

Dialogue on Reverse-Engineering Assessment and Methods

The DREAM of High-Throughput Pathway Inference

GUSTAVO STOLOVITZKY,^a DON MONROE,^b AND ANDREA CALIFANO^c

^a*IBM Computational Biology Center, Yorktown Heights, New York 10598, USA*

^b*Science Writer, Berkeley Heights, New Jersey, USA*

^c*Department of Biomedical Informatics, Columbia University, New York, New York, USA*

ABSTRACT: The biotechnological advances of the last decade have confronted us with an explosion of genetics, genomics, transcriptomics, proteomics, and metabolomics data. These data need to be organized and structured before they may provide a coherent biological picture. To accomplish this formidable task, the availability of an accurate map of the physical interactions in the cell that are responsible for cellular behavior and function would be exceedingly helpful, as these data are ultimately the result of such molecular interactions. However, all we have at this time is, at best, a fragmentary and only partially correct representation of the interactions between genes, their byproducts, and other cellular entities. If we want to succeed in our quest for understanding the biological whole as more than the sum of the individual parts, we need to build more comprehensive and cell-context-specific maps of the biological interaction networks. DREAM, the Dialogue on Reverse Engineering Assessment and Methods, is fostering a concerted effort by computational and experimental biologists to understand the limitations and to enhance the strengths of the efforts to reverse engineer cellular networks from high-throughput data. In this chapter we will discuss the salient arguments of the first DREAM conference. We will highlight both the state of the art in the field of reverse engineering as well as some of its challenges and opportunities.

KEYWORDS: reverse engineering; pathway inference; DREAM conference

Address for correspondence: Gustavo Stolovitzky, IBM Computational Biology Center, P.O. Box 218, Yorktown Heights, NY 10598. Voice: (914) 945-1292; fax: (914) 945-4217.
gustavo@us.ibm.com

Ann. N.Y. Acad. Sci. 1115: 1–22 (2007). © 2007 New York Academy of Sciences.
doi: 10.1196/annals.1407.021

INTRODUCTION

The idea of a “systems” or holistic approach to biology was very much in the minds of physiologists several decades ago. So what is today’s systems-biology buzz all about? The driving force behind this burgeoning discipline is, in great part, the development of new tools that have opened the floodgates to data flowing from molecular events in the cellular realm.

Concerted efforts are now being made to create tools to consolidate, standardize, visualize, and manipulate knowledge from multiple sources.^{1,2} This should make the data easily available to improve understanding and to attempt the decoding of the design principles of biological pathways.

Although not a hard and fast rule, we can think of the many efforts in the field of systems biology as attempts to infer networks of interactions of cellular components from biological data, and use those networks as scaffolds on top of which increasingly more accurate predictive models can be built. For instance, one may attempt to use such predictive models to make hypotheses of how cells would react to exogenous perturbations and to design appropriate experiments to validate the model.

The complexities of biological systems are orchestrated by networks comprising thousands of interacting molecular species, including DNA, RNA, proteins, and smaller molecules. One of the goals of systems biology is thus to map these networks in ways that may provide fundamentally new understanding of cell biology at the molecular level. Unfortunately, even though modern biotechnology can provide rich data sets by simultaneously monitoring thousands of different types of molecules, dissecting the underlying network of physical interactions from these observations—that is, reverse engineering—remains a daunting challenge.

It is clear that a sense of the reliability of the reconstructed networks can only arise from our understanding of the limitations of reverse-engineering methods from precise quantitative metrics. In order to address this simple requirement we conceived the initiative we call DREAM (Dialogue on Reverse Engineering Assessment and Methods). The fundamental question for DREAM is simple: How can reverse engineers assess how well they are doing? The answer is not so simple. Researchers have used a variety of algorithms to deduce the structure of very different biological and artificial networks, and evaluated their success using various metrics. What is still needed, and what DREAM aspires to, is a fair comparison of the strengths and weaknesses of the methods and a clear sense of the reliability of the network models they produce.

On September 7–8, 2006, about 150 researchers gathered to discuss the reverse engineering of biological networks in the Wave Hill Conference Center, overlooking the Hudson River in the Bronx, New York. This first DREAM conference was co-organized by Gustavo Stolovitzky (IBM Research), Andrea Califano (Columbia University), and Jim Collins (Boston University). The meeting was sponsored by the Center for Discrete Mathematics and

Theoretical Computer Science (DIMACS), the Columbia University Center for the Multiscale Analysis of Genetic Networks (MAGNet), the NIH Roadmap Initiative, and the IBM Computational Biology Center.

DREAM conference attendees included computer scientists, physicists, mathematicians, and experimental biologists. In fact, systematically reaching out broadly to the community of experimental biologists is a key aim of the DREAM effort. During this meeting, although participants often emphasized contrasting aspects of the reverse-engineering challenge, some common themes emerged, both during the lectures and informal discussions:

- Even though many of the networks inferred by reverse-engineering methods may not be “realistic” (in the sense of mechanistically accurate), they can still be useful to understand cell function and behavior.
- It is critical to explore biological networks within a specific cellular/molecular context, because network topology and properties can be quite different in different tissues and under diverse biological conditions.
- Some well-understood biological networks may serve as “gold standards” for assessing the quality of reverse-engineering algorithms, but their identity is not easily hidden from researchers to create a truly “blind” test.
- Artificial networks, which are easily generated and blinded, will continue to be important for understanding the strengths and weaknesses of algorithms. However, experimental biologists are unlikely to trust any conclusions reached via these approaches.
- Assessment requires balancing errors in predicting interactions that are present as well as those that are not, because biological networks are sparsely connected.
- It may be time to take the first steps toward a formal assessment process, in which researchers test their methods against a set of common challenges.

