ELSEVIER

A biological approach to computational models of proteomic networks

Kevin A Janes and Douglas A Lauffenburger

Computational modeling is useful as a means to assemble and test what we know about proteins and networks. Models can help address key questions about the measurement, definition and function of proteomic networks. Here, we place these biological questions at the forefront in reviewing the computational strategies that are available to analyze proteomic networks. Recent examples illustrate how models can extract more information from proteomic data, test possible interactions between network proteins and link networks to cellular behavior. No single model can achieve all these goals, however, which is why it is critical to prioritize biological questions before specifying a particular modeling approach.

Addresses
Biological Engineering Division and Cell Decision Processes Center, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

Corresponding author: Janes, Kevin A (kjanes@mit.edu)

Current Opinion in Chemical Biology 2006, **10**:73–80

This review comes from a themed issue on
Proteomics and genomics
Edited by Garry P Nolan and Emanuel F Petricoin

Available online 6th January 2006

1367-5931/$ – see front matter

DOI 10.1016/j.cbpa.2005.12.016

Introduction

Our current understanding of the proteins, interactions and pathways that comprise signaling networks is detailed, yet it remains incomplete. Recent experimental techniques for unraveling intricate signaling networks have become increasingly quantitative and multiplex. New approaches are now needed to compile the existing quantitative biological knowledge and to maximize the information extracted from large-scale signaling and proteomic datasets. Computational models formalize a complex biological or experimental process mathematically, which can be useful for assembling and analyzing quantitative data. Modeling is thus critical for fields such as proteomics, genomics and systems biology.

As a discipline, biology thrives on clarity through consensus (take, for instance, the central dogma). To model biological networks, however, we and others have argued against a consensus 'one size fits all' philosophy, favoring instead a spectrum of computational techniques [1,2]. Admittedly, the full breadth of computational modeling approaches can be daunting [3], and all techniques are not equally valid for all questions. If the choice of model is flexible but not arbitrary, then which modeling approaches are appropriate for which biological applications? Here, we attempt to answer this question through examples of recently published proteomic-network models. Rather than organizing the review around a sequence of modeling techniques, we focus on the important biological problems to which different network models have contributed. This biology-centric approach might clarify how proteomic-network models can be versatile but not all-encompassing.

Anchoring model sophistication with experimental data

Proteomics research is clearly directed at uncovering more biological detail within networks — new proteins, new interactions, new complexes [4]. For computational models of such networks, however, the level of detail must be constrained by the availability of supporting quantitative data. Models with very little detail are nicely constrained but oversimplified, whereas highly sophisticated models lack the experiments needed to specify the detailed mechanisms realistically. Model 'believability' is therefore highest when the model contains only the detail needed to capture and predict the experiments of interest for a biological question.

How does increasing the level of model detail decrease believability? With model detail come parameters. In a model, these parameters might define a signaling protein's starting concentration, rate of turnover or diffusivity through the cytoplasm. Model parameters are frequently unknown and must therefore be estimated from data, which reduces believability. Importantly, the number of required parameters multiplies as more biological detail is added (Table 1). Spatial detail, for example, requires substantial parameterization, limiting the scope of believable models to only a few proteins. This does not imply that proteomic-network models should never include details of protein localization. Clearly, spatial detail has been essential for answering certain biological questions, such as those involving nucleocytoplasmic shuttling [5•–7•]. Model sophistication ultimately evolves via a series of choices that are based on the particular problem at hand (Figure 1). An advantage of analyzing proteomic networks computationally is that these choices (and all of their underlying assumptions) are hard-coded into the mathematics of the model.

Table 1

Parameter considerations for proteomic network models.

