used. We will discuss three major ones: SBML (33), CellML (34) and BioPAX (35). All three of these formats are utilize XML, an open standard for creating markup languages (36). The advantage of using XML

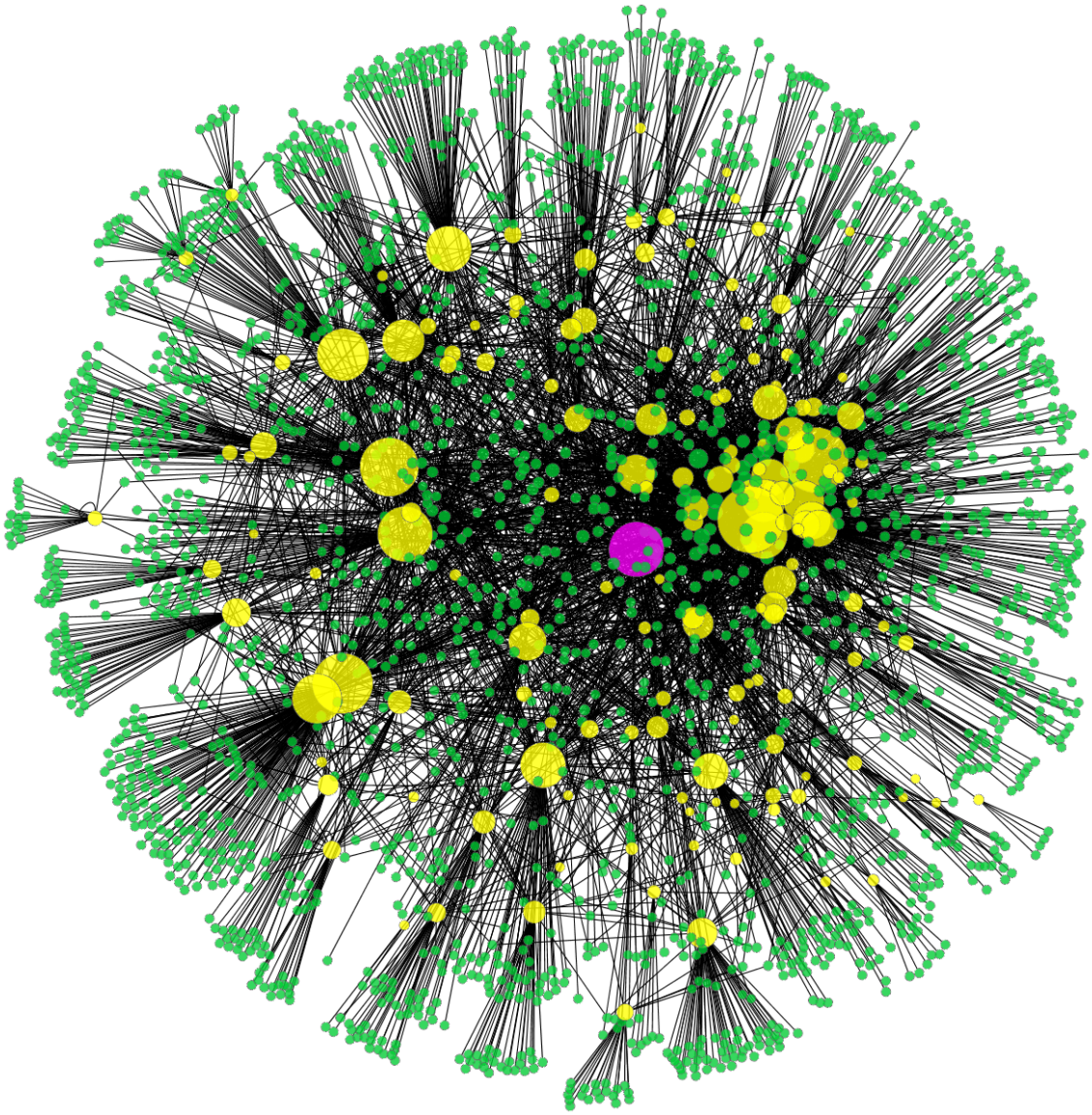

Figure 3: Visualization of the PPIs within two steps away from the epidermal growth factor receptor (EGFR) in humans extracted from the Human Protein Reference Database (31). The nodes are proteins and the edges represent interactions between them. There are 2,331 proteins and 5,255 interactions. EGFR is in magenta, proteins directly interacting with EGFR are in yellow and proteins two steps away are in green. The size of the node is proportional to its degree. This figure was made using Cytoscape (11).

lies in ability to use well established software tools developed to parse and validate XML based languages. While these formats have many similarities, they also address different needs in the community and may be more or less appropriate based on the type of data. Additionally, all three of these formats are under active development and are sure to evolve with the needs of the community.

CellML and SBML have similar goals to provide a way to represent and exchange mathematical models of biological processes. They often use similar methods and technologies to represent data, for example both models use MathML 2.0 to represent mathematical expressions. However, the two models differ on the specifics of how the data is conceptualized and represented.

The decision on whether to use CellML or SBML will depend on the specific model being worked on and compatibility with the desired software tools. Tools are being developed to convert between the two formats and work to some extent. However, it is possible to represent data in one format which cannot be converted to the other and thus the choice of format is still relevant.

BioPAX is different from SBML and CellML in that its goal is not to represent mathematical models, but rather to provide a standard exchange format for biological pathway data. BioPAX Level 1 was designed for metabolic pathways, Level 2 extended to contain molecular interactions, sequence features and hierarchical pathways. BioPAX Level 3 is under development and will extend to include signaling pathways, gene regulation and genetic interactions.

2.6. Multiple Types of Pathways and Integration Across Databases

While the division of biological pathway information into separate categories is convenient, it is often an artificial division and makes integrating multiple types of data difficult. As a result, several resources have been started to help gather, integrate and provide access to multiple types of pathways.

Some of these resources remain organism specific, but try to provide information about all of the different types of pathways in an organism. The Reactome project is such a resource for human pathway (37). Reactome seeks to be a expertly curated, comprehensive and detailed database for human biologic processes. This spans all of the biological networks discussion in the previous sections along with cross referencing to other databases in order to help with data integration. More recently, Reactome has added computationally inferred pathways and reactions from other organisms based on the human pathways. In addition to their website, where one can browse and search the data, Reactome has a variety of human and machine readable files for downloads.

Other resources function more as repositories where researchers can upload pathway models. Often the repositories offer both curated and uncurated pathway models. These typically revolve around a certain pathway format, for example the CellML Model Repository (38) or the BioModels Database (39) which is SBML oriented, but also provides other formats.

Several different ways have emerged to create a centralized resource for pathway information. WikiPathways is an open and public resource based on wiki model where the community is responsible for submission and curation

of pathways (40). WikiPathways was originally seeded with pathways from another data source, GenMAPP, but the goal is to mainly promote user submission of pathways.

Another approach is taken by Pathway Commons (41), which hopes to provide a common access point for existing biological pathway information. They act as a centralized source for information in other databases, and do not store or curate the data themselves. In 2008, Pathway Commons accesses data from Cancer Cell Map, HPRD, HumanCyc, IntAct,MINT, NCI/Nature Pathway Interaction and Reactome, with plans for more to come. All data is provided in BioPAX format. In addition to their website interface, Pathway Commons provides a web service API for programmatic access.

3. Topological Structure of Biological Networks

The availability of data on biological networks has led to new questions such as: Are they random? Does the structure of the network have implications for its functions? There is strong evidence that they are not fully random, but the exact nature of their structure is still not completely answered. Biological networks fit nicely into the general study of complex networks and so the analysis of their properties both drew upon and contributed to network and graph theory.

3.1. Graph Representations

In general, the elements in the networks, whether they be proteins, chemicals, genes or other biological entities, are usually the vertices of the graph and the edges represent some sort of interactions between them. Representing biological networks as graphs allows us to tap into the rich field of graph

theory to understand the properties and behaviors of these systems. However, the exact graph representation can vary depending on the data and in turn this can affect the choice of algorithms and conclusions reached about the structure of the networks.

3.2. Scale-Free and Small-World

One of the first questions to be addressed was how biological networks compare to random graphs. Random graphs are surrounded by well established theory beginning with the work of Erdős and Rényi (42). In 2000, a systematic analysis of the metabolic networks of 43 organisms was performed. The conclusions of this work was that topology of the metabolic networks are not random but rather "scale-free" and "small-world" (43). A scale-free network is one whose degree distribution is characterized by a power law of the form $P(k) \sim k^{-\gamma}$ (44). A small-world network is one where any two vertices in the network can be connected by a relatively short path. These properties of networks are of interest because a number of real-world networks, such as the Internet and citation networks, appear to have similar properties (45). A number of other papers found that other biological networks, such as PPI, transcription and signaling network also are scale-free and small-world and there have been number of review written about the subject (46; 47; 45; 48).

However, these properties have also been a point of contention and the debate surrounding them highlights how the properties of biological networks can change depending on how they are viewed. A recent review of the structure of metabolic pathways organizes arguments against biological networks being scale-free into four major categories (49): the quality of the data, the methodology used to determine scale-freeness, the graph representation of

biological networks and whether scale-free is a meaningful property.

The first argument is that biological data is incomplete or noisy and therefore one cannot conclude that the networks are scale-free. This does not prove that biological networks are not scale-free, but it is important to note that random subnets of scale-free networks are not scale-free (50). Additionally, one study sampled from theoretical biological networks and demonstrated that a limited sampling may provide misleading results (51). These results have important implications when dealing with incomplete biological networks and hopefully the picture will become clearer as the available data improves.

Methodological arguments say that different types of fitting and comparison can change whether the network is considered scale-free or not. In the area of PPI networks, it was argued that the networks fit a geometric random graph model better than a scale-free graph model (52). Instead of just looking at the degree distributions the article compares the distribution of "graphlets" which are all possible combinations of 3, 4, and 5 node graphs. The comparison of the frequencies of each of these graphlets in the PPI networks to random graphs, including random scale-free networks, revealed that the geometric random graph model provided the best fit. Therefore the conclusion was that the geometric random graph model is a better fit for PPI networks than the scale-free model.

It is argued that many graph representations of biological information abstract away potentially important differences between biological entities. This is the basis of the conclusion that the metabolic network of *Escherichia coli* is not small-world found in (53). Instead of allowing any path between

two nodes, this study only allowed a path from metabolite X to Y if at least one carbon atom in X reached Y. This approach brings the biochemical meaning of the metabolic pathways into the structural analysis. If one defines paths in this way, then the average path length of the *Escherichia coli* metabolic network is eight, larger than a random graph of the same size (53). It is also noted that in this type of representation of the metabolic network makes it inappropriate to evaluate scalefree-ness.