Even though researchers have proposed a variety of algorithms for network reconstruction, comparing them on an objective basis can prove quite difficult. As the conference made clear, some of the challenges arise because the various methods target different problems. The problems at the time of algorithm-performance comparison differ in many dimensions, including:

- (1) sources of biological data (e.g., high-throughput screens or carefully validated interactions);
- (2) biological function of the network (e.g., metabolic, signaling, transcriptional, protein–protein interactions, etc.)
- (3) generative network (biological, mathematical, or *in vivo* synthetic);
- (4) types of perturbation (e.g., varied environmental conditions or knocking out entire nodes);
- (5) network behavior modeled (steady-state versus dynamic, continuum versus discrete, topological versus differential equations);

- (6) quality of data (including noise and coverage); and
- (7) metrics (e.g., topological versus quantitative, tradeoffs of accuracy and discrimination).

In the following sections we describe the salient arguments that emerged during the conference. They constitute a good snapshot of the state of the field, and at the same time they help understand the areas where a concerted effort will help advance the field.

TYPES OF NETWORK INFORMATION

What Data Best Capture Network Behavior?

Biological networks involve interactions between several types of molecular species, including DNA, RNA, proteins, and metabolites. The best-developed high-throughput technology, microarrays, lets researchers monitor average mRNA concentrations in a cell population on a genome-wide scale. Other high-throughput techniques, such as transcription-factor binding microarrays (ChIP-chip), compound library screening, protein and metabolite profiles, and genome-wide identification of protein–protein interactions and post-translational modifications, are less mature but improving quickly.

Researchers can also exploit reliable quantitative data from lower-throughput, more laborious techniques, such as functional reporter-gene assays. As they aid in reverse engineering massive networks, these smaller but better-understood systems can provide critical sanity checking. Other important clues come from evolutionary conservation, a property often exploited in other areas of biological research, where known conservation of an interaction in some species may be used to predict its existence in another species.

Most pathway inference approaches address a specific “layer” of biochemical interactions, such as transcriptional (protein–DNA), signaling (protein–protein, protein–DNA), protein complex (protein–protein), or metabolic (protein–metabolite) interactions. Different layers can have a different network character than that of gene expression networks. For example, as discussed by Pedro Mendes³ and Ilya Nemenman,⁴ metabolic networks tend to be resistant to change. They are constrained by a few co-factors that influence myriad reactions, and the corresponding graphs have many branches and loops.

The deepest biological insight will require combining different layers, including data on gene expression, proteins, and small molecules, among others. Several participants described efforts to do this: Allister Bernard,⁵ for example, combined electrophysiological and expression underlying songbird behavior in songbirds, with information on gene expression and protein–DNA interaction in yeast. Michael Samoilov⁶ also described an algorithm for reverse-engineering combined data.

Combining these layers can yield surprises, even when all the ingredients are known. For example, Pedro Mendes³ described a network including 20 genes, 23 proteins, and 16 metabolites, interacting by transcription, translation, metabolism, and signal transduction. Although simpler than real biological networks, this network showed unexpected behavior, including apparent connections between unconnected species, and proved quite resistant to early attempts to reverse engineer it. This issue begs the simple question of how to draw the basic network of interactions that best represents the behavior of the set of chemical reactions.

The emergence of RNA interference has highlighted a previously unrecognized layer of biological control. Networks that omit this layer must incorporate its influence through effective but indirect interactions. Indeed, since all data sources will be incomplete, a critical feature of reverse-engineering algorithms is whether they deal gracefully with missing data, whether from the same layer or from other layers. This issue of how to deal with a network with latent variables, that is, cellular players that are unknown and thus are not represented in the network, is addressed in Adam Margolin's article⁷ in this volume.

Types of Network Models

Biological networks are often represented as graphs, with “edges” connecting “nodes” (FIG. 1). This is not necessarily an ideal representation as often it may not even be clear whether the nodes correspond to a unique molecular species, such as a specific protein isoform in a given cellular compartment. Researchers, such as Dana Pe'er,⁸ for example, have found that reverse engineering is more robust at the level of “modules” of coherently regulated species. In that case, a node may represent an entire collection of mRNAs. Other researchers have suggested the use of individual domains and binding sites as network-nodes, achieving greater granularity of the representation.

Similarly, complex multivariate dependencies cannot be adequately represented and are shown as collection of pairwise relationships. In other cases, such as for protein–protein “interactomes,” an edge may have a more obvious meaning, indicating for instance that two proteins have been shown to form a complex in laboratory assays. The true significance of an edge, however, is more subtle, especially when multiple layers interact (for instance, a directed edge may indicate that the kinetics of one species is a function that depends on other species). As the field matures, researchers will need to be increasingly more precise about what “edges” represent in a particular context.

In many applications of reverse engineering, researchers attempt only to reconstruct the topology of the networks (FIG. 1A) rather than the nature of the individual relationships (i.e., the type of interactions and its kinetic constants). In these cases a graph, possibly a directed one indicating the direction of influence, constitutes an adequate representation, as in FIGURE 1B. A more

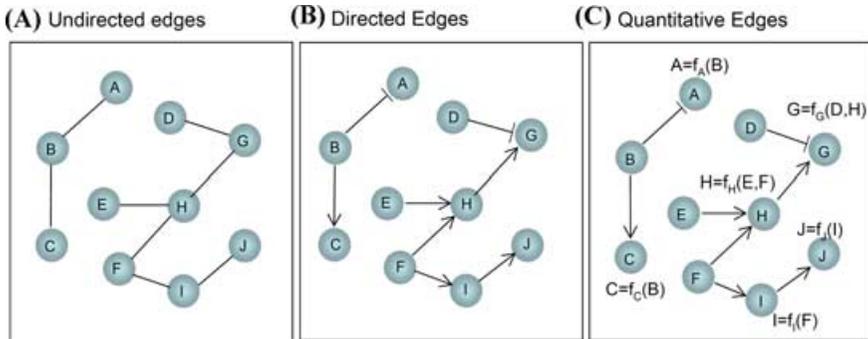


FIGURE 1. Types of network models in biology. Biological networks are often represented as graphs, with “edges” connecting “nodes.” The nodes can represent genes, proteins, metabolites, or other biological entities. The connections represent physical interactions (i.e., protein–protein binding or substrate–product linkages) or functional linkages (i.e., expression of gene D inhibits expression of gene G). In increasing order of detail, the edges can be: **(A) Undirected edges:** connections between the interacting nodes. **(B) Directed edges:** connections showing an indication of how incoming edges affect (possibly in a causal way) nodes. As is typically shown, arrows indicate positive or activating connections, and lines terminated with perpendicular segments indicate negative or inhibitory connections. **(C) Quantitative edges:** edges can also be endowed with a quantitative function of how the inputs control the output. For example, $G = f_G(D, H)$ would explicitly describe the co-dependence of the expression gene G with respect to expression levels of genes D and H.

complete recapitulation of the behavior of a biological network, however, requires a quantitative description (FIG. 1C), such as that provided by a set of ordinary differential equations, or even partial differential equations if the volume of the cell must be accounted for. Reverse engineering these models equates to not only finding the topology of the network, but also finding the kinetic constants in the equations. Both goals are obviously important, but the algorithms that attempt to achieve them must be assessed with different metrics.