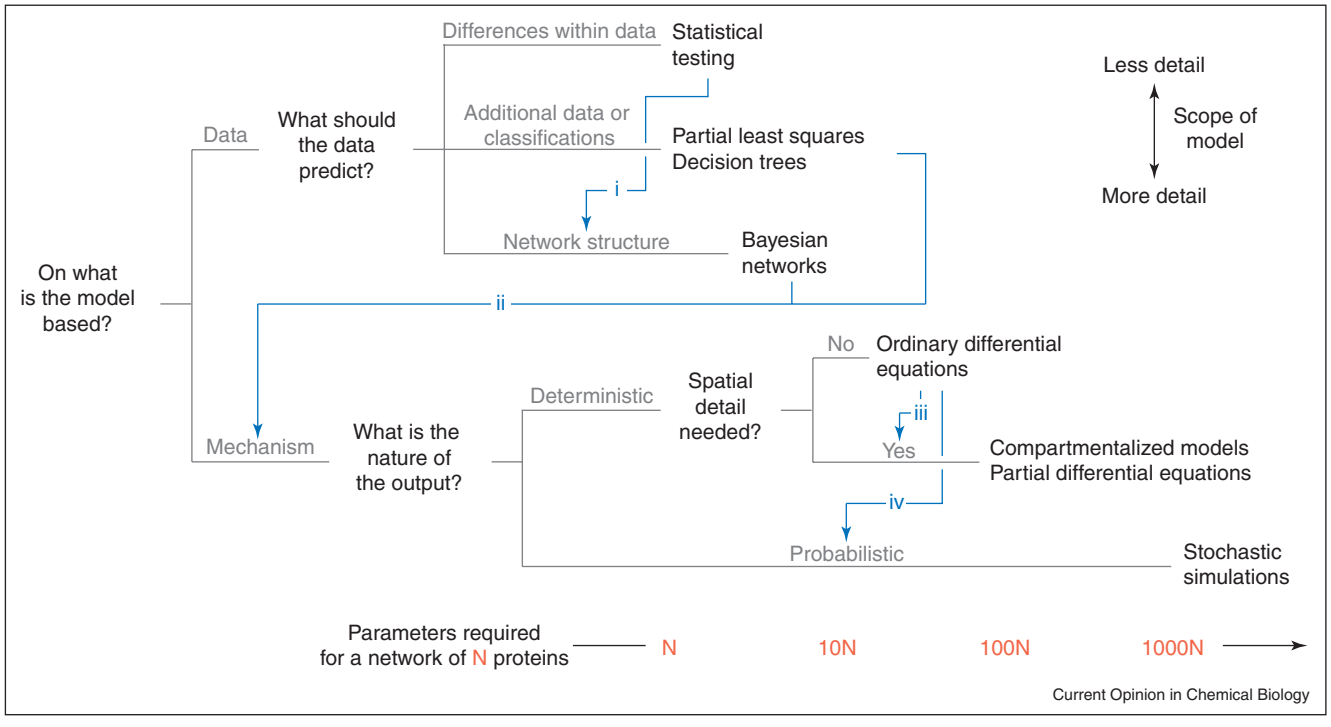
Included in model	Parameters required	Comments and assumptions
Proteins	N	Protein concentrations for a network of N proteins.
Interactions	+ 5N	Average number of interactions per protein [51].
Reaction kinetics	+ 2.5 × 5N	Reaction-rate parameters for each connection, assuming equal proportion of binding (two-parameter) and enzyme-catalyzed (three-parameter) reactions.
Spatial detail	+ (C – 1) × N	Additional protein concentrations for each of C topologically distinct compartments (organelles, surfaces, etc.); C can be as large as 20 [52].
	+ N	Diffusion coefficients for each of N proteins, assuming equal diffusivity in different compartments.
	+ 2 × C × N	Boundary conditions needed to solve two-dimensional partial differential equations in each of C compartments.
Stochasticity	× M × N	For an exact stochastic solver simulating the dynamics of M individual molecules [42]; stochasticity thought to become important when M is less than 100 per cell.

Organizing computational models from a biological perspective

Different types of network models can be separated into mathematical classes: deterministic, probabilistic and statistical (Figure 2a, Box 1). Although useful for organizing the computational algorithms, this model-centric view hides the interesting biological applications behind equations and nomenclature. Unanswered biological questions are the critical motivation for proteomic-network modeling, but their importance is de-emphasized when computation is placed at the forefront. A model-centric view creates models in search of questions.