The final argument is that even if biological networks are scale-free, this does not tell us much about the network. One aspect of this argument is that scale-free networks are perhaps not as new and unsuspected as the literature would suggest (54). Another is that knowing a network is scale-free does not necessarily give any additional biological insight into the properties and behavior of the network (55). Therefore, researchers have moved towards understanding in more detail the structure of these networks. One popular way of doing this is to look at motifs found in the graphs.

4. Identifying Network Motifs

In the study of biological networks, "network motifs", often just called motifs, are defined as subgraphs which occur in significantly higher numbers than expected by random chance (56; 57). It has also been proposed that it is also important to look at rare subgraphs. These subgraphs are just identified by their topology; any biological information is ignored and just the structure of the graph is analyzed. Finding motifs in the networks helps to provide a better understanding of the structural elements of biological networks beyond just looking at global properties. For example, motifs have

been used to predict PPIs (58). The existence of different motifs may also help understand the role of modularity in biological networks and how it relates to the evolution of biological networks (59; 60).

The first type of networks to undergo motif analysis were transcription regulatory networks. As mentioned previously, these networks are usually represented as directed graphs where the vertices are transcription factors and their targets. In initial studies they exhaustively searched the networks for all subgraphs of size 3 and 4 (56; 57). Then by comparing the number of subgraphs found to the number of subgraphs found in random graph they were able to detect motifs. They found two motifs in both the *Escherichia coli* and *Saccharomyces cerevisiae* networks which are now widely known: the "feedforward loop" and the "bi-fan", which have interesting biological meaning (56; 61). This type of motif analysis has been extended to all types of biological networks.

Finding motifs can be broken down into three steps. First is determining the frequencies of subgraphs in the input network. Second is identifying and grouping isomorphic subgraphs. Last is then determining which subgraphs occur at significantly higher frequencies in the input graph than in a random graph. It should also be noted that the level of overlap two subgraphs are allowed to have can affect both results and performance. There are three different ways which the subgraphs can overlap: arbitrary overlaps of nodes and edges, only overlaps of nodes, and no overlaps (62). Motif finding algorithms in biological networks usually allow for arbitrary overlaps. The computational demands of these these three steps limit exact methods to motifs of 3-5; however using sampling techniques motif finding techniques

can find motifs of up to size 12-15 (63).

In order to overcome computational limitations, many algorithms use some form of sampling to estimate the frequencies of the subgraphs for the first step of motif finding. One way to do this is to randomly pick nodes from the network, which provides uniform sampling but will also sample disconnected subgraphs which are not considered motifs (64). Another motif finder, called mfinder (65), takes a different approach by picking a random edge and iteratively expanding by choosing an adjacent node until obtaining a subgraph of size n . The final sampled subgraph contains the selected nodes and all of edges between them in the original graph. This form of sampling produces a biased sampling and so the algorithm uses weights to correct the bias. This is done by calculating the probability P of sampling the specific subgraph and assigning it a weight $W = 1/P$. The weights are then used to calculate a significance score for a motif. The results from mfinder show that it is able to extract rare motifs and that relatively few samples are needed to obtain good results. Addtionally, mfinder is able to scale well with the overall size of the network, but does not scale well for larger motifs. Other criticisms of mfinder include the fact that the sampling bias is not necessarily eliminated by the weighting scheme and that computing these weights is expensive. Another sampling-based algorithm, called RAND-ESU addresses some of these issues, but still requires empirically calculated probabilities (66).

Non-sampling approaches include: partitioning the network as in NeMoFINDER (67) and mapping the network onto an embedded space (68). NeMoFINDER is able to find motifs up to size 12 by partitioning a network into a set of

graphs based on repeated trees. Other algorithms take a motif-centric view, where the algorithm is focused on finding a single motif instead of all motifs of a certain size. By taking this approach and enumerating all of the subgraphs and search for one at a time, motifs of size 15 in the PPI and transcription regulatory network of *Saccharomyces cerevisiae* were found (69).

The second step of motif finding requires graph isomorphism, a well-researched and difficult problem in graph theory. Therefore, many motif finding algorithms use established graph isomorphism algorithms, both exact and approximate. However, many graph isomorphism algorithms are aimed at comparing two relatively large graphs. While in motif finding, one must compare a large number of relatively small graphs. Therefore, the methods and heuristics may vary. One approach is to develop a label assignment based on the topology of each subgraph. This label is designed such that if two graphs are isomorphic they have the same labeling, and are likely, but not guaranteed, to have different labeling if they are not isomorphic (64). NeMoFINDER introduces the concept of "graph cousins", which is based on the similarity of the graph's canonical adjacency matrix (67). Large speed ups have also been achieved by using symmetry-breaking techniques to avoid repeatedly finding a subgraph due to its symmetry (69). These methods share similar themes to traditional graph isomorphism heuristics, but tailor them specifically to the motif finding application.

The third and final piece needed for motif finding is to determine an appropriate null model in order to determine which subgraphs occur at a frequency higher than expected. Most algorithms generate an ensemble of random graphs and run their subgraph finding methods on the random graphs.

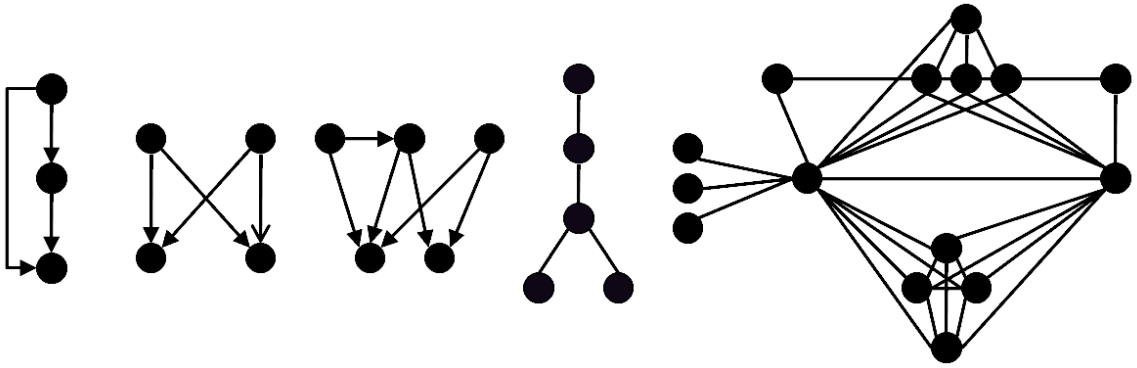

Figure 4: Some examples of directed and undirected motifs. The first three from left to right were identified in transcription regulatory networks: "feedforward loop" (57),"bi-fan" (56) and one of size 5 (65). The next two are from PPI networks: one of size 5 (52) and one of size 15 (69).

However, the choice of random graph is not necessarily straightforward as there are a number of random graph models and which ones are appropriate to use for biological networks is not always clear (70; 71).

Identifying motifs provides a more complete picture of an individual network, and research demonstrates that certain motifs appear to be common across a number of different networks. This has led to hypothesis and studies on on how motifs are related to evolution and modularity of biological networks. While motifs are an important tool to understand and compare the topology of networks, they do not explicitly incorporate biological information. Algorithms developed to compare biological as well as topological information of biological networks are usually classified as network alignment or network querying algorithms.

5. Biological Network Comparison

Today, biological sequence comparison algorithms, such as BLAST (72), are standard tools used by researchers. They provide an efficient way to identify sequence similarity which may have functional and/or evolutionary implications. In turn, the sequence alignments provide the basis for many other endeavors such as phylogeny and structure prediction algorithms. As

discussed previously, there is an increasing amount of biological network data available, and so questions arise if network comparison can be done to understand the similarities and differences between biological networks. This has also been termed "comparative interactomics" (73).

Network alignment differs from motif finding because it is not purely topological. Instead, network alignment includes both topology and a notion of biological similarity between the matched nodes. Network alignment requires both a scoring function and a search procedure. The similarity measure used is highly dependent on the type of network, for example in metabolic networks it could be the enzyme classification numbers (74), while in PPI networks it is usually based on sequence alignments (75).

Network comparison algorithms can be roughly classified as pairwise alignment, multiple alignment or querying methods. Pairwise network alignment is where two networks are globally compared to one another to identify regions of similarity. Figure 5 depicts a generic pairwise alignment. Multiple network alignment is a generalization of pairwise alignment, where more than two networks are compared. Network querying is finding parts of a larger network which are similar to a smaller query network. In the rest of this section, we briefly describe a few examples of network comparison algorithms.

5.1. Pairwise Network Alignment

PATHBLAST is an example of a pairwise alignment method (76). It takes as input two PPI networks and outputs conserved pathways between the two networks. This is done by first merging the input graphs based on protein sequence similarity into a "global alignment graph". Any two proteins from

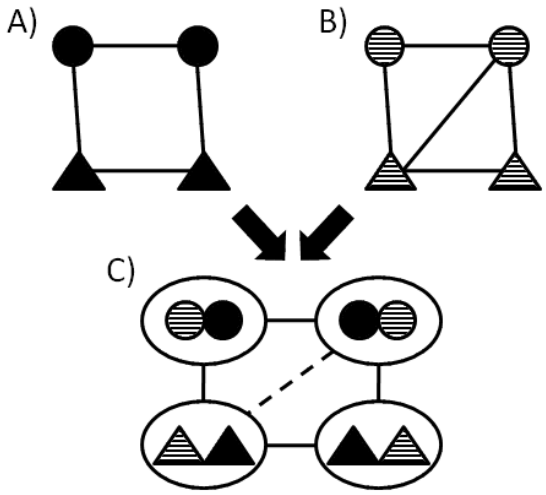

Figure 5: The similarity between nodes is represented by the shapes, in reality this would be some measure of biological similarity. Note that graph A and B would not be considered the same motif because the topology is different, but they can be aligned because of the similarity between the shapes. An alignment graph C is then constructed a solid edge is drawn if the corresponding edges if found in both A and B and a dashed line if the corresponding edge is only found in one of the graphs.

the original graphs which have a specified level of sequence similarity are merged into one node in the global alignment graph. The edges in the global alignment graph are labeled as direct, graph or mismatch, based on the existence of connections in the original graphs. A direct connection means the pair of nodes from the first graph and the pair of the nodes from the second graph are directly connected in their respective graphs. A gap means that one of the pairs of nodes is not directly connected, but are within two edges of each other. A mismatch means both of the pairs of nodes are within two edges of each other in their original graphs. And no edge is drawn if both of the pairs are further than two edges of each other.