Similarly, reproducing the dynamic response of a network emphasizes different aspects of a model than capturing its steady-state behavior. Representing network response discretely, perhaps as overexpression, underexpression, or no change, as done by Kundaje *et al.*⁹ is not as rich as a full quantitative description, but may be more robust in light of limited and noisy experiments.

How to Perturb Networks?

A classic way to reverse-engineering cellular networks or to test their power as predictive models is to perturb the cellular system and observe its response. A large set of gene expression profiles, for instance, corresponding to distinct

biochemical or environmental perturbations, can activate distinct pathways, forcing the cell to find new equilibrium points and thus providing much greater information about its dynamic landscape. Similarly, network models may be used to infer how the cell would react to specific perturbations, and thus one may be able to experimentally validate the predictions. There are basically three kinds of perturbations: biochemical, genetic, and environmental ones.

Biochemical perturbations are based on the administration of chemical compounds and tend to directly affect one or more pathways. For instance, one may use a ligand that binds a specific receptor or a nonspecific inhibitor, such as TSA, which affects an entire class of HDAC enzymes. Such an approach was recently used by the Golub lab¹⁰ to dissect the activity profile of a wide class of compounds in different cell lines.

Genetic perturbations are based on the use of specific molecules that can inhibit an individual gene (knockout) or on the introduction of constructs in the cell that allow the exogenous expression of a gene (knockin). This is often accomplished by transfecting cell lines with siRNA or expression vector constructs. Jim Collins, for example, described a set of experiments to use such techniques to study the regulatory networks of *E. coli* bacteria. While these perturbations should in theory be razor sharp, affecting only the gene of interest, the complexity of cellular networks can result in complex and unexpected effects that can both help and hinder the process of reverse engineering from perturbation data. Arnie Levine,^{11,12} for instance, emphasized that the effects of gene knockouts in mice are sometimes masked by the homeostatic machinery of the cell so that the transgenic animal is undistinguishable from its wild-type counterpart.

Environmental perturbations are related to the change of physiochemical variables in the cellular environment (e.g., growth media). For instance, changes in nutrient concentration, heat-shock, or oxygen concentration changes can all activate alternative “programs” of cellular responses, leading to a vast change in the repertoire of expressed genes.

Choosing the right perturbation set depends on the specific cellular context of interest since a compound that is active in a B cell may have no effect in a fibroblast.

HOW SHOULD THE DATA FOR REVERSE-ENGINEERING ASSESSMENT BE GENERATED?

Before we can assess how well or how poorly a reverse-engineering algorithm is performing on a network, we must ensure that we have accurate knowledge about the network itself. Specifically, we must make sure that we know precisely whether a specific interaction is present or not within the network. The collection of all interactions that are known to be present in a network is called a positive gold standard (GS+), while the collection of all

interactions that are known not to be present in a network is called the negative gold standard (GS⁻). Moreover, if we want the test to be truly blind, we must ensure that only the evaluation team knows the identity of the interactions in the gold-standard sets.

Creating an adequate set of gold standards for different problems in reverse engineering is probably the single biggest challenge of the DREAM initiative. The reason for this is manifold. First, most of the known networks are in the public domain. Thus it would be close to impossible to generate blind GS sets from them. Second, interactions that are experimentally validated may only be valid in a specific cellular context and may not be used in other contexts. Third, even experimentally validated interactions may be false positives. For instance, it is expected that close to 20% of all interactions predicted by the most accurate yeast-2-hybrids technology are false positives. Thus the positive gold standard may contain a significant number of interactions that are not really present in the network. Fourth, and most important, our network structure knowledge is highly incomplete and anywhere from 50% to 95% of the true interactions may be missing, leading to difficulties in compiling the negative gold-standard (i.e., which interactions are not present in the network).

The following sections in this chapter will describe some of the attempts by the DREAM community to work around these obstacles.

Synthetic in Silico Networks

In silico networks offer an ideal model of the network. Basically, all interactions are known exactly and thus the positive and negative gold-standard sets contain no errors. Furthermore, they can be simulated under any condition or perturbation, yielding a virtually infinite flexibility of the reverse-engineering assessment studies. Thus, these models are currently the only ones that offer a “guaranteed correct” gold standard to test reverse-engineering algorithms and that can be guaranteed to be really “blind” to anybody but the evaluation team.

However, even the most biologically inspired among synthetic models is far removed from an actual biological counterpart. They may omit important aspects of regulation that are indeed present in real, multilayered biological systems. They may also omit the natural biological noise as well as the experimental noise present in any measurement setting. In addition, since they have not been honed by evolution, they may lack important features of real networks, such as redundancy and canalization, or have other features that real networks lack. But the most important challenge is a cultural one. Experimental biologists, as it became obvious during the first DREAM meeting, are not ready to trust any reverse-engineering algorithm based on its performance on synthetic data. Thus, their use at this time is more targeted toward computational researchers that may hone their techniques using this kind of model.

Some of the richest *in silico* networks were described by Pedro Mendes,³ whose group has made artificial networks as realistic as possible. For example, Mendes believes that saturation of response is an important feature of biological networks, so it should be part of model networks as well.