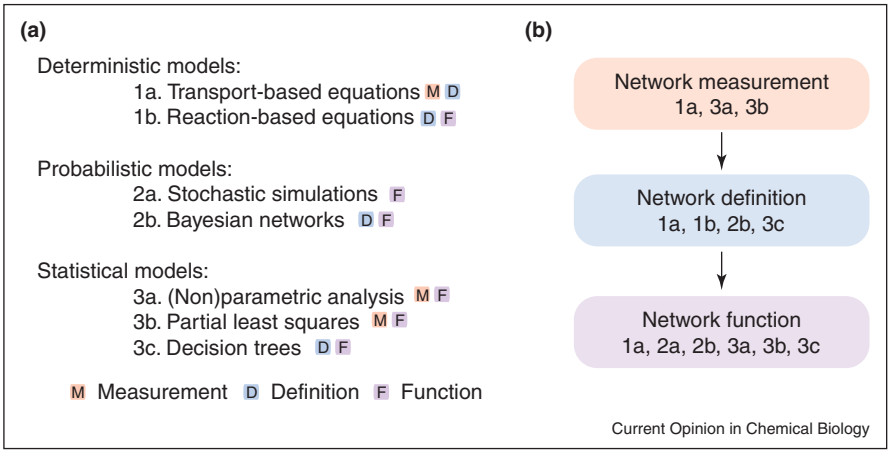
Moving biology ahead of algorithms creates a different classification of proteomic-network models (Figure 2b). First, there arise questions of network measurement. How can proteomic measurements be made more reliable, quantitative and reproducible? Are some measurements more informative than others? Second, there are questions of network definition. Is our current under-

Figure 1



Decision tree for selecting different modeling approaches. Models with differing levels of detail are organized vertically, and parameter requirements are arranged horizontally on a logarithmic scale according to the estimates in Table 1. Blue arrows illustrate potential interconnections between models: (i) statistical tests leading to new insights into network structure; (ii) data-driven models that reveal new network mechanisms; and ordinary differential equations that test the need for (iii) spatial detail or (iv) stochasticity. See Box 1 for a glossary of terms.

Figure 2



Two distinct perspectives on computational models of proteomic networks. **(a)** A model-centric view. Overall categories are organized by model type and applied to the measurement ('M'), definition ('D') or function ('F') of proteomic networks. Note that no single type of model is appropriate for all three areas. **(b)** A problem-centric view. Overall categories are organized by biological question. Numbers and letters refer to the approaches listed in (a). See Box 1 for a glossary of terms.

standing of a proteomic network sufficient to explain counterintuitive results? Can we deduce new interactions from large-scale measurements alone? Finally, there are questions of network function. How are phenotypes coordinated by a proteomic network? Where should we perturb networks for therapy? For each of these questions, there exists a panel of computational approaches from which to choose (Figure 2b). Model selection within each panel can then be based on the specifics of the biological system, its current extent of understanding and the data available (Figure 1).

Box 1 Glossary.

Bayesian network: A graphical modelling technique for calculating the most likely set of protein–protein interactions given a set of experimental measurements of these proteins.
Compartmentalized model: A model based on ordinary differential equations that incorporates spatial detail by considering proteins in different organelles as distinct measured variables; for example, $x_{cytoplasm}$ and $x_{nucleus}$.
Decision tree: A learning algorithm that approximates a cellular output by constructing a 'tree', where the 'branches' classify experiments based on the levels of measured variables and the 'leaves' at the end of the branches predict the output.
Deterministic model: A model that will give the same output each time when provided the same starting conditions.
Ordinary differential equation: An equation containing derivatives with respect to a single experimental variable, such as time.
Partial differential equation: An equation containing derivatives with respect to multiple experimental variables, such as both time and space.
Partial least squares: A regression technique for relating independent and dependent measured variables when the number of variables exceeds the number of experimental observations.
Pearson correlation: A statistical term that quantifies the degree to which two measured variables are linearly proportional to one another.
Probabilistic model: A model that can give different outputs from the same starting conditions, because outputs are based on the probability of certain outcomes occurring.
Stochasticity: Randomness, used here to refer to the heterogeneity of outcomes that can occur as a result of reactions involving small numbers of proteins.

Network-measurement models

Modeling has played an increasingly important role in the measurement of proteomes and networks. Measurement models are a useful way to condense different methodological considerations into a single quantitative description of an experiment (Table 2). High-throughput and multiplex techniques can be particularly sensitive to methodological details if the goal is to fuse results from many samples and many experiments into a common dataset. Models are valuable to convert raw measurement data into useful information or to quantify confidence within large-scale datasets.