Paths in the global alignment graph represent a potential alignment and each path has a score based on how likely the two proteins in a node are homologous and how likely the PPIs are real. Therefore, the objective is to find the highest scoring path. If the graph is acyclic, the high scoring path can be found in linear time via dynamic programming, but the global alignment graph normally contains cycles. Therefore, PATHBLAST generates $5L!$ acyclic subgraphs, where L is the is the length of the path to be found, and then finds the highest scoring path for each subgraphs. The expense of this operation limits the practical size of path. The method has been extended to three networks to find dense clusters of interactions in addition to the short linear paths (77).

5.2. Multiple Network Alignment

Multiple network alignment is often viewed as an extension of pairwise network alignment. One example is Graemlin (78), an algorithm developed for multiple alignment on large PPI networks that is based on performing

multiple pairwise alignments . The order that the pairwise alignments are performed is based on phylogenetic information about the species that the networks come from. Parsimonious ancestral history is also used in Graemlin's node scoring scheme along with the BLAST bitscore between proteins. The edge scoring parameterization can be adjusted based on the user's goals. Graemlin performs a pairwise alignment by generating a set of seeds, or "d-clusters", as all the nodes some d distance away from each node in the network. These clusters are then compared across the networks and clusters which score higher than some threshold T are kept and expanded. The method stops when no expansion can increase the score of the alignment. Multiple alignment is progressively done using the pairwise method and the order of pairwise comparisons are given by a phylogenetic tree, hence more closely related species are compared first. One difficulty with network alignment methods is creating an objective way to evaluate the quality of the results. The authors of Graemlin present measures using KEGG pathways as a way to evaluate the specificity and sensitivity of the alignment methods.

5.3. Network Querying

The final type of network comparison discussed in this section is network querying. In the case of network alignment a global comparison between one or more network is being done to find areas of similarity. In contrast, network querying takes a large biological network and a much smaller subnetwork or path of interest and returns parts of the larger network which have high similarity to the provided subnetwork or path.

PathMatch and GraphMatch are related network querying methods (79). The PathMatch algorithm only takes in a linear path to match against a

large biological network. By taking advantage of the fact that the query is a linear path, the algorithm is able to run in polynomial time. The path matching problem can be reduced to finding the longest weighted path in a directed acyclic graph. This is done by constructing a DAG containing all of the nodes in the network which correspond to the nodes in the pathway. The graph query method, GraphMatch, is an exact algorithm with time complexity $O(2^{|V_0|}|V_0|^2)$ where V_0 is the number of vertices in the query graph, G_0. These are efficient algorithms, but they are limited because they assume that the edge scoring scheme penalizes equally for mismatches and indels in the case of PathMatch and ignores mismatches in the case of GraphMatch. This assumption is unlikely to be true for a number of biological networks. The formulation of these algorithms demonstrate how the scoring scheme can effect the algorithm design and speed. If the scoring scheme does not penalize equally then the presented algorithms can no longer be used and it is likely that a more expensive algorithm is required.

Biological network comparison is still a relatively new field with a lot of potential, but it has many challenges to overcome before becoming a standard tool like sequence comparison. A more through discussion on the challenges that network comparison faces can be found in a recent review (75). While the development of comparison algorithms will greatly further the understanding of biological networks, especially how they relate to each other, it is still a static analysis of a dynamic system. Therefore, modeling and simulation play a key role in systems biology in order to understand the dynamics.

6. Biological Network Modeling and Simulation

While the term computational systems biology has had a recent resurgence, the idea of modeling biological systems has been around for many decades. For example, Turing worked on the idea of "mathematical biology" in the 1950's, publishing a paper on models for morphogenesis (80). Since then, developing computational models and simulating biology systems has become an area of increasing interest. As mentioned in the previous sections, understanding and comparing the structure of biological networks does bring important insight, but the next step beyond is understanding the dynamics of the system. A major goal, especially when it comes to disease and therapeutics, is to understand how the system changes when perturbed and how to modify the system to achieve a desired outcome.

Good computational modeling and simulation should be integrated tightly with experimental laboratory work. Experimental work provides information for creating the model, the computational work then generates hypothesis or points out where data is missing and provide information for which experiments should be done next, whose results are then used to improves the model. The research area concerned with modeling biological networks is vast. There are a number of different modeling and simulation types used in systems biology, depending on the type of system, the data available and results desired. These span from detailed continuous models such as ordinary differential equations to simple discrete models such as boolean networks. Mirroring the large number of options for modeling, a large number of software packages have been developed specifically for modeling and simulating biological networks and systems (81; 82; 83; 84). We will only touch on a

few techniques and examples here. Other recent reviews cover a number of methods and models used for gene regulatory networks (85; 86), signaling networks (87) and metabolic networks (88; 89).

6.1. Boolean Networks

The idea of using boolean networks to model has been around for several decades now, at least since 1969 when it was proposed to model a gene as "binary (on/off) device" in order to understand regulatory behavior (90). Boolean models are often cited as one of the "simplest" types of models because each entity in the model only has two states. This type of modeling fits naturally with gene regulatory networks and signaling networks where the states can represent represent active and inactive states. The relationships between the entities are then modeled using logical functions. The states of the entities are then updated synchronously or asynchronously based on the logical functions. Synchronous update is faster and more straightforward, but it is argued that biological systems do not operate synchronously and therefore asynchronous methods are needed (91; 92). There is also work to combine the two methods of updating for applications in biological networks (93). Boolean networks have also been extended to probabilistic boolean networks to deal with uncertainty (94).

Despite, or perhaps because of, their simplicity boolean models can provide insights into the dynamics of biological networks. Random boolean networks have been used to understand the robustness and stability of yeast transcriptional network (95). By generating a number of random boolean networks based on yeast transcriptional network data, it was concluded that the network is stable and robust under randomization of the boolean func-

tions. The results indicate that this may be because the structure of the network causes many nodes to be in a fixed state.

Other studies model systems more directly to understand the specific behavior of a network of interest. A recent example constructs a boolean network model for the T cell large granular lymphocyte (T-LGL) survival signaling network (96). Using this model the authors made predictions on what could cause deregulation in leukemic T-LGL. In turn these predictions were verified experimentally, demonstrating the validity and usefulness of the original model. Just a few examples of other biological networks modeled using boolean networks include mammalian cell cycle (92), yeast cell cycle (97), T-cell receptor signaling (98) and stomatal signaling networks in *Arabidopsis thaliani* (99).

6.2. Petri Nets

Standard petri nets provide a different way to model related discrete events asynchronously. The origins of petri nets are found in work by Carl Petri in the early 1960s to model asynchronous distributed systems (100). Since then petri nets have been applied to a number of different systems. Therefore, much work has gone into developing an underlying theory in order to understand their properties and behaviors (101). This preexisting framework can be a large advantage when modeling biological systems. The basic petri net is a directed bipartite graph consisting of place and transition nodes. Each place can be marked by a number of tokens, the state of the system represented by the number of tokens on each place is called a marking. Edges connect places to transitions if the transition is dependent on the state of the place. Transitions are connected to places if the transition has some

effect on the place. The weight of the edges represent the number of tokens that need to be transferred. If each of the inputs to a transition contain at least $w(p,t)$ tokens, where $w(p,t)$ is the weight of the edges, from the place to the transition, then the transition is said to be enabled. This allows the transition to be fired. A firing of a transition moves the $w(p,t)$ from each input and increments the number of tokens in each output place $w(t,p)$, where $w(t,p)$ is the weight from the transition to the output place. The petri net is initialized with an initial marking and allowed to evolve based on the firing of transitions and therefore the dynamics of the system can be studied.

There are a number of extensions to the basic petri net model, such as colored (102; 103), stochastic (104; 105) and hybrid petri nets (106). Many of these variations have found applications in modeling different types of biological networks, and in turn biological networks have been the driving force behind other petri net extensions such as hybrid functional petri nets (107) and signaling petri nets (108). Petri nets are applied most naturally to systems where there is some discrete unit being transferred, such as metabolic networks where the flow of tokens can represent fluxes (109; 110; 111; 112; 113; 114). They have also been applied to gene regulatory networks (115; 116; 117) and signaling networks (108; 118; 119; 120). A full review of petri nets and their applications to biological networks are outside the scope of this review. There are a number of reviews and books on petri nets (121; 101) as well as the application of petri nets to biological networks (122; 123; 124; 125).

6.3. Ordinary Differential Equations

Ordinary differential equations (ODEs) have a long history of being used to model dynamical systems, and biological networks are no exception. ODEs have been used heavily over the years to model biological systems (126). Differential equation modeling is appealing because it has deep mathematical roots which can be used to analyze and understand the system and its properties, such as robustness. Additionally, many software packages can be used for this type of modeling, from standard ones like Matlab and Mathematica to a growing number of custom packages aimed at biological networks (81; 84).

Kinetic models of biological networks are constructed using rate equations which describe the reaction rates of interactions in the system. This set of equations can be solved numerically by a number of methods to determine the concentration of molecular species over time. One of the problems which plague differential equation modeling is the lack of parameters, and work on handling incomplete information is important to overcoming this limitation (127). However, models using experimentally determined parameters are currently limited to small systems, such as glycolysis in yeast (128).

Many studies instead focus on how the dynamics of the system change over a range of parameter values. Bifurcation analysis and sensitivity analysis are established ways to understand how the dynamics of the system change in relation to their parameters (2). The results of these types of analysis can in turn be compared to what is known experimentally to make predictions about the system and general hypothesis for further study. There have been many ODE models of various biological systems with varying level of success,

just one example of a predictive model is that of the cell cycle, first developed for frog eggs, expanded to budding yeast and then generalized to eukaryotes (129; 130; 131).

One weakness of ODEs is that they ignore the randomness and noise found in biological networks, which may have important implications in their function (132; 133; 134; 135). This has caused a number of other methods to be developed which incorporate the stochasticity in the simulation of biological systems.

6.4. Stochastic Simulation

One way to add noise to differential models is to use stochastic differential equations (SDEs), where a noise term is added to the differential equations. However, differential equation models assume the concentration of molecular species to be continuous, which is reasonable when populations are large, but does not hold when discrete events or small populations need to be accounted for (136; 137). Therefore, methods which use a discrete formulation have become a popular way to stochastically model biological networks. Many of these methods are based on Gillespie's stochastic simulation algorithm (138).