An advantage of synthetic-network models is that they may allow the researchers to rapidly explore what data types may be more convenient for network inference: is it mRNA expression profiles, proteomics data, measurements of phosphorylation states? They also can be used to assess how sensitive their methods are to changes in topology, parameter values, kinetics, data availability, and noise.^{13–17}

In general it was thought that, despite their limitations, *in silico* networks have an important role to play in the assessment of reverse-engineering methods and will be included in the first round of DREAM challenges.

Experimental Biology

A major stimulus for the reverse-engineering community has been the availability of high-throughput data, for mRNA transcription as well as other types of data. At the simplest level of analysis, efforts in this direction looked for correlations in the mRNA levels of different genes as evidence of linkage between them. These techniques are very powerful and have driven the initial spurt of systems biology results. However, new integrative methods are proving to be better suited to reverse engineering.

For example, correlations are of limited use without knowing the *direction* of influence. Eric Schadt¹⁸ describes how genetic-sequence data help to solve this problem, even for the complex expression patterns of real tissues from animals or humans with disease. Because genetic variation influences, but is not influenced by, variations in expression, these data provide the same sort of directional cause-and-effect information. The DNA data help distinguish between “causal” expression changes and those that are merely reacting to the disease state. Interestingly, the DNA changes that predict disease almost all involve non-protein-coding regions. These methods go under the name of genetical genomics.

Another example of powerful biological tools used to probe signaling networks was discussed by Dana Pe’er, who, with her collaborators,^{8,19} approach the problem of reconstruction a signaling network using single-cell monitoring of phosphorylated proteins in T cells. Although the technique probes only a dozen species, she and her colleagues probed the patterns of variation in large numbers of individual cells using flow cytometry. They also used a host of small-molecule interventions that target specific parts of the network, allowing them to distinguish correlation from causation in the post-translational modifications that drive signaling networks.

In yet in another example of biological experimentation for reverse engineering, Ilya Nemenman⁴ is exploring networks that go beyond expression data. He describes plans for a robotically sampled chemostat to monitor hundreds of metabolites in small batches of cells. The metabolic networks he probes represent another important layer in the complex reality of biological systems.

Synthetic in Vivo Networks

Artificially designed synthetic networks, embedded into real biological systems, play an intermediate role between the control promised by *in silico* systems and the complexities of real biology. To combine the strengths of the two types of systems, the tools of bioengineering can be used to insert relatively simple “designer networks” in an organism, such as yeast, or cell lines. Unlike natural systems, such bioengineered systems would be almost perfectly known. (The “almost” is because remnant interactions between the synthetic network and the host network might create indirect unintended influences that affect the synthetic network.) Moreover, researchers who want to extend the results will not be limited to analyzing experimental data: they can be given the actual yeast strain and the corresponding perturbation plasmids. Results from such *in vivo* networks may be more likely to be accepted as relevant by the experimental biology community than purely *in silico* models. The criticism might be that this approach will not scale to larger networks easily.

Diego di Bernardo and his colleagues²⁰ integrated a synthetic regulatory network into the genome of budding yeast. This network is small, comprising only five added genes, but it provides a uniquely well-understood *in vivo* network for assessing reverse-engineering techniques according to di Bernardo. The researchers selected five interacting genes and introduced new promoter/gene combinations directly into the chromosomal DNA.

METRICS FOR ASSESSMENT

The evaluation of the performance of reverse-engineering methods depends to a great measure on the metrics used to evaluate them. Since there are many types of data, and many ways to organize these data into networks and models, coming up with metrics that are good for all methods is extremely hard. Comparison between reverse-engineering methods are subject to the same caveats.

The most basic comparison between a gold-standard network and its target is the topology: which nodes connect to which others? One popular test of a putative network is whether the inferred “edges” between nodes exist or do not exist in the gold-standard network. *Precision* indicates the percent of inferred interactions that exist in the gold-standard, while *Recall* indicates the percent

of interactions in the gold-standard network that were inferred by the method. Thus plotting the precision versus the recall performance of a method as a function of some of the method's parameters provides an excellent metrics for comparison.

In general, as emphasized by Mark Gerstein,²¹ making a test more stringent eliminates false positives (higher precision), but at the cost of an increasing number of false negatives (lower recall). Therefore metrics should be equally sensitive to both false positives and false negatives. Because biological networks tend to be sparse, there are typically many more negative pairs of nodes not joined by an edge than otherwise. This imbalance can influence the metrics for network reconstruction. In this sense PRC (precision–recall curve) plots are preferable to standard ROC (receiver operating curve) plots where recall is plotted against the false positive rate (i.e., the fraction of negative examples that are misclassified as positive), even when there is a close similarity between ROC and PRC plots.²² This is so because as the number of negative examples greatly exceeds the number of positive examples in typical biological networks, a large change in the number of false positives can lead to a small change in the false positive rate used in ROC analysis. On the other hand, precision captures the effect of the large number of negative examples on the algorithm's performance by comparing false positives to the number of inferred edges rather than to the much larger number of true negatives.

The PRC plot shows quantitatively how a reverse-engineering algorithm trades precision for recall: neither precision nor recall, by themselves, provides useful characterization. The typical summarization of this trade-off between precision and recall is achieved by estimating the area under the PRC, or AU-PRC. The closer to 1 this number is, the better the predictions of the reverse-engineering algorithm.

Pedro Mendes and collaborators^{3,17} have used the “confusion matrix” as a metric of performance. A confusion matrix organizes in a tabular format the information about actual and predicted classifications done by a classification system. When the classification system is binary (e.g., positives and negative cases), a confusion matrix contains the information to draw just one point in either the ROC or the PRC. As it is unclear what the best point is to compute the confusion matrix in the PRC, it is customary to choose 3 or 4 values of recall where to predict the precision. For example, a valid metric of comparison can be the precision at 10%, 20%, 50%, and 80% recall.

An alternative test, which recognizes the lack of a gold standard for real biological networks, is to compare observed behavior (which could be qualitative, such as over- and underexpression, or quantitative, such as response to perturbations, time courses, etc.) between the model and data collected from its target. The resulting metric will depend critically on what importance is ascribed to different features of agreement and disagreement.