In the field of 'global' proteomics, no other experimental method has had as significant an impact as mass spectrometry (MS) [8]. MS-based proteomic datasets still require significant technical expertise to decipher the complexities of upstream fractionation, MS instrumentation and downstream peptide (or protein) identification that are involved in each experiment [9]. However, emerging from this expertise are empirical statistical models, where anecdotal information and earlier experiments are converted mathematically into prior knowledge about the current dataset. MS-relevant examples of prior knowledge include using (i) peptide size and hydrophobicity to model retention times on separating columns [10], (ii) database scoring parameters to discriminate correct and incorrect peptide assignments from MS spectra [11] and (iii) sibling and degenerate peptides to predict the pre-

Table 2

Classes and subtypes of proteomic network models.

Model class	Model subtype	Goal or purpose	Examples
Network measurement	Methodology	Improve the accuracy, sensitivity or stability of a proteomic network measurement	HPLC modeling [10] Peptide identification [11] Protein identification [12•]
	Information	Assess the quantitative or relative value of a proteomic network measurement	Proteomic compendia [15••] Interactome compendia [14•]
Network definition	Reconstruction	Define a consensus mechanism summarizing the existing knowledge of a proteomic network	Growth factors: EGF [21], NGF [22], PDGF [23] Cytokines: FasL [24,25], TNF [26] Morphogens: Wnt [27], Shh [28] Signaling modules: JAK-STAT [6,44], NF-κB [45], TCR [46]
	Inference	Deduce protein-network structure from a series of experimental measurements or perturbations	T cells [30••] Stem cells [47] Endothelial cells [31••,48]
Network function	Response	Predict graded cellular outputs from network signals	Migration [38•] Cytokine secretion [35] Contractility [37] Size [49]
	Decision	Predict cellular phenotypes from network signals	Apoptosis [36,39••] Differentiation [32] Proliferation [50] Viral infectivity [41]

EGF, epidermal growth factor; HPLC, high-performance liquid chromatography; JAK, Janus kinase; NF-κB, nuclear factor-κB; NGF, nerve growth factor; PDGF, platelet-derived growth factor; Shh, Sonic hedgehog; STAT, signal transducer and activator of transcription; TCR, T-cell receptor; TNF, tumor necrosis factor; Wnt, Wingless/Int-1.

sence of proteins in a complex starting mixture [12•]. Importantly, prior knowledge can be updated during iterations between model and experiment, such that the first model output from the data becomes part of the prior knowledge during a second pass through the model. An excellent demonstration of iterative modeling for MS-based measurements is the work by Keller *et al.* [11], which used prior MS measurements as training data to fit an initial frequency distribution of peptides whose assignments were correct and incorrect. The initial distributions were used as prior information in a model that calculates the probability of correct peptide assignment given an MS spectrum. Running the model through all of the spectra in an MS experiment generates a new distribution of (probably) correct and (probably) incorrect peptides, which can update the prior information for the next iteration through the model. In this way, the model learns the most likely peptide assignments from the spectra itself, with initialization provided by a high-quality training dataset. The resulting peptide probabilities can then be fed into downstream models that calculate protein assignments from a set of likely peptides [12•].

Measurement models are also useful for analyzing data quality itself. Often, quality is synonymous with information [13], and recently modeling has been used to identify the information-rich subset of measurements within proteomic datasets (Table 2). Gunsalus *et al.* [14•] selected for high-quality proteomic data by calculating the intersection of large-scale phenotypic, transcriptional and interaction datasets in *Caenorhabditis elegans*. Using the overlap among the measured networks, the Gunsalus *et al.* model was shown to be enriched in proteins sharing common biological functions. Gaudet *et al.* [15••] used the predictive ability of a model to quantify network information content directly from a proteomic measurement set. A key conclusion from this work was the importance of measurement combinations. Different types of assays (kinase activity assays, quantitative western blotting, etc.) used over a range of time points were critical to accurately predicting the response of cells treated with multiple experimental stimuli. As quantitative MS-based experiments evolve [16,17], it will be important to use similar strategies to quantify how information distributes among different precursor-ion spectra and to examine how the information from MS experiments compares with other large-scale measurement techniques [18,19].