The original formulation of the Gillespie algorithm is computationally demanding, but as interest has picked up, a number of improvements have been made to reduce the time needed for these types of simulations (139). Others have using field programmable gate arrays (FPGAs) in order to speed up the simulations by an order of magnitude (140). Stochastic simulations still remain slower than other techniques, and therefore there has been great interest in combining then with other techniques in order to model larger systems (141). This is part of a trend towards hybrid models as a way to

overcome limitations of current methods.

6.5. Hybrid Models

Hybrid models merge together discrete and continuous dynamics into a single dynamical model of a system. This encapsulates a broad space of models and systems, but a number of useful general frameworks have been developed (142; 143). The details of the model is dependent on the specific system, but researchers can build upon and draw from foundational work in the field of hybrid systems (144). In turn, problems from systems biology have created new challenges for the hybrid systems community and may drive the development of new models and algorithms (145; 146). Hybrid models are appealing for biological systems because they contain discrete events, such as a activation and inactivation, as well as properties that follow continuous dynamics, such as concentration (147).

A variety of hybrid system models have been developed for biological networks. Drawing on previous work on hybrid systems in areas outside of biology, some of these models have been used to perform reachability analysis to elucidate biologically meaningful properties. In the case of the *lac* operon, a relatively efficient reachability algorithm was used to study how the values of parameters effect the bi-stability of the system. The *lac* operon system has been well studied both experimentally and using continuous models (148; 149). Therefore, the hybrid model and use of the reachability algorithm were validated by comparing with experimental data and continuous models (150). Just a few other examples of biological hybrid system models analyzed in similar ways include Delta-Notch decision process (151; 152), bacterial stringent response (153) and genetic regulatory networks of carbon starvation

(154) and nutritional stress response (155) in *Escherichia coli*.

Hybrid petri nets are one specific example of a hybrid modeling tool used to study the dynamics of biological networks. They are a graphical modeling tool which allow the petri net to contain continuous places and transitions as well as the standard discrete elements in order to model hybrid systems (156; 157; 158). One study modeled several relatively small gene regulatory networks using hybrid petri nets in order to study their dynamics (117). In their operon model, consisting of the transcription of translation of two genes, continuous variables represented the concentration of proteins and mRNA and discrete variables represented the binding of RNA polymerase to initiate transcription. The discrete transition of the RNA polymerase binding and unbinding has a delay time associated with it which represented the time of transcription. When the transcription of a gene is finished, a function connects this discrete event to increasing the continuous variable of concentration of mRNA. Continuous variables and transitions then handle modeling the translation and degradation of mRNA and the corresponding proteins. Other work has extended hybrid petri nets to add additional constructs specifically for biological networks, called hybrid functional petri nets (107).

As modeling and simulation of biological networks continue to advance, researchers will face new challenges to develop meaningful, predictive models of biological systems. Additionally, new efficient algorithms which can handle these complex systems will be required. These algorithms will be required not only to merge discrete and continuous dynamics but also to assist in the incorporation of multiple levels of data operating in a wide range of time

scales.

7. Moving Forward: Biological Networks Do Not Stand Alone

The field of systems biology is founded on the principles of integration and considering the system as a whole. In practice, this has often meant looking at a single biological network at a time. This approach has provided a number of important insights into the workings of biological networks and still contains a number of challenges. However, as the amount of biological data available continues to increase it will also be increasingly important to integrate this data for computational analysis. One aspect of this data integration is incorporating similar data found across separate databases. For example, STRING's effort to integrate PPI databases as discussed in Section 2.4. Another, perhaps more challenging, aspect is the incorporation of additional levels of heterogeneous biological data to the graph representation of biological networks.

The standard representations of biological networks abstract away several layers of biological information. This abstraction is reasonable in a number of cases where the information is missing or incomplete. However, biological networks are found in a complex physical environment and as new data becomes available incorporating it into the model may help improve analysis and predictions about the system. One area which has had rapid growth is the availability of three-dimensional structures of proteins. The number of experimentally determined three dimensional protein structures has been increasing dramatically (159). This experimental information along with improvements in computational structural biology have made the use

of structural details of increasing interest for the field of systems biology (160; 161; 162; 163). For example, one study characterized PPIs using the structures of the proteins (164). The interactions were classified as either simultaneously possible, that is they use different interaction sites, or mutually exclusive, that is they use the same interaction site. This adds additional information to the original interaction graph and may have implications for the behavior of the system.

Adding protein structure information to biological systems is just one example of numerous ways in which models of biological systems can be augmented. A sampling of other information include temporal and spatial location (165; 166; 167), functional context (168; 169), environmental conditions (170; 171) and evolutionary context (172). Combining this heterogeneous data creates a number of challenges, both from a computational and an experimental point of view. An important study demonstrated the feasibility of combining a number of data sets for galactose utilization in yeast (173; 174), but the work in this area is far from done. Addressing how to coherently integrate multiple different levels and types of data will be an important factor in advancing systems biology.

8. Concluding Remarks

This review has presented a number of computational challenges in the field of systems biology. We began with an overview of different types of relevant data and highlighted a few online resources for each type. We presented a variety of ways developed to analyze and model both the static and dynamic properties of biological networks. We then provided a discussion

on the importance of integration of many different types of data in systems biology, as the quantity and quality of available data continues to increase. In order to further the knowledge of biological systems, computational work is required to understand the different levels and type of data, distill out the relevant data, and develop algorithms which can incorporate and analyze this data in a meaningful way.

Acknowledgments

This work has been supported by a Hamill Innovation Grant.

References

[1] T. Ideker, T. Galitski, L. Hood, A new approach to decoding life: systems biology., Annu Rev Genomics Hum Genet 2 (2001) 343–372. doi:http://dx.doi.org/10.1146/annurev.genom.2.1.343.

[2] H. Kitano, Systems biology: a brief overview., Science 295 (5560) (2002) 1662–1664. doi:http://dx.doi.org/10.1126/science.1069492.

[3] Z. N. Oltvai, A. L. Barabási, Systems biology. life's complexity pyramid., Science 298 (5594) (2002) 763–764. doi:http://dx.doi.org/10.1126/science.1078563.

[4] A. Aderem, Systems biology: Its practice and challenges, Cell 121 (2005) 511–513. doi:http://dx.doi.org/10.1016/j.cell.2005.04.020.

[5] A. Trewavas, A brief history of systems biology, Plant Cell 18 (10) (2006) 2420–2430. doi:http://dx.doi.org/10.1105/tpc.106.042267.

[6] G. D. Bader, M. P. Cary, C. Sander, Pathguide: a pathway resource list., Nucleic Acids Res 34 (Database issue) (2006) D504–D506. doi:http://dx.doi.org/10.1093/nar/gkj126.
URL `http://pathguide.org`

[7] M. Kanehisa, M. Araki, S. Goto, M. Hattori, M. Hirakawa, M. Itoh, T. Katayama, S. Kawashima, S. Okuda, T. Tokimatsu, Y. Yamanishi, Kegg for linking genomes to life and the environment, Nucl. Acids Res. 36 (Database issue) (2008) D480–D484. doi:http://dx.doi.org/10.1093/nar/gkm882.

[8] R. Caspi, H. Foerster, C. A. Fulcher, P. Kaipa, M. Krummenacker, M. Latendresse, S. Paley, S. Y. Rhee, A. G. Shearer, C. Tissier, T. C. Walk, P. Zhang, P. D. Karp, The metacyc database of metabolic pathways and enzymes and the biocyc collection of pathway/genome databases, Nucl. Acids Res. 36 (Database issue) (2008) D623–D631. doi:http://dx.doi.org/10.1093/nar/gkm900.

[9] P. D. Karp, C. A. Ouzounis, C. M. Kochlacs, L. Goldovsky, P. Kaipa, D. Ahren, S. Tsoka, N. Darzentas, V. Kunin, N. L. Bigas, Expansion of the biocyc collection of pathway/genome databases to 160 genomes, Nucl. Acids Res. 33 (19) (2005) 6083–6089. doi:http://dx.doi.org/10.1093/nar/gki892.

[10] M. Kanehisa, S. Goto, S. Kawashima, A. Nakaya, The kegg databases at genomenet, Nucl. Acids Res. 30 (1) (2002) 42–46. doi:http://dx.doi.org/10.1093/nar/30.1.42.
URL `http://www.genome.ad.jp/kegg/`

[11] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, Cytoscape: a software environment for integrated models of biomolecular interaction networks., Genome Res 13 (11) (2003) 2498–2504. doi:http://dx.doi.org/10.1101/gr.1239303.
URL http://cytoscape.org/

[12] Science signaling: Signal transduction knowledge environment.
URL http://stke.sciencemag.org

[13] Pathway interaction database.
URL http://pid.nci.nih.gov

[14] R. Elkon, R. Vesterman, N. Amit, I. Ulitsky, I. Zohar, M. Weisz, G. Mass, N. Orlev, G. Sternberg, R. Blekhman, J. Assa, Y. Shiloh, R. Shamir, Spike - a database, visualization and analysis tool of cellular signaling pathways, BMC Bioinformatics 9 (1) (2008) 110. doi:http://dx.doi.org/10.1186/1471-2105-9-110.
URL http://www.cs.tau.ac.il/ spike

[15] T. I. Lee, N. J. Rinaldi, F. Robert, D. T. Odom, Z. B. Joseph, G. K. Gerber, N. M. Hannett, C. T. Harbison, C. M. Thompson, I. Simon, J. Zeitlinger, E. G. Jennings, H. L. Murray, B. D. Gordon, B. Ren, J. J. Wyrick, J. B. Tagne, T. L. Volkert, E. Fraenkel, D. K. Gifford, R. A. Young, Transcriptional regulatory networks in saccharomyces cerevisiae, Science 298 (5594) (2002) 799–804. doi:http://dx.doi.org/10.1126/science.1075090.
URL http://web.wi.mit.edu/young/regulatory_network/

[16] S. G. Castro, V. J. Jacinto, M. P. Gil, A. S. Zavaleta, Monica, B. C. Moreira, J. S. Salazar, L. M. Rascado, I. M. Flores, H. Salgado, C. B. Martinez, C. A. Goodger, C. R. Penagos, J. M. Rios, E. Morett, E. Merino, A. M. Huerta, L. T. Quintanilla, J. C. Vides, Regulondb (version 6.0): gene regulation model of escherichia coli k-12 beyond transcription, active (experimental) annotated promoters and textpresso navigation, Nucl. Acids Res. 36 (Database issue) (2008) D120–D124. doi:http://dx.doi.org/10.1093/nar/gkm994.

[17] B. A. Shoemaker, A. R. Panchenko, Deciphering protein-protein interactions. part i. experimental techniques and databases, PLoS Computational Biology 3 (3) (2007) e42. doi:http://dx.doi.org/10.1371/journal.pcbi.0030042.