In silico models of real biological processes are quite sensitive to the values of a handful of parameters. Most other parameters, however, can vary over wide

ranges without any obvious effect. It is important to recognize this limitation when assessing reverse-engineering algorithms, although it is often hard to know in advance which parameters will be poorly determined.

The precise form of the metric depends on the type of network model the methods generate. In a few cases, such as the gap-gene system described by Theodore Perkins,^{23,24} this model includes differential equations characterizing the interactions. Many of the reverse-engineering methods, however, only aim to predict a graph that summarizes which nodes are connected to which other nodes by an edge. For these models, the metric should quantify how well the method classifies edges as present or absent.

SPECIFIC ALGORITHMS

This symposium included some descriptions of specific algorithms to infer network structure. Many participants employed Bayesian network analysis, which goes beyond simple pair-wise correlations. By examining the conditional dependence and independence of a quantity in the context of other variables, it can pick up associations that are not obvious in the correlations because of confounding variables, including omitted nodes.

As these techniques typically start from a random seed, and each seed may lead to slightly different results, researchers replicate the same method hundreds of times starting with different seeds. Each individual replica is assigned a score according to how well the converged network explains the existing data. The final result is summarized in a consensus network, whose edges are those that appear in a large fraction of the resulting high-scoring networks. A pure Bayesian network approach reports statistical dependences, with directionality. Both Dana Pe'er⁸ and Eric Schadt¹⁸ showed that some types of data allowed them to deduce the direction of influence.

One shortcoming of Bayesian network analysis is that each one of the individually inferred networks has no feedback loops, in which the influence of one node on others eventually acts back on the original node. Thus, the final consensus network will likely miss feedback loops that are central to the behavior of many real networks. One way to deal with that weakness is to “unfold” the data in time. In this “dynamic” Bayesian network analysis, as described by Chris Wiggins,^{25,26} each node affects others only at a later time step, so loops at any time step are impossible. This requires time-series data, however, which is often hard to get at the appropriate time-resolution for complex, living organisms.

Regression-based algorithms are also a method of choice to reverse-engineer networks. Jim Collins, Tim Gardner, and collaborators have used this class of methods to identify the targets of antibiotics in *E. coli*.^{27,28}

HOW WELL CAN WE DO?

Going for the Gold Standard

Reverse engineers would like to test their methods against a “gold-standard” biological network whose detailed interactions are perfectly known (or at least known with high confidence) about both linkages and their absence. To serve as the most objective test, a gold standard would be perfectly unknown to those trying to reverse engineer it, although this “blinding” might deprive researchers of relevant biological context. There are many well-understood biological systems, but they are also widely publicized. Conference participants described many sources of extensively validated network information.

Gap Genes in Fruit Flies

Theodore Perkins^{23,24} described the gap gene system of *Drosophila melanogaster*, a very well-understood developmental network. The gap genes are active in the earliest stages of *Drosophila* development, and control the establishment of anterior-posterior structure. The developing spatial patterns of expression emerge as the gene products act as transcription factors for one another. Researchers have modeled early development using systems of coupled differential equations, which capture both local expression changes and spatial diffusion of proteins. By developing methods to measure evolving expression profiles for each gene during development, experimentalists create a huge number of conditions (concentrations varying in both space and time) to constrain the network model and equations. Of course, the results are widely known, and so could not be used as blinded tests of reverse-engineering methods.

Physical Interactions of Proteins

Joel Bader³⁵ asserted that a gold standard needs to have a measurable physical reality. Influence networks, such as Bayesian networks, he noted, are prone to uncertainties, depending on which nodes are included in the modeling. Even if you believe there is an interaction between two molecular species, you do not know what intermediates the interaction goes through. Bader proposed assessing methods based on the “physical reality of biochemical interactions between proteins and proteins and [between] proteins and DNA.” The protein–protein interaction data are available from high-throughput two-hybrid screens, while information about protein–DNA interactions comes from chromatin immunoprecipitation (ChIP).

Established Pathways in Cancer

Arnold Levine^{11,12} described the signal transduction pathway associated with the p53 tumor suppressor gene. The interactions have been painstakingly validated, and researchers have clarified the detailed biological mechanisms along much of the pathway. A similarly detailed pathway was summarized by Riccardo Dalla-Favera³⁰ in the context of B cells. These well-established and robust pathways bring both the confidence and the interest of a larger biological community.

Genetic Expression in E. coli

Boris Hayete^{32,33} used the RegulonDB network of genetic expression in *E. coli* for comparing different reverse-engineering algorithms. This database has modest coverage, however, so finding an edge that it doesn't contain is not necessarily a false positive.

The issue of blinding the data remains contentious, since experimentalists have little reward for withholding their hard-won results for the benefit of reverse engineers. Withholding a fraction of high-throughput data is more promising. Some researchers, however, questioned whether removing data from their biological context even makes sense. The second DREAM conference will contain blinded data removed from biological context, as indicated in FIGURE 2. Therefore the usefulness or lack thereof of these kinds of exercises will be demonstrated.

Can We Ever Get Enough Data?

Even the best data may not differentiate between different network models. Thomas Maiwald³⁶ built a software framework to test numerically how accumulating experiments discriminate between proposed models. He found that some kinds of data are never adequate to let researchers distinguish between some models.

Winfried Just^{37,38} took a mathematical approach, developing theorems for a system in which concentrations assume discrete values and evolve in discrete time steps, but he said that many of his conclusions should apply to other types of dynamic systems as well. Given a set of observations of the concentrations, the task is to find a rule, relating the values to those at the preceding time step, that reiterates the observations. The problem of generalizing to a rule from a finite number of observations is well known. For this type of problem, Just's "no free lunch" theorem states that a huge number of rules will match a data set equally well, unless there are prior constraints on the structure of the rule or the nature of the data. At first blush, this theorem appears to doom the entire reverse-engineering endeavor to failure. In practice, however, biologically