Network-definition models

An important goal for computational models is to define mathematically the proteins and pathways that constitute a signaling network. Modeling strategies for addressing the question of network definition can be subdivided into two categories: reconstruction models, which build networks from previously reported mechanisms; and inference models, which deduce network structure from large-scale datasets (Table 2). Network reconstruction

[20•] has a long history dating back to metabolic engineering and has been widely applied to proteomic signaling networks involving growth factors [21,22•,23], cytokines [24–26] and morphogens [27,28].

What can be learned from these complex models founded on highly parameterized systems of differential equations (Figure 1)? Ironically, reconstruction models tend to reveal more about the underlying network when they cannot capture the observed biology. A model by Hua *et al.* [25] of the prodeath cytokine, FasL, was at first unable to predict correctly how the activation of the apoptotic protease caspase-3 would be affected by overexpression of the antiapoptotic protein Bcl-2. The initial model had specified that overexpressed Bcl-2 interacted exclusively with the prodeath molecule Bax. Using the model, alternative Bcl-2 interaction scenarios were investigated. The *in silico* network that best matched experiment specified for Bcl-2 to interact with both Bax and another prodeath Bcl-2 family member, truncated Bid. In this way, reconstruction models can evolve with the emergence of new data and biological mechanisms. A revised nerve growth factor (NGF) model by Sasagawa *et al.* [22•], for example, implicated Rap1 as the critical small G protein that mediates sustained extracellular-regulated kinase activation after NGF stimulation. The perpetual refinement of this subtype of models suggests that open distribution and sharing of network reconstructions will be important in the future [29] (see also Update).

Many biological networks lack the in-depth mechanistic understanding needed for a plausible network reconstruction. With these networks, inferential modeling approaches can be used to suggest connections between molecules (Table 2). Unlike reconstruction models, inference models rely heavily on analysis of a core dataset (upper half of Figure 1). Different mathematical algorithms can help reveal networks from data, but they are all fundamentally based on identifying covariations among measurements. A Bayesian network approach [30••] might identify different relationships than would a Pearson correlation analysis [31••] (Box 1), but it is the design of the original experiments that determines whether either computational approach will identify new mechanistic links in the network. Recent evidence suggests two useful strategies to design experiments for network definition: stimulate cells with diverse combinations of extracellular stimuli [15••,32] to reduce the likelihood of chance correlations; and disrupt many individual proteins genetically or pharmacologically [30••,31••] to focus on points of control within the network. A probabilistic model by Sachs *et al.* [30••] inferred the connectivity of 11 signaling molecules from a single set of experiments involving primary T cells treated with combinations of pathway activators and inhibitors. By using a dataset based on interventions, the model recapitulated the mechanisms established from decades of signaling biochemistry and also suggested new links within the network. Another nice example is the approach of Plavec *et al.* [31••], which used protein overexpression combined with various cytokines to deduce statistically the interconnections among proteins in the Ras and NF-κB pathways. An increase in multiparameter, systems thinking about signaling networks [33•] suggests that many experimental datasets will soon be available for inference modeling. In the future, more complete and reliable network definitions will probably be achieved by applying multiple computational techniques (Figure 2b) to the same network and dataset of interest.

Network-function models

Proteomic networks are important because they ultimately control cellular functions. Diverse extracellular stimuli converge upon a common intracellular network, which can mediate an array of cellular responses [34•]. Function models aim to address questions of how proteomic networks control graded cellular responses, such as the amount of a secreted cytokine [35], or binary cellular decisions, such as death versus survival [36] (Table 2). Some functions, like the contraction of a cardiomyocyte [37], are extremely well characterized and can be approached with a highly detailed mechanistic model (lower half of Figure 1). Most function models, though, are correlative and seek to link a measured intracellular network with measured cellular phenotypes. Hautaniemi *et al.* [38•] used decision-tree modeling (Box 1) to characterize cell migration based on the phosphorylation levels of five key intracellular proteins. The resulting 'branches' of the decision tree identified the sequence of conditional molecular statements that best predicted low, medium or high cellular speed — for instance, IF extracellular-regulated kinase phosphorylation is low AND IF myosin light-chain phosphorylation is high THEN migration speed is high. It would be interesting to use this approach in larger networks while constraining decision-tree branchpoints based on the approximate positions of molecules in the network (first membrane transducers, then initiators, then effectors, etc.).