[18] C. von Mering, R. Krause, B. Snel, M. Cornell, S. G. Oliver, S. Fields, P. Bork, Comparative assessment of large-scale data sets of protein-protein interactions., Nature 417 (6887) (2002) 399–403. doi:http://dx.doi.org/10.1038/nature750.

[19] J. S. Bader, A. Chaudhuri, J. M. Rothberg, J. Chant, Gaining confidence in high-throughput protein interaction networks., Nat Biotechnol 22 (1) (2004) 78–85. doi:http://dx.doi.org/10.1038/nbt924.

[20] H. Yu, P. Braun, M. A. Yildirim, I. Lemmens, K. Venkatesan, J. Sahalie, T. Hirozane-Kishikawa, F. Gebreab, N. Li, N. Simonis, T. Hao, J.-F. Rual, A. Dricot, A. Vazquez, R. R. Murray, C. Simon, L. Tardivo, S. Tam, N. Svrzikapa, C. Fan, A.-S. de Smet, A. Motyl, M. E. Hudson, J. Park, X. Xin, M. E. Cusick, T. Moore,

C. Boone, M. Snyder, F. P. Roth, A.-L. Barabasi, J. Tavernier, D. E. Hill, M. Vidal, High-quality binary protein interaction map of the yeast interactome network, Science 322 (5898) (2008) 104–110. doi:http://dx.doi.org/10.1126/science.1158684.

[21] The international molecular exchange consortium.
URL `http://imex.sourceforge.net`

[22] L. Salwinski, C. S. Miller, A. J. Smith, F. K. Pettit, J. U. Bowie, D. Eisenberg, The database of interacting proteins: 2004 update, Nucl. Acids Res. 32 (Database issue) (2004) D449–D451. doi:http://dx.doi.org/10.1093/nar/gkh086.

[23] Kerrien, S., Alam-Faruque, Y., Aranda, B., Bancarz, I., Bridge, A., Derow, C., Dimmer, E., Feuermann, M., Friedrichsen, A., Huntley, R., Kohler, C., Khadake, J., Leroy, C., Liban, A., Lieftink, C., Montecchi-Palazzi, L., Orchard, S., Risse, J., Robbe, K., Roechert, B., Thorneycroft, D., Zhang, Y., Apweiler, R., Hermjakob, H., Intact–open source resource for molecular interaction data, Nucleic Acids Research 35 (Database issue) (2007) D561–D565. doi:http://dx.doi.org/10.1093/nar/gkl958.
URL `http://www.ebi.ac.uk/intact`

[24] A. C. Aryamontri, A. Ceol, L. M. Palazzi, G. Nardelli, M. V. Schneider, L. Castagnoli, G. Cesareni, Mint: the molecular interaction database, Nucl. Acids Res. 35 (Database issue) (2007) D572–574. doi:http://dx.doi.org/10.1093/nar/gkl950.
URL `http://mint.bio.uniroma2.it/mint/`

[25] U. Güldener, M. Münsterkötter, M. Oesterheld, P. Pagel, A. Ruepp, H. W. Mewes, V. Stümpflen, Mpact: the mips protein interaction resource on yeast., Nucleic Acids Res 34 (Database issue). doi:http://dx.doi.org/10.1093/nar/gkj003.
URL http://mips.gsf.de/genre/proj/mpact

[26] C. Stark, B. J. Breitkreutz, T. Reguly, L. Boucher, A. Breitkreutz, M. Tyers, Biogrid: a general repository for interaction datasets., Nucleic Acids Res 34 (Database issue). doi:http://dx.doi.org/10.1093/nar/gkj109.
URL http://www.thebiogrid.org

[27] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, The protein data bank., Nucleic Acids Res 28 (1) (2000) 235–242. doi:http://dx.doi.org/10.1093/nar/28.1.235.
URL http://www.pdb.org

[28] S. Mathivanan, B. Periaswamy, T. K. B. Gandhi, K. Kandasamy, S. Suresh, R. Mohmood, Y. L. Ramachandra, A. Pandey, An evaluation of human protein-protein interaction data in the public domain, BMC Bioinformatics 7 (Suppl 5) (2006) S19. doi:http://dx.doi.org/10.1186/1471-2105-7-S5-S19.

[29] L. J. Jensen, M. Kuhn, M. Stark, S. Chaffron, C. Creevey, J. Muller, T. Doerks, P. Julien, A. Roth, M. Simonovic, P. Bork, C. Mering, String 8–a global view on proteins and their functional interactions

in 630 organisms, Nucl. Acids Res. 37 (Database issue) (2008) D412–D416. doi:http://dx.doi.org/10.1093/nar/gkn760.
URL http://string-db.org/

[30] C. von Mering, L. J. Jensen, M. Kuhn, S. Chaffron, T. Doerks, B. Kruger, B. Snel, P. Bork, String 7–recent developments in the integration and prediction of protein interactions, Nucl. Acids Res. 35 (Database issue) (2007) D358–D362. doi:http://dx.doi.org/10.1093/nar/gkl825.

[31] K. T. S. Prasad, R. Goel, K. Kandasamy, S. Keerthikumar, S. Kumar, S. Mathivanan, D. Telikicherla, R. Raju, B. Shafreen, A. Venugopal, L. Balakrishnan, A. Marimuthu, S. Banerjee, D. S. Somanathan, A. Sebastian, S. Rani, S. Ray, H. C. J. Kishore, S. Kanth, M. Ahmed, M. K. Kashyap, R. Mohmood, Y. L. Ramachandra, V. Krishna, A. B. Rahiman, S. Mohan, P. Ranganathan, S. Ramabadran, R. Chaerkady, A. Pandey, Human protein reference database–2009 update, Nucl. Acids Res. 37 (Database issue) (2008) D767–D772. doi:http://dx.doi.org/10.1093/nar/gkn892.
URL http://www.hprd.org/

[32] E. Klipp, W. Liebermeister, A. Helbig, A. Kowald, J. Schaber, Systems biology standards the community speaks, Nature Biotechnology 25 (4) (2007) 390–391. doi:http://dx.doi.org/10.1038/nbt0407-390.

[33] M. Hucka, A. Finney, H. M. Sauro, H. Bolouri, J. C. Doyle, H. Kitano, The, A. P. Arkin, B. J. Bornstein, D. Bray, C. A. Bowden, A. A. Cuellar, S. Dronov, E. D. Gilles, M. Ginkel, V. Gor, I. I.

Goryanin, W. J. Hedley, T. C. Hodgman, J. H. Hofmeyr, P. J. Hunter, N. S. Juty, J. L. Kasberger, A. Kremling, U. Kummer, N. Le Novere, L. M. Loew, D. Lucio, P. Mendes, E. Minch, E. D. Mjolsness, Y. Nakayama, M. R. Nelson, P. F. Nielsen, T. Sakurada, J. C. Schaff, B. E. Shapiro, T. S. Shimizu, H. D. Spence, J. Stelling, K. Takahashi, M. Tomita, J. Wagner, J. Wang, The systems biology markup language (sbml): a medium for representation and exchange of biochemical network models, Bioinformatics 19 (4) (2003) 524–531. doi:http://dx.doi.org/10.1093/bioinformatics/btg015.
URL `http://www.sbml.org/`

[34] A. Garny, D. P. P. Nickerson, J. Cooper, R. W. D. W. Santos, A. K. K. Miller, S. McKeever, P. M. F. M. Nielsen, P. J. J. Hunter, Cellml and associated tools and techniques., Philosophical transactions. Series A, Mathematical, physical, and engineering sciences 366 (1878) (2008) 3017–3043. doi:http://dx.doi.org/10.1098/rsta.2008.0094.
URL `http://www.cellml.org/`

[35] Biopax.
URL `http://www.biopax.org`

[36] Extensible markup language (xml).
URL `http://www.w3.org/XML/`

[37] I. Vastrik, P. D'Eustachio, E. Schmidt, G. J. Tope, G. Gopinath, D. Croft, B. de Bono, M. Gillespie, B. Jassal, S. Lewis, L. Matthews, G. Wu, E. Birney, L. Stein, Reactome: a knowledge base

of biologic pathways and processes, Genome Biology 8 (R39). doi:http://dx.doi.org/10.1186/gb-2007-8-3-r39.

[38] C. M. M. Lloyd, J. R. R. Lawson, P. J. J. Hunter, P. F. F. Nielsen, The cellml model repository., Bioinformatics 24 (18) (2008) 2122–2123. doi:http://dx.doi.org/10.1093/bioinformatics/btn390.
URL http://www.cellml.org/models

[39] N. Le Novère, B. Bornstein, A. Broicher, M. Courtot, M. Donizelli, H. Dharuri, L. Li, H. Sauro, M. Schilstra, B. Shapiro, J. L. Snoep, M. Hucka, Biomodels database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems, Nucleic Acids Res 34 (Database issue) (2006) D689–D691. doi:http://dx.doi.org/10.1093/nar/gkj092.
URL http://www.ebi.ac.uk/biomodels/

[40] A. R. Pico, T. Kelder, M. P. van Iersel, K. Hanspers, B. R. Conklin, C. Evelo, Wikipathways: Pathway editing for the people, PLoS Biology 6. doi:http://dx.doi.org/10.1371/journal.pbio.0060184.
URL http://wikipathways.org/

[41] Pathway commons.
URL http://www.pathwaycommons.org

[42] P. Erdős, A. Rényi, On random graphs, Publicationes Mathematicae Debrecen 6 (1959) 290–297.

[43] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, A. L. Barabasi, The

large-scale organization of metabolic networks, Nature 407 (2000) 651–654. doi:http://dx.doi.org/10.1038/35036627.

[44] A. L. Barabási, R. Albert, Emergence of scaling in random networks, Science 286 (5439) (1999) 509–512. doi:http://dx.doi.org/10.1126/science.286.5439.509.

[45] R. Albert, A. L. Barabási, Statistical mechanics of complex networks, Reviews of Modern Physics 74 (1) (2002) 47–97. doi:http://dx.doi.org/10.1103/RevModPhys.74.47.

[46] R. Albert, Scale-free networks in cell biology, J Cell Sci 118 (2005) 4947–4957. doi:http://dx.doi.org/10.1242/jcs.02714.

[47] A. L. Barabasi, Z. N. Oltvai, Network biology: understanding the cell's functional organization, Nat Rev Genet 5 (2004) 101–113. doi:http://dx.doi.org/10.1038/nrg1272.

[48] S. H. Strogatz, Exploring complex networks, Nature 410 (2001) 268–276. doi:http://dx.doi.org/10.1038/35065725.