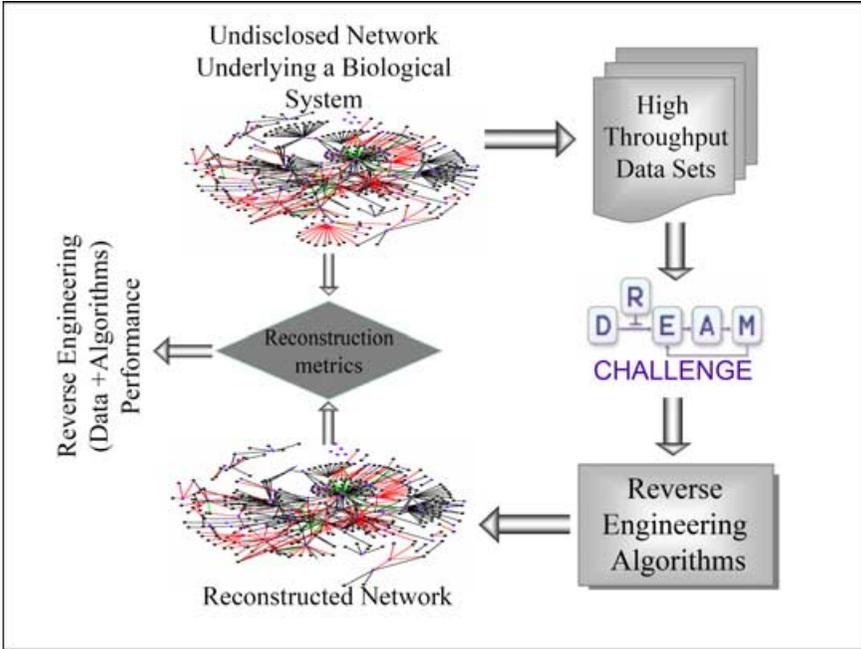


FIGURE 2. The algorithm performance assessment method proposed in the context of the DREAM project. A biological system is used to produce a high-throughput data set. The data set is distributed through the DREAM project web site to the participants of the DREAM challenge, without disclosing the identity of the network underlying the biological system. The predictors are invited to infer the network from which the data originated. The actual networks will be pitted against the reconstructed networks through a number of network comparison metrics to determine the quality of the network reconstruction algorithm and of the appropriateness of the data to reconstruct the underlying network.

reasonable constraints and experimental protocols provide important “priors” that may limit the number of choices.

BIOLOGICAL RELEVANCE

High-throughput data collection is blind to important aspects of biology, including cellular geometry and compartmentalization as well as large protein complexes. A network-centric view of biology is bound to have serious flaws if too much biological context is removed in lieu of simple models. Some speakers were invited to remind the participants of the larger biological picture.

Leslie Loew coordinates the Virtual Cell project, which aims to simulate realistic spatial features of biology along with the reaction kinetics. In his presentation,³⁹ Loew described the growth of actin filaments which depends on much more than mass-action kinetics typical of bulk chemistry. He also

described the long-term depression of nerve signals in Purkinje cells, which occurs when two types of fibers fire simultaneously. An accurate model of this process requires specification of the geometry of the dendritic spines.

In another important talk, Michael Rout reminded participants that many of the proteins in the cell interact by forming complexes with precise spatial relationships. In particular, the nuclear pore complex, which transports macromolecules between the nucleus and the cytoplasm, includes at least thirty different proteins.⁴⁰ To understand how they fit together Rout's team breaks cells apart and rapidly freezes them. This method identifies interacting proteins that a more generic technique might well miss.

Throughout the molecular-biology era, biologists have been steadily uncovering the molecular events and interactions that govern diverse phenomena. Traditionally, they depict these interactions as chains or "pathways," although such a linear picture often oversimplifies the true networks. Arnold Levine surveyed the traditional^{11,12} techniques for elucidating pathways. In particular, he focused on signal transduction pathways, especially the one involving the p53 cancer-suppressor gene. Biologists have now clarified dozens of such pathways, in which post-translational modifications play a critical role.

Riccardo Dalla-Favera³⁰ describes the power of reverse engineering in clarifying the networks of B cells. Unlike traditional gene expression profiling, the network analysis gives researchers like him a clear indication of promising potential targets.

WHAT IS TO BE DONE?

In a final discussion session of the First DREAM conference, and in a follow-up questionnaire,⁴¹ the participants in the first DREAM conference expressed their opinions about how to move the process forward. Although they expressed enthusiasm for the goals of evaluating the reverse-engineering enterprise, they held different views on how—and even whether—to create any formal process.

CASP-Like Competition?

One possible role model is the biannual CASP (Critical Assessment of Protein Structure Prediction) challenge for evaluating protein structure predictions.⁴² In CASP, once proteins' structures are solved experimentally, competitors make predictions based on the sequence data, over a prescribed period. These data are "blinded," in the sense that no structural information is made available. The experimental structure is then revealed and the deviations of the various predictions are computed (FIG. 2).

Participants noted many objections to applying this model to reverse engineering:

- There are many relevant types of input data, and there is no consistent format.
- Data on networks often emerges in stages, not all at once.
- Different underlying network models will produce different types of detail in their predictions.
- Unlike protein structure, there is no “right answer”; as yet there is no gold-standard experiment on which to compare predictions.
- The only precisely understood data sources are those with questionable biological relevance.
- Experimenters may be reluctant to sequester their hard-won results while the competition proceeds.
- Studying “blind” data, with restricted scope and removed from its biological context, runs the danger of becoming an abstract exercise with no relevance to the larger community, as well as degrading the quality of predictions.

An incremental approach would start with a formal, blinded assessment process based on limited data from artificial or synthetic networks. This assessment may have little importance for biologists at first. However, internal quality control within the reverse engineering community could develop the procedures, and buy credibility among experimenters to address problems that are more relevant.

A Repository

DREAM could provide a repository for two types of data: model predictions for validation by experimenters, and experimental data for evaluation by modelers. Establishing and enforcing standards of statistical rigor could improve the quality of papers.

Many attendees felt that DREAM could provide a clearinghouse for ongoing reverse-engineering activities. One idea that found a good following suggested maintaining a list of predictions made by various modeling papers, which could be a resource for experimentalists. There could also be a complementary list of biological networks under active investigation.