Prediction of cellular functions can also be achieved more quantitatively by training models on measurements of the upstream signaling network. We have used partial least squares modeling to predict 12 measured apoptotic responses from 19 time-dependent signaling profiles [39••]. This particular modeling approach calculates the most informative combinations of signals that together predict cellular functions. The combinations of stress, prodeath and prosurvival signals identified by the model were consistent with known mechanisms but could not have been predicted by inspection. Focusing on intracellular signals with recognized but complex roles in cell death thus allowed the model to identify new mechanisms of apoptosis control within the currently understood network. Very recently, we have found that this approach

to modeling network function could effectively capture cytokine-induced apoptotic responses that differ between diverse cell types (K Miller-Jensen, KA Janes, DA Lauffenburger, unpublished data). This suggests that different cell types might share a common network that converts signals into cellular responses. If true, then reconstruction models (Table 2) could soon encompass cellular functions, as has been done for genetic networks in prokaryotes [40]. Further modeling efforts aimed at predicting cellular functions will help understand how proteomic networks can be targeted therapeutically to improve or correct cell behavior.

Conclusions

The measurement, definition and function of proteomic networks must be addressed in such a way that models complement experiment. These networks remain too unconstrained to study by mathematics alone yet have become too complex to understand completely by intuition. We believe that insights here will come about by approaching models through biological questions. In line with this view, our review has focused on less detailed models with strong foundations in data (Figure 1). Very detailed models incorporating stochasticity have been successfully developed in simpler systems [41], and recent computational advances [42] might allow low-abundance signaling molecules to be examined in mammalian networks. We expect that modeling approaches will readily keep pace in providing network insight as proteomic techniques become more quantitative, high-throughput and sensitive [8,43].

Update

Recent work has provided new examples of how existing reconstruction models can be further explored and refined to aid biological discovery. Cheong *et al.* [53•] used an NF-κB signaling model [45] to identify the dynamics of inhibitor of NF-κB kinase (IKK) that were required for the dose-independent activation of NF-κB in cells stimulated with tumor necrosis factor. Werner *et al.* [54••] and Covert *et al.* [55••] separately revised the same NF-κB model [45] to study lipopolysaccharide-induced IKK signaling through Toll-like receptor 4.

Acknowledgements

We thank Suzanne Gaudet for critically reviewing this manuscript. The work in our laboratory described in this review was supported by grants from the National Institutes of Health (P50-GM68762 and U54-GM64346). KAJ is a postdoctoral fellow of the American Cancer Society (New England Division – SpinOdyssey 2005, PF-05-224-01-MGO).

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Ideker T, Lauffenburger D: **Building with a scaffold: emerging strategies for high- to low-level cellular modeling**. *Trends Biotechnol* 2003, **21**:255-262.

2. Eungdamrong NJ, Iyengar R: **Computational approaches for modeling regulatory cellular networks**. *Trends Cell Biol* 2004, **14**:661-669.

3. Vayttaden SJ, Ajay SM, Bhalla US: **A spectrum of models of signaling pathways**. *ChemBioChem* 2004, **5**:1365-1374.

4. Tyers M, Mann M: **From genomics to proteomics**. *Nature* 2003, **422**:193-197.

5. • Riddick G, Macara IG: **A systems analysis of importin-α-β mediated nuclear protein import**. *J Cell Biol* 2005, **168**:1027-1038.
See annotation to [7•].

6. • Swameye I, Muller TG, Timmer J, Sandra O, Klingmuller U: **Identification of nucleocytoplasmic cycling as a remote sensor in cellular signaling by databased modelling**. *Proc Natl Acad Sci USA* 2003, **100**:1028-1033.
See annotation to [7•].

7. • Smith AE, Slepchenko BM, Schaff JC, Loew LM, Macara IG: **Systems analysis of Ran transport**. *Science* 2002, **295**:488-491.
[5•–7•] are excellent examples of network models that include spatial detail. These studies use two-compartment models and partial differential equations (Figure 1) to examine the protein networks involved in nucleocytoplasmic shuttling and signaling.