[49] V. Lacroix, L. Cottret, P. Thébault, M. F. Sagot, An introduction to metabolic networks and their structural analysis, IEEE/ACM Transactions on Computational Biology and Bioinformatics 28.

[50] M. P. H. Stumpf, C. Wiuf, Robert, Subnets of scale-free networks are not scale-free: Sampling properties of networks, Proceedings of the National Academy of Sciences 102 (12) (2005) 4221–4224.

[51] J.-D. J. Han, D. Dupuy, N. Bertin, M. E. Cusick, M. Vidal, Effect of sampling on topology predictions of protein-protein interaction networks, Nature Biotechnology 23 (7) (2005) 839–844. doi:http://dx.doi.org/10.1038/nbt1116.

[52] N. Przulj, D. G. Corneil, I. Jurisica, Modeling interactome: scale-free or geometric?, Bioinformatics 20 (18) (2004) 3508–3515. doi:http://dx.doi.org/10.1093/bioinformatics/bth436.

[53] M. Arita, The metabolic world of escherichia coli is not small, PNAS 101 (6) (2004) 1543–1547.

[54] E. F. Keller, Revisiting "scale-free" networks, BioEssays 27 (10) (2005) 1060–1068. doi:http://dx.doi.org/10.1002/bies.20294.

[55] M. Arita, Scale-freeness and biological networks, J Biochem 138 (1) (2005) 1–4. doi:http://dx.doi.org/10.1093/jb/mvi094.

[56] R. Milo, S. S. Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, U. Alon, Network motifs: Simple building blocks of complex networks, Science 298 (2002) 824–827. doi:http://dx.doi.org/10.1126/science.298.5594.824.

[57] Shai, R. Milo, S. Mangan, U. Alon, Network motifs in the transcriptional regulation network of escherichia coli, Nat Genet 31 (2002) 64–68. doi:http://dx.doi.org/10.1038/ng881.

[58] I. Albert, R. Albert, Conserved network motifs allow protein-protein interaction prediction, Bioinformatics 20 (18) (2004) 3346–3352. doi:http://dx.doi.org/10.1093/bioinformatics/bth402.

[59] A. Vespignani, Evolution thinks modular, Nat Genet 35 (2) (2003) 118–119. doi:http://dx.doi.org/10.1038/ng1003-118.

[60] S. Wuchty, Z. N. Oltvai, A. L. Barabasi, Evolutionary conservation of motif constituents in the yeast protein interaction network, Nat Genet 35 (2) (2003) 176–179. doi:http://dx.doi.org/10.1038/ng1242.

[61] G. C. Conant, A. Wagner, Convergent evolution of gene circuits, Nat Genet 34 (2003) 264–266. doi:http://dx.doi.org/10.1038/ng1181.

[62] F. Schreiber, H. Schwöbbermeyer, Frequency concepts and pattern detection for the analysis of motifs in networks, Transactions on Computational Systems Biology III 3737 (2005) 89–104. doi:http://dx.doi.org/10.1007/11599128_7.

[63] G. Ciriello, C. Guerra, A review on models and algorithms for motif discovery in protein-protein interaction networks, Brief Funct Genomic Proteomic 7 (2) (2008) 147–156. doi:http://dx.doi.org/10.1093/bfgp/eln015.

[64] K. Baskerville, M. Paczuski, Subgraph ensembles and motif discovery using an alternative heuristic for graph isomorphism, Physical Review E (Statistical, Nonlinear, and Soft Matter Physics) 74 (2006) 051903–11. doi:http://dx.doi.org/10.1103/PhysRevE.74.051903.

[65] N. Kashtan, S. Itzkovitz, R. Milo, U. Alon, Efficient sampling algorithm for estimating subgraph concentrations and detecting network motifs, Bioinformatics 20 (11) (2004) 1746–1758. doi:http://dx.doi.org/10.1093/bioinformatics/bth163.

[66] S. Wernicke, F. Rasche, Simple and fast alignment of metabolic pathways by exploiting local diversity, Bioinformatics 23 (15) (2007) 1978–1985. doi:http://dx.doi.org/10.1093/bioinformatics/btm279.

[67] J. Chen, W. Hsu, M. L. Lee, S. K. Ng, Nemofinder: dissecting genome-wide protein-protein interactions with meso-scale network motifs, Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining (2006) 106–115doi:http://dx.doi.org/10.1145/1150402.1150418.

[68] E. Ziv, R. Koytcheff, M. Middendorf, C. Wiggins, Systematic identification of statistically significant network measures, Physical Review E (Statistical, Nonlinear, and Soft Matter Physics) 71 (2005) 016110–8. doi:http://dx.doi.org/10.1103/PhysRevE.71.016110.

[69] J. Grochow, M. Kellis, Network motif discovery using subgraph enumeration and symmetry-breaking, Research in Computational Molecular Biology 4453 (2007) 92–106. doi:http://dx.doi.org/10.1007/978-3-540-71681-5_7.

[70] M. E. J. Newman, The structure and function of complex networks, SIAM Review 45 (2) (2003) 167–256.

[71] Y. A. Randrup, S. J. Fleishman, N. B. Tal, L. Stone, Comment on "network motifs: Simple building blocks of complex networks" and "superfamilies of evolved and designed networks", Science 305 (5687). doi:http://dx.doi.org/10.1126/science.1099334.

[72] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool., J Mol Biol 215 (3) (1990) 403–410. doi:http://dx.doi.org/10.1006/jmbi.1990.9999.

[73] G. Cesareni, A. Ceol, C. Gavrila, L. M. Palazzi, M. Persico, M. V. Schneider, Comparative interactomics, FEBS Letters 579 (8) (2005) 1828–1833. doi:http://dx.doi.org/10.1016/j.febslet.2005.01.064.

[74] R. Y. Pinter, O. Rokhlenko, E. Yeger-Lotem, M. Ziv-Ukelson, Alignment of metabolic pathways, Bioinformatics 21 (16) (2005) 3401–3408. doi:http://dx.doi.org/10.1093/bioinformatics/bti554.

[75] R. Sharan, T. Ideker, Modeling cellular machinery through biological network comparison, Nature Biotechnology 24 (4) (2006) 427–433. doi:http://dx.doi.org/10.1038/nbt1196.

[76] B. P. Kelley, R. Sharan, R. M. Karp, T. Sittler, D. E. Root, B. R. Stockwell, T. Ideker, Conserved pathways within bacteria and yeast as revealed by global protein network alignment, Proceedings of the National Academy of Sciences 100 (20) (2003) 11394–11399.

[77] R. Sharan, S. Suthram, R. M. Kelley, T. Kuhn, S. Mccuine, P. Uetz, T. Sittler, R. M. Karp, T. Ideker, Conserved patterns of protein interaction in multiple species, Proceedings of the National Academy of Sciences 102 (6) (2005) 1974–1979. doi:http://dx.doi.org/10.1073/pnas.0409522102.

[78] J. Flannick, A. Novak, B. S. Srinivasan, H. H. Mcadams, S. Batzoglou, Graemlin: General and robust alignment of multiple

large interaction networks, Genome Res. 16 (2006) 1169–1181. doi:http://dx.doi.org/10.1101/gr.5235706.

[79] Q. Yang, S. H. Sze, Path matching and graph matching in biological networks, Journal of Computational Biology 14 (1) (2007) 56–67.

[80] A. M. Turing, The chemical basis of morphogenesis, Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences (1934-1990) 237 (641) (1952) 37–72. doi:http://dx.doi.org/10.1098/rstb.1952.0012.

[81] R. Alves, F. Antunes, A. Salvador, Tools for kinetic modeling of biochemical networks, Nature Biotechnology 24 (6) (2006) 667–672. doi:http://dx.doi.org/10.1038/nbt0606-667.

[82] S. Hoops, S. Sahle, R. Gauges, C. Lee, J. Pahle, N. Simus, M. Singhal, L. Xu, P. Mendes, U. Kummer, Copasia complex pathway simulator, Bioinformatics 22 (24) (2006) 3067–3074. doi:http://dx.doi.org/10.1093/bioinformatics/btl485.

[83] S. Ramsey, D. Orrell, H. Bolouri, Dizzy: stochastic simulation of large-scale genetic regulatory networks., J Bioinform Comput Biol 3 (2) (2005) 415–436.

[84] A. Pettinen, T. Aho, O.-P. Smolander, T. Manninen, A. Saarinen, K.-L. Taattola, O. Yli-Harja, M.-L. Linne, Simulation tools for biochemical networks: evaluation of performance and usability, Bioinformatics 21 (3) (2005) 357–363. doi:http://dx.doi.org/10.1093/bioinformatics/bti018.

[85] G. Karlebach, R. Shamir, Modelling and analysis of gene regulatory networks, Nat Rev Mol Cell Biol 9. doi:http://dx.doi.org/10.1038/nrm2503.

[86] H. de Jong, Modeling and simulation of genetic regulatory systems: a literature review., J Comput Biol 9 (1) (2002) 67–103. doi:http://dx.doi.org/10.1089/10665270252833208.

[87] D. Gilbert, H. Fuß, X. Gu, R. Orton, S. Robinson, V. Vyshemirsky, M. J. J. Kurth, C. S. S. Downes, W. Dubitzky, Computational methodologies for modelling, analysis and simulation of signalling networks., Brief Bioinform 7 (4) (2006) 339–353. doi:http://dx.doi.org/10.1093/bib/bbl043.

[88] J. A. Papin, J. Stelling, N. D. Price, S. Klamt, S. Schuster, B. O. Palsson, Comparison of network-based pathway analysis methods., Trends Biotechnol 22 (8) (2004) 400–405. doi:http://dx.doi.org/10.1016/j.tibtech.2004.06.010.

[89] J. A. Papin, N. D. Price, S. J. Wiback, D. A. Fell, B. O. Palsson, Metabolic pathways in the post-genome era., Trends Biochem Sci 28 (5) (2003) 250–258.

[90] S. A. Kauffman, Metabolic stability and epigenesis in randomly constructed genetic nets, Journal of Theoretical Biology 22 (3) (1969) 437–467. doi:http://dx.doi.org/10.1016/0022-5193(69)90015-0.

[91] M. Chaves, R. Albert, E. D. Sontag, Robustness and fragility of boolean

models for genetic regulatory networks., J Theor Biol 235 (3) (2005) 431–449. doi:http://dx.doi.org/10.1016/j.jtbi.2005.01.023.