DREAM Database

The DREAM project will maintain a database that will serve as a clearinghouse for three important classes of data. One is a “database of high-probability predicted interactions” in search of validation by experimental biologists. Crisp predictions about the results of removing specific interactions in a real biological context should also get the attention of experimental biologists. The second

is “a database of high-value cellular circuitry data that requires concerted attention by the reverse-engineering community.” The third one is a database of predictions of observable changes in cell behavior resulting from specific cellular perturbations. These resources along with the recurring workshops will help unite computational and experimental biologists as a community around the reverse-engineering goal.

Challenge for the DREAM Database

The diversity of data sources and their uneven (albeit rapidly improving) quality pose a challenge for DREAM in providing a repository for data useful for the task of inference of biological networks. A mechanism must be created to determine which data sets will be initially be made available to the community. The availability of predictions from a consistent set of data will make it somewhat easier to compare results in the absence of a perfect gold standard. What procedures can best ensure the quality of data in the DREAM database? This question needs further exploration.

CONCLUSIONS

At the time of this writing, we are the midst of the organization of DREAM2, the Second DREAM conference, where we create what seems to be a suitable set of gold standards to explore the strengths and limitations of different algorithms through network inference challenges. The challenges are:

- the identification of the actual targets of a transcription factor in a mammalian cell, in a set containing both actual targets and decoys;
- the inference of a protein–protein interaction sub-network from the list of yeast genes;
- the network of interaction in a synthetic biology circuit transfected to an *in vivo* organism, from both gene expression and qPCR data;
- the reverse engineering of three *in silico*–generated networks, with qualitatively different topologies from *in silico*–generated gene expression data; and
- the gene regulatory network of a model organism from high-throughput gene expression data.

The participants in this exercise will be tasked to use reverse-engineering algorithms to infer the connectivity of the networks underlying the curated data sets. We will adopt a double-blind competitive assessment style in which the participants will not know the structure of the network they are trying to predict, and the evaluators will not know who made the predictions. We expect that these efforts will enhance our understanding of the merits and limitations of different methods. We also hope that these challenges will help us to advance by a notch our understanding and thus organization of the current

bewildering plethora of reverse-engineering methods as well as test the limits of the whole conception of reverse engineering from integrated data sets of cellular networks.

As we continue to collect information and consolidate the knowledge emerging from these networks, it might be possible to extract from them some clues to the basic design principles of these circuits. Some hints that there might be some general design principles have emerged from the analysis of biological networks using the tools of graph theory. For example, it has been observed that many biological networks have the scale-free property,⁴³ meaning that a few cellular components (nodes in the graph) are highly connected to other cellular components, much more so than most of the other nodes in the graph. These highly connected nodes tend to be essential for cellular function. Another level of network analysis has identified local network properties, such as topological motifs. A topological motif is a recurring sub-structure in the network that may represent a reused circuit pattern within the network appearing at a significantly higher rate than would be expected by chance.⁴⁴

The network of interactions provides us with a global view of the connectivity of cellular components and hints of generic design principles. These networks are useful to make general inferences, and generate some qualitative hypotheses. Ultimately, however, quantitative models with some degree of predictive power and mechanistic detail best demonstrate that a complex system is understood well enough to resolve how subtle differences in the cellular machinery may result in substantially different behavior. Reaching such understanding is arguably the ultimate goal of systems biology.

ACKNOWLEDGMENTS

We thank the support of the NIH Roadmap Initiative, the Columbia University Center for Multiscale Analysis Genomic and Cellular Networks (MAG-Net), the IBM Computational Biology Center, and The New York Academy of Sciences for their financial and logistic support of the DREAM project.

REFERENCES

1. SAURO, H.M., M. HUCKA, A. FINNEY, *et al.* 2003. Next generation simulation tools: the Systems Biology Workbench and BioSPICE integration. *OMICS* **7**: 355–372.
2. SHANNON, P., A. MARKIEL, O. OZIER, *et al.* 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**: 2498–2504.
3. MENDES, P. 2006. Web lecture delivered at the First DREAM Conference. www.nyas.org/ebriefreps/ebrief/000596/presentations/mendes/player.html.

4. NEMENMAN, I., G.S. ESCOLA, W.S. HLAVACEK, P.J. Unkefer, C.J. Unkefer & M.E. Wall. 2007. Reconstruction of metabolic networks from high throughput metabolic profiling data: *in silico* analysis of red blood cell metabolism. *Ann. N. Y. Acad. Sci.* **1115**: 102–115. This volume.
5. BERNARD, A. 2006. Web lecture delivered at the First DREAM Conference. www.nyas.org/ebriefreps/ebrief/000596/presentations/bernard/player.html.
6. SAMOILOV, M. 2006. Web lecture delivered at the First DREAM Conference. www.nyas.org/ebriefreps/ebrief/000596/presentations/samoilov/player.html.
7. MARGOLIN, A. & A. CALIFANO. 2007. Theory and limitations of genetic network inference from microarray data. *Ann. N. Y. Acad. Sci.* **1115**: 51–72. This volume.
8. PE'ER, D. 2006. Web lecture delivered at the First DREAM Conference. www.nyas.org/ebriefreps/ebrief/000596/presentations/peer/player.html.
9. KUNDAJE, A., S. LIANOGLU, X. LI, *et al.* 2007. Learning regulatory programs that accurately predict differential expression with MEDUSA. *Ann. N. Y. Acad. Sci.* 178–202. This volume.
10. LAMB, J., E.D. CRAWFORD, D. PECK, *et al.* 2006. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* **313**: 1929–1935.
11. LEVINE, A.J. 2006. Web lecture delivered at the First DREAM Conference. www.nyas.org/ebriefreps/ebrief/000596/presentations/levine/player.html.
12. LEVINE, A.J., W. HU, Z. FENG & G. GIL. 2007. Reconstructing signal transduction pathways: challenges and opportunities. *Ann. N. Y. Acad. Sci.* **1115**: 32–50. This volume.
13. MENDES, P., W. SHA & K. YE. 2003. Artificial gene networks for objective comparison of analysis algorithms. *Bioinformatics (Suppl 2)*: ii122–129.
14. RICE, J.J., Y. TU & G. STOLOVITZKY. 2005. Reconstructing biological networks using conditional correlation analysis. *Bioinformatics* **21**: 765–773.
15. MARGOLIN, A.A., I. NEMENMAN, K. BASSO, *et al.* 2006. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* **7**(Suppl. 1): S7.
16. BANSAL, M., V. BELCASTRO, A. AMBESI-IMPIOMBATO & D. DI BERNARDO. 2007. How to infer gene networks from expression profiles. *Mol. Syst. Biol.* **3**: 78.
17. CAMACHO, D., P. VERA LICONA, P. MENDES & R. LAUBENBACHER. 2007. Comparison of reverse-engineering networks using an *in silico* network. *Ann. N. Y. Acad. Sci.* **1115**: 73–89. This volume.
18. SCHADT, E. 2006. Web lecture delivered at the First DREAM Conference. <http://www.nyas.org/ebriefreps/ebrief/000596/presentations/schadt/player.html>.
19. SACHS, K., O. PEREZ, D. PE'ER, *et al.* 2005. Causal protein-signaling networks derived from multiparameter single-cell data. *Science* **308**: 504–506.
20. DI BERNARDO, D. 2006. Web lecture delivered at the First DREAM Conference. <http://www.nyas.org/ebriefreps/ebrief/000596/presentations/bernardo/player.html>.
21. GERSTEIN, M. 2006. Web lecture delivered at the First DREAM Conference. www.nyas.org/ebriefreps/ebrief/000596/presentations/gerstein/player.html.
22. DAVIS, J. & GOADRICH, M. 2006. The relationship between precision-recall and ROC curves. *Proceedings of the Twenty-Third International Conference on Machine Learning (ICML'06)*. June 25–29, 2006, Pittsburgh, PA. ACM International Conference Proceedings Series; Vol. 148, pp. 233–240 (2006).
23. PERKINS, T.J. 2006. Web lecture delivered at the First DREAM Conference. www.nyas.org/ebriefreps/ebrief/000596/presentations/perkins/player.html.