8. Aebersold R, Mann M: **Mass spectrometry-based proteomics**. *Nature* 2003, **422**:198-207.

9. Nesvizhskii AI, Aebersold R: **Interpretation of shotgun proteomic data: the protein inference problem**. *Mol Cell Proteomics* 2005, **4**:1419-1440.

10. Krokhin OV, Craig R, Spicer V, Ens W, Standing KG, Beavis RC, Wilkins JA: **An improved model for prediction of retention times of tryptic peptides in ion pair reversed-phase HPLC: its application to protein peptide mapping by off-line HPLC-MALDI MS**. *Mol Cell Proteomics* 2004, **3**:908-919.

11. Keller A, Nesvizhskii AI, Kolker E, Aebersold R: **Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search**. *Anal Chem* 2002, **74**:5383-5392.

12. • Nesvizhskii AI, Keller A, Kolker E, Aebersold R: **A statistical model for identifying proteins by tandem mass spectrometry**. *Anal Chem* 2003, **75**:4646-4658.
[11] and [12•] use in-depth prior knowledge about the details of peptide and protein assignment from MS spectra to construct statistical models that quantify confidence in results derived from an MS experiment.

13. Martens H, Martens M: **Multivariate Analysis of Quality: An Introduction**John Wiley & Sons; 2001.

14. • Gunsalus KC, Ge H, Schetter AJ, Goldberg DS, Han JD, Hao T, Berriz GF, Bertin N, Huang J, Chuang LS *et al.*: **Predictive models of molecular machines involved in *Caenorhabditis elegans* early embryogenesis**. *Nature* 2005, **436**:861-865.
See annotation to [15•].

15. •• Gaudet S, Janes KA, Albeck JG, Pace EA, Lauffenburger DA, Sorger PK: **A compendium of signals and responses triggered by prodeath and prosurvival cytokines**. *Mol Cell Proteomics* 2005, **4**:1569-1590.
[14•] and [15••] are two recent reports that used models in combining datasets of heterogeneous measurements. Gunsalus et al. focus on the intersection of three measurement types (phenotypic, transcriptional and interaction), whereas Gaudet et al. focus on the union of four measurement types (kinase assays, immunoblotting, antibody arrays and flow cytometry).

16. Blagoev B, Ong SE, Kratchmarova I, Mann M: **Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics**. *Nat Biotechnol* 2004, **22**:1139-1145.

17. Zhang Y, Wolf-Yadlin A, Ross PL, Pappin DJ, Rush J, Lauffenburger DA, White FM: **Time-resolved mass spectrometry of tyrosine phosphorylation sites in the epidermal growth factor receptor signaling network reveals dynamic modules**. *Mol Cell Proteomics* 2005, **4**:1240-1250.

18. Janes KA, Albeck JG, Peng LX, Sorger PK, Lauffenburger DA, Yaffe MB: **A high-throughput quantitative multiplex kinase assay for monitoring information flow in signaling networks:**

application to sepsis-apoptosis. *Mol Cell Proteomics* 2003, **2**:463-473.

19. Nielsen UB, Cardone MH, Sinskey AJ, MacBeath G, Sorger PK: **Profiling receptor tyrosine kinase activation by using Ab microarrays**. *Proc Natl Acad Sci USA* 2003, **100**:9330-9335.

20. • Papin JA, Hunter T, Palsson BO, Subramaniam S: **Reconstruction of cellular signalling networks and analysis of their properties**. *Nat Rev Mol Cell Biol* 2005, **6**:99-111.
This review nicely compiles order-of-magnitude estimates for many important modeling parameters (Table 1).

21. Schoeberl B, Eichler-Jonsson C, Gilles ED, Muller G: **Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors**. *Nat Biotechnol* 2002, **20**:370-375.

22. • Sasagawa S, Ozaki Y, Fujita K, Kuroda S: **Prediction and validation of the distinct dynamics of transient and sustained ERK activation**. *Nat Cell Biol* 2005, **7**:365-373.
This study combines quantitative experiments and mechanistic modeling to investigate the dynamics of ERK activation after stimulation with epidermal growth factor (EGF) or NGF.