[92] A. Fauré, A. Naldi, C. Chaouiya, D. Thieffry, Dynamical analysis of a generic boolean model for the control of the mammalian cell cycle., Bioinformatics 22 (14). doi:http://dx.doi.org/10.1093/bioinformatics/btl210.

[93] A. Garg, A. Dicara, I. Xenarios, L. Mendoza, G. D. D. Micheli, Synchronous vs. asynchronous modeling of gene regulatory networks., Bioinformatics 24 (17) (2008) 1917–1925. doi:http://dx.doi.org/10.1093/bioinformatics/btn336.

[94] I. Shmulevich, E. R. Dougherty, S. Kim, W. Zhang, Probabilistic boolean networks: a rule-based uncertainty model for gene regulatory networks., Bioinformatics 18 (2) (2002) 261–274. doi:http://dx.doi.org/10.1093/bioinformatics/18.2.261.

[95] S. Kauffman, C. Peterson, B. Samuelsson, C. Troein, Random boolean network models and the yeast transcriptional network., Proc Natl Acad Sci U S A 100 (25) (2003) 14796–14799. doi:http://dx.doi.org/10.1073/pnas.2036429100.

[96] R. Zhang, M. V. Shah, J. Yang, S. B. Nyland, X. Liu, J. K. Yun, R. Albert, T. P. Loughran, Network model of survival signaling in large granular lymphocyte leukemia, Proceedings of the National Academy of Sciences 105 (42) (2008) 16308–16313. doi:http://dx.doi.org/10.1073/pnas.0806447105.

[97] F. Li, T. Long, Y. Lu, Q. Ouyang, C. Tang, The yeast cell-cycle network is robustly designed, PNAS 101 (14) (2004) 4781–4786. doi:http://dx.doi.org/10.1073/pnas.0305937101.

[98] S. Klamt, J. Saez-Rodriguez, J. A. Lindquist, L. Simeoni, E. D. Gilles, A methodology for the structural and functional analysis of signaling and regulatory networks., BMC Bioinformatics 7 (1). doi:http://dx.doi.org/10.1186/1471-2105-7-56.

[99] S. Li, S. M. Assmann, R. Albert, Predicting essential components of signal transduction networks: A dynamic model of guard cell abscisic acid signaling, PLoS Biology 4 (10) (2006) e312. doi:http://dx.doi.org/10.1371/journal.pbio.0040312.

[100] C. A. Petri, Kommunikation mit Automaten, Ph.D. thesis, Institut für instrumentelle Mathematik, Bonn (1962).

[101] T. Murata, Petri nets: Properties, analysis and applications, Proceedings of the IEEE 77 (4) (1989) 541–580. doi:http://dx.doi.org/10.1109/5.24143.

[102] L. M. Kristensen, S. Christensen, K. Jensen, The practitioners guide to coloured petri nets, International Journal on Software Tools for Technology Transfer (STTT) 2 (2) (1998) 98–132. doi:http://dx.doi.org/10.1007/s100090050021.

[103] K. Jensen, Coloured petri nets, Petri Nets: Central Models and Their Properties 254 (1987) 248–299. doi:http://dx.doi.org/10.1007/BFb0046842.

[104] G. Balbo, Introduction to generalized stochastic petri nets, Formal Methods for Performance Evaluation 4486 (2007) 83–131. doi:http://dx.doi.org/10.1007/978-3-540-72522-0_3.

[105] F. Bause, P. S. Kritzinger, Stochastic petri nets: An introduction to the theory, SIGMETRICS Perform. Eval. Rev. 26 (2) (1998) 2–3. doi:http://dx.doi.org/10.1145/288197.581194.

[106] H. Alla, R. David, Continuous and hybrid petri nets, Journal of Circuits, Systems, and Computers 8 (1) (1998) 159–188.

[107] H. Matsuno, Y. Tanaka, H. Aoshima, A. Doi, M. Matsui, S. Miyano, Biopathways representation and simulation on hybrid functional petri net, In Silico Biol 3 (3) (2003) 389–404.

[108] D. Ruths, M. Muller, J. T. Tseng, L. Nakhleh, P. T. Ram, The signaling petri net-based simulator: a non-parametric strategy for characterizing the dynamics of cell-specific signaling networks, PLoS computational biology 4 (2). doi:http://dx.doi.org/10.1371/journal.pcbi.1000005.

[109] E. Simão, E. Remy, D. Thieffry, C. Chaouiya, Qualitative modelling of regulated metabolic pathways: application to the tryptophan biosynthesis in e.coli, Bioinformatics 21 Suppl 2 (2005) 190–196. doi:http://dx.doi.org/10.1093/bioinformatics/bti1130.

[110] I. Koch, B. H. Junker, M. Heiner, Application of petri net theory for modelling and validation of the sucrose breakdown pathway in the potato tuber, Bioinformatics 21 (7) (2005) 1219–1226. doi:http://dx.doi.org/10.1093/bioinformatics/bti145.

[111] R. Küffner, R. Zimmer, T. Lengauer, Pathway analysis in metabolic databases via differential metabolic display (dmd)., Bioinformatics 16 (9) (2000) 825–836. doi:http://dx.doi.org/10.1093/bioinformatics/16.9.825.

[112] Z. I. Oancea, S. Schuster, Topological analysis of metabolic networks based on petri net theory, In Silico Biol 3 (3) (2003) 323–345.

[113] M. Chen, R. Hofestädt, Quantitative petri net model of gene regulated metabolic networks in the cell, In Silico Biol 3 (3) (2003) 347–365.

[114] K. Voss, M. Heiner, I. Koch, Steady state analysis of metabolic pathways using petri nets, In Silico Biol 3 (3) (2003) 367–387.

[115] J. Kielbassa, R. Bortfeldt, S. Schuster, I. Koch, Modeling of the u1 snrnp assembly pathway in alternative splicing in human cells using petri nets, Computational Biology and Chemistry 33 (1) (2008) 46–61. doi:http://dx.doi.org/10.1016/j.compbiolchem.2008.07.022.

[116] L. J. Steggles, R. Banks, O. Shaw, A. Wipat, Qualitatively modelling and analysing genetic regulatory networks: a petri net approach, Bioinformatics 23 (3) (2007) 336–343. doi:http://dx.doi.org/10.1093/bioinformatics/btl596.

[117] H. Matsuno, A. Doi, M. Nagasaki, S. Miyano, Hybrid petri net representation of gene regulatory network, Pac Symp Biocomput (2000) 341–352.

[118] A. Sackmann, M. Heiner, I. Koch, Application of petri net based analysis techniques to signal transduction pathways, BMC Bioinformatics 7 (1) (2006) 482. doi:http://dx.doi.org/10.1186/1471-2105-7-482.

[119] C. Li, S. Suzuki, Q. W. Ge, M. Nakata, H. Matsuno, S. Miyano, Structural modeling and analysis of signaling pathways based on petri nets, J Bioinform Comput Biol 4 (5) (2006) 1119–1140.

[120] T. T. Igarashi, Ontology based standardization of petri net modeling for signaling pathways, In Silico Biol 5 (2005) 0047.

[121] R. David, H. Alla, Discrete, Continuous, and Hybrid Petri Nets, 1st Edition, Springer, 2004.

[122] C. Chaouiya, Petri net modelling of biological networks, Brief Bioinform 8 (4) (2007) 210–219. doi:http://dx.doi.org/10.1093/bib/bbm029.

[123] M. Peleg, D. Rubin, R. B. Altman, Using petri net tools to study properties and dynamics of biological systems., J Am Med Inform Assoc 12 (2) (2005) 181–199. doi:http://dx.doi.org/10.1197/jamia.M1637.

[124] S. Hardy, P. N. Robillard, Modeling and simulation of molecular biology systems using petri nets: modeling goals of various approaches, J Bioinform Comput Biol 2 (4) (2004) 595–613.

[125] J. W. Pinney, D. R. Westhead, G. A. Mcconkey, Petri net representations in systems biology, Biochem Soc Trans 31 (Pt 6) (2003) 1513–1515. doi:http://dx.doi.org/10.1042/BST0311513.

[126] E. Klipp, R. Herwig, A. Kowald, C. Wierling, H. Lehrach, Systems Biology in Practice: Concepts, Implementation and Application, Wiley-VCH, 2005.

[127] H. de Jong, D. Ropers, Strategies for dealing with incomplete information in the modeling of molecular interaction networks., Brief Bioinform 7 (4) (2006) 354–363. doi:http://dx.doi.org/10.1093/bib/bbl034.

[128] B. Teusink, J. Passarge, C. A. Reijenga, E. Esgalhado, C. C. van der Weijden, M. Schepper, M. C. Walsh, B. M. Bakker, K. van Dam, H. V. Westerhoff, J. L. Snoep, Can yeast glycolysis be understood in terms of in vitro kinetics of the constituent enzymes? testing biochemistry, European Journal of Biochemistry 267 (17) (2000) 5313–5329. doi:http://dx.doi.org/10.1046/j.1432-1327.2000.01527.x.

[129] M. Borisuk, Bifurcation analysis of a model of mitotic control in frog eggs, Journal of Theoretical Biology 195 (1) (1998) 69–85. doi:http://dx.doi.org/10.1006/jtbi.1998.0781.

[130] K. C. Chen, A. Csikasz-Nagy, B. Gyorffy, J. Val, B. Novak, J. J. Tyson, Kinetic analysis of a molecular model of the budding yeast cell cycle, Mol. Biol. Cell 11 (1) (2000) 369–391.

[131] A. Csikasz-Nagy, D. Battogtokh, K. C. Chen, B. Novak, J. J. Tyson, Analysis of a generic model of eukaryotic cell-cycle regulation, Biophys. J. 90 (12) (2006) 4361–4379. doi:http://dx.doi.org/10.1529/biophysj.106.081240.

[132] W. J. Blake, G. Balazsi, M. A. Kohanski, F. J. Isaacs, K. F. Murphy, Y. Kuang, C. R. Cantor, D. R. Walt, J. J. Collins, Phenotypic consequences of promoter-mediated transcriptional noise, Molecular Cell 24 (2006) 853–865. doi:http://dx.doi.org/10.1016/j.molcel.2006.11.003.

[133] J. M. Raser, E. K. O'Shea, Control of stochasticity in eukaryotic gene expression, Science 304 (5678) (2004) 1811–1814. doi:http://dx.doi.org/10.1126/science.1098641.

[134] W. J. Blake, M. Kaern, C. R. Cantor, J. J. Collins, Noise in eukaryotic gene expression, Nature 422 (2003) 633–637. doi:http://dx.doi.org/10.1038/nature01546.

[135] H. Mcadams, A. Arkin, Stochastic mechanisms in geneexpression, PNAS 94 (3) (1997) 814–819.