24. PERKINS, T.J. 2007. The gap gene system of *Drosophila melanogaster*: model-fitting and validation. *Ann. N. Y. Acad. Sci.* **1115**: 116–131. This volume.
25. WIGGINS, C.H. 2006. Web lecture delivered at the First DREAM Conference. <http://www.nyas.org/ebriefreps/ebrief/000596/presentations/wiggins/player.html>.
26. DAVID, L.A. & C.H. WIGGINS. 2007. Benchmarking of dynamic Bayesian networks inferred from stochastic time-series data. *Ann. N. Y. Acad. Sci.* **1115**: 90–101. This volume.
27. GARDNER, T.S., D. DI BERNARDO, D. LORENZ & J.J. COLLINS. 2003. Inferring genetic networks and identifying compound mode of action via expression profiling. *Science* **301**: 102–105.
28. DI BERNARDO, D., M.J. THOMPSON, T.S. GARDNER, *et al.* 2005. Chemogenomic profiling on a genome-wide scale using reverse-engineered gene networks. *Nat. Biotechnol.* **23**: 377–383.
29. BASSO, K., A.A. MARGOLIN, G. STOLOVITZKY, *et al.* 2005. Reverse engineering of regulatory networks in human B cells. *Nat. Genet.* **37**: 382–390.
30. DALLA-FAVERA, R. 2006. Web lecture delivered at the First DREAM Conference. www.nyas.org/ebriefreps/ebrief/000596/presentations/favera/player.html.
31. NEMENMAN, I. 2006. Web lecture delivered at the First DREAM Conference. <http://www.nyas.org/ebriefreps/ebrief/000596/presentations/nemenman/player.html>.
32. HAYETE, B. 2006. Web lecture delivered at the First DREAM Conference. www.nyas.org/ebriefreps/ebrief/000596/presentations/hayete/player.html.
33. FAITH, J.J., B. HAYETE, J.T. THADEN, *et al.* 2007. Large-scale mapping and validation of *Escherichia coli* transcriptional regulation from a compendium of expression profiles. *PLoS Biol.* **5**: e8.
34. STIGLER, B. 2006. Web lecture delivered at the First DREAM Conference. www.nyas.org/ebriefreps/ebrief/000596/presentations/stigler/player.html.
35. BADER, J. 2006. Web lecture delivered at the First DREAM Conference. www.nyas.org/ebriefreps/ebrief/000596/presentations/bader/player.html.
36. MAIWALD, T., C. KREUTZ, A.C. PFEIFER, *et al.* 2007. Dynamic pathway modeling: Feasibility analysis and optimal experimental design. *Ann. N. Y. Acad. Sci.* **1115**: 212–220. This volume.
37. JUST, W. 2006. Web lecture delivered at the First DREAM Conference. www.nyas.org/ebriefreps/ebrief/000596/presentations/just/player.html.
38. JUST, W. 2007. Data requirements of reverse-engineering algorithms. *Ann. N. Y. Acad. Sci.* **1115**: 142–153. This volume.
39. LOEW, L. 2006. Web lecture delivered at the First DREAM Conference. www.nyas.org/ebriefreps/ebrief/000596/presentations/loew/player.html.
40. ROUT, M.P., J.D. AITCHISON, A. SUPRAPTO, *et al.* 2000. The yeast nuclear pore complex: composition, architecture, and transport mechanism. *J. Cell Biol.* **148**: 635–651.
41. [http://www.nyas.org/ebriefreps/ebrief/000596/presentations/DREAM_](http://www.nyas.org/ebriefreps/ebrief/000596/presentations/DREAM_Questionnaire_Responses.ppt)
Questionnaire Responses.ppt.
42. MOULT, J. 2006. Lessons from a decade of Critical Assessments of Structure Predictions. <http://www.nyas.org/ebriefreps/ebrief/000534/presentations/moult/player.html>.
43. JEONG, H., B. TOMBOR, R. ALBERT, *et al.* 2000. The large-scale organization of metabolic networks. *Nature* **407**: 651–654.
44. SHEN-ORR, S.S., R. MILO, S. MANGAN & U. ALON. 2002. Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat. Genet.* **31**: 64–68.