[136] C. V. Rao, D. M. Wolf, A. P. Arkin, Control, exploitation and tolerance of intracellular noise, Nature 420 (6912) (2002) 231–237. doi:http://dx.doi.org/10.1038/nature01258.

[137] H. H. Mcadams, A. Arkin, It's a noisy business! genetic regulation at the nanomolar scale, Trends in Genetics 15 (2) (1999) 65–69. doi:http://dx.doi.org/10.1016/S0168-9525(98)01659-X.

[138] D. T. Gillespie, A general method for numerically simulating the stochastic time evolution of coupled chemical reactions, Journal of Computational Physics 22 (4) (1976) 403–434. doi:http://dx.doi.org/10.1016/0021-9991(76)90041-3.

[139] D. T. Gillespie, Stochastic simulation of chemical kinetics., Annu Rev Phys Chem 58 (2007) 35–55. doi:http://dx.doi.org/10.1146/annurev.physchem.58.032806.104637.

[140] L. Salwinski, D. Eisenberg, In silico simulation of biological network dynamics, Nat Biotech 22 (8) (2004) 1017–1019. doi:http://dx.doi.org/10.1038/nbt991.

[141] Y. Kaznessis, Multi-scale models for gene network engineering, Chemical Engineering Science 61 (3) (2006) 940–953. doi:http://dx.doi.org/10.1016/j.ces.2005.06.033.

[142] T. A. Henzinger, The theory of hybrid automata, Proceedings, Eleventh Annual IEEE Symposium on Logic in Computer Science (LICS) (1996) 278–292doi:http://dx.doi.org/10.1109/LICS.1996.561342.

[143] R. Alur, C. Courcoubetis, N. Halbwachs, T. A. Henzinger, Ho, X. Nicollin, A. Olivero, J. Sifakis, S. Yovine, The algorithmic analysis of hybrid systems, Theoretical Computer Science 138 (1) (1995) 3–34. doi:http://dx.doi.org/10.1016/0304-3975(94)00202-T.

[144] P. Lincoln, A. Tiwari, Symbolic systems biology: Hybrid modeling and analysis of biological networks, Hybrid Systems: Computation and Control 2993 (2004) 147–165.

[145] A. Casagrande, C. Piazza, A. Policriti, Discreteness, hybrid automata, and biology, 9th International Work-

shop on Discrete Event Systems (WODES) (2008) 281–286doi:http://dx.doi.org/10.1109/WODES.2008.4605960.

[146] C. Piazza, M. Antoniotti, V. Mysore, A. Policriti, F. Winkler, B. Mishra, Algorithmic algebraic model checking i: Challenges from systems biology, Computer Aided Verification 3576 (2005) 5–19. doi:http://dx.doi.org/10.1007/11513988_3.

[147] J. Fisher, T. A. Henzinger, Executable cell biology, Nat Biotech 25 (11) (2007) 1239–1249. doi:http://dx.doi.org/10.1038/nbt1356.

[148] M. Santillán, M. C. Mackey, Quantitative approaches to the study of bistability in the lac operon of escherichia coli., Journal of the Royal Society, Interface / the Royal Society 5 Suppl 1. doi:http://dx.doi.org/10.1098/rsif.2008.0086.focus.

[149] B. Müller-Hill, The Lac Operon: A Short History of a Genetic Paradigm, Walter de Gruyter, 1996.

[150] A. Halasz, V. Kumar, M. Imielinski, C. Belta, O. Sokolsky, S. Pathak, H. Rubin, Analysis of lactose metabolism in e.coli using reachability analysis of hybrid systems, IET Systems Biology 1 (2) (2007) 130–148. doi:http://dx.doi.org/10.1049/iet-syb:20060035.

[151] R. Ghosh, A. Tiwari, C. Tomlin, Automated symbolic reachability analysis; with application to delta-notch signaling automata, Hybrid Systems: Computation and Control 2623 (2003) 233–248. doi:http://dx.doi.org/10.1007/3-540-36580-X_19.

[152] R. Ghosh, C. Tomlin, Symbolic reachable set computation of piecewise affine hybrid automata and its application to biological modelling: Delta-notch protein signalling, Systems Biology 1 (1) (2004) 170–183. doi:http://dx.doi.org/10.1049/sb:20045019.

[153] C. Belta, P. Finin, L. C. Habets, A. M. Halász, M. Imielinski, V. R. Kumar, H. Rubin, Understanding the bacterial stringent response using reachability analysis of hybrid systems, Hybrid Systems: Computation and Control 2993 (2004) 107–116.

[154] S. Drulhe, G. Ferrari-Trecate, H. de Jong, A. Viari, Reconstruction of switching thresholds in piecewise-affine models of genetic regulatory networks, Hybrid Systems: Computation and Control 3927 (2006) 184–199. doi:http://dx.doi.org/10.1007/11730637_16.

[155] G. Batt, D. Ropers, H. de Jong, J. Geiselmann, M. Page, D. Schneider, Qualitative analysis and verification of hybrid models of genetic regulatory networks: Nutritional stress response in escherichia coli, Hybrid Systems: Computation and Control 3414 (2005) 134–150. doi:http://dx.doi.org/10.1007/b106766.

[156] R. David, Modeling of hybrid systems using continuous and hybrid petri nets, Proceedings of the Seventh International Workshop on Petri Nets and Performance Models (1997) 47–58doi:http://dx.doi.org/10.1109/PNPM.1997.595536.

[157] S. Pettersson, B. Lennartson, Hybrid modelling focused on hybrid petri

nets, 2nd European Workshop on Real-time and Hybrid systems (1995) 303–309.

[158] R. David, H. Alla, Petri nets and Grafcet : tools for modelling discrete event systems, Prentice Hall, New York, 1992.

[159] J.-M. Chandonia, S. E. Brenner, The impact of structural genomics: Expectations and outcomes, Science 311 (5759) (2006) 347–351. doi:http://dx.doi.org/10.1126/science.1121018.

[160] M. Stein, R. R. Gabdoulline, R. C. Wade, Bridging from molecular simulation to biochemical networks., Curr Opin Struct Biol 17 (2) (2007) 166–172. doi:http://dx.doi.org/10.1016/j.sbi.2007.03.014.

[161] D. Devos, R. B. B. Russell, A more complete, complexed and structured interactome, Curr Opin Struct Biol 17 (3) (2007) 370–377. doi:http://dx.doi.org/10.1016/j.sbi.2007.05.011.

[162] P. Beltrao, C. Kiel, L. Serrano, Structures in systems biology, Curr Opin Struct Biol 17 (3) (2007) 378–384. doi:http://dx.doi.org/10.1016/j.sbi.2007.05.005.

[163] P. Aloy, R. B. Russell, Structural systems biology: modelling protein interactions, Nat Rev Mol Cell Biol 7 (2006) 188–197. doi:http://dx.doi.org/10.1038/nrm1859.

[164] P. M. Kim, L. J. Lu, Y. Xia, M. B. Gerstein, Relating three-dimensional structures to protein networks provides evolutionary insights., Science 314 (5807) (2006) 1938–1941. doi:http://dx.doi.org/10.1126/science.1136174.

[165] C. C. Fowlkes, C. L. Hendriks, S. V. Keränen, G. H. Weber, O. Rübel, M. Y. Huang, S. Chatoor, A. H. DePace, L. Simirenko, C. Henriquez, A. Beaton, R. Weiszmann, S. Celniker, B. Hamann, D. W. Knowles, M. D. Biggin, M. B. Eisen, J. Malik, A quantitative spatiotemporal atlas of gene expression in the drosophila blastoderm., Cell 133 (2) (2008) 364–374. doi:http://dx.doi.org/10.1016/j.cell.2008.01.053.

[166] J.-D. J. Han, N. Bertin, T. Hao, D. S. Goldberg, G. F. Berriz, L. V. Zhang, D. Dupuy, A. J. M. Walhout, M. E. Cusick, F. P. Roth, M. Vidal, Evidence for dynamically organized modularity in the yeast protein-protein interaction network, Nature 430 (6995) (2004) 88–93. doi:http://dx.doi.org/10.1038/nature02555.

[167] B. C. Goodwin, S. A. Kauffman, Spatial harmonics and pattern specification in early drosophila development. part i. bifurcation sequences and gene expression., Journal of theoretical biology 144 (3) (1990) 303–319.

[168] C. L. Myers, O. G. Troyanskaya, Context-sensitive data integration and prediction of biological networks, Bioinformatics 23 (17) (2007) 2322–2330. doi:http://dx.doi.org/10.1093/bioinformatics/btm332.

[169] J. Rachlin, D. D. Cohen, C. Cantor, S. Kasif, Biological context networks: a mosaic view of the interactome, Mol Syst Biol 2 (2006) 66. doi:http://dx.doi.org/10.1038/msb4100103.

[170] C.-S. Chin, V. Chubukov, E. R. Jolly, J. Derisi, H. Li, Dynamics and design principles of a basic regulatory architecture con-

trolling metabolic pathways, PLoS Biology 6 (6) (2008) e146+. doi:http://dx.doi.org/10.1371/journal.pbio.0060146.

[171] T. A. Knijnenburg, L. F. A. Wessels, M. J. T. Reinders, Combinatorial influence of environmental parameters on transcription factor activity, Bioinformatics 24 (13) (2008) i172–181. doi:http://dx.doi.org/10.1093/bioinformatics/btn155.

[172] V. Spirin, M. S. S. Gelfand, A. A. A. Mironov, L. A. A. Mirny, A metabolic network in the evolutionary context: Multiscale structure and modularity, Proc Natl Acad Sci U S A 103 (23) (2006) 8774–8779. doi:http://dx.doi.org/10.1073/pnas.0510258103.

[173] D. Hwang, A. G. G. Rust, S. Ramsey, J. J. J. Smith, D. M. M. Leslie, A. D. D. Weston, P. de Atauri, J. D. D. Aitchison, L. Hood, A. F. F. Siegel, H. Bolouri, A data integration methodology for systems biology, Proc Natl Acad Sci U S A 102 (48) (2005) 17296–17301. doi:http://dx.doi.org/10.1073/pnas.0508647102.

[174] D. Hwang, J. J. Smith, D. M. Leslie, A. D. Weston, A. G. Rust, S. Ramsey, P. de Atauri, A. F. Siegel, H. Bolouri, J. D. Aitchison, L. Hood, A data integration methodology for systems biology: Experimental verification, PNAS 102 (48) (2005) 17302–17307. doi:http://dx.doi.org/10.1073/pnas.0508649102.