23. Park CS, Schneider IC, Haugh JM: **Kinetic analysis of platelet-derived growth factor receptor/phosphoinositide 3-kinase/Akt signaling in fibroblasts**. *J Biol Chem* 2003, **278**:37064-37072.

24. Bentele M, Lavrik I, Ulrich M, Stosser S, Heermann DW, Kalthoff H, Krammer PH, Eils R: **Mathematical modeling reveals threshold mechanism in CD95-induced apoptosis**. *J Cell Biol* 2004, **166**:839-851.

25. Hua F, Cornejo MG, Cardone MH, Stokes CL, Lauffenburger DA: **Effects of bcl-2 levels on fas signaling-induced caspase-3 activation: molecular genetic tests of computational model predictions**. *J Immunol* 2005, **175**:985-995.

26. Cho KH, Shin SY, Lee HW, Wolkenhauer O: **Investigations into the analysis and modeling of the TNF alpha-mediated NF-kappa B-signaling pathway**. *Genome Res* 2003, **13**:2413-2422.

27. Lee E, Salic A, Kruger R, Heinrich R, Kirschner MW: **The roles of APC and Axin derived from experimental and theoretical analysis of the Wnt pathway**. *PLoS Biol* 2003, **1**:E10.

28. Lai K, Robertson MJ, Schaffer DV: **The sonic hedgehog signaling system as a bistable genetic switch**. *Biophys J* 2004, **86**:2748-2757.

29. Sorger PK: **A reductionist's systems biology**. *Curr Opin Cell Biol* 2005, **17**:9-11.

30. •• Sachs K, Perez O, Pe'er D, Lauffenburger DA, Nolan GP: **Causal protein-signaling networks derived from multiparameter single-cell data**. *Science* 2005, **308**:523-529.
This study uses a Bayesian network approach to infer the connectivity of 11 signaling molecules by using single-cell data from polychromatic flow cytometry. The network includes seven protein kinases, as well as phosphoinositides, kinase substrates and phospholipases.

31. •• Plavec I, Sirenko O, Privat S, Wang Y, Dajee M, Melrose J, Nakao B, Hytopoulos E, Berg EL, Butcher EC: **Method for analyzing signaling networks in complex cellular systems**. *Proc Natl Acad Sci USA* 2004, **101**:1223-1228.
This study describes a combined experimental-computational approach for network inference called biologically multiplexed activity profiling (BioMAP). The BioMAP approach focuses on a defined protein subnetwork, which is then interrogated experimentally by overexpression in the context of various extracellular stimuli. [48] describes a later version of BioMAP applied to studying mechanisms of drug action.

32. Prudhomme W, Daley GQ, Zandstra P, Lauffenburger DA: **Multivariate proteomic analysis of murine embryonic stem cell self-renewal versus differentiation signalling**. *Proc Natl Acad Sci USA* 2004, **101**:2900-2905.

33. • Pawson T: **Specificity in signal transduction: from phosphotyrosine-SH2 domain interactions to complex cellular systems**. *Cell* 2004, **116**:191-203.
This outstanding historical review describes the emergence of systems thinking as it evolved from the earliest studies of phospho-dependent protein–protein interactions.

34. • Oda K, Matsuoka Y, Funahashi A, Kitano H: **A comprehensive pathway map of epidermal growth factor receptor signaling**. *Mol Syst Biol* 2005. doi:10.1038/msb4100014.
This review presents a high-quality literature-based network reconstruction of EGF signaling. Assembly of the known interactions reveals an hourglass shape in the network, with many sensors and initiators, relatively few transducers and many effectors.

35. Lee SJ, Hori Y, Groves JT, Dustin ML, Chakraborty AK: **Correlation of a dynamic model for immunological synapse formation with effector functions: two pathways to synapse formation**. *Trends Immunol* 2002, **23**:492-499.

36. Janes KA, Kelly JR, Gaudet S, Albeck JG, Sorger PK, Lauffenburger DA: **Cue-signal-response analysis of TNF-induced apoptosis by partial least squares regression of dynamic multivariate data**. *J Comput Biol* 2004, **11**:544-561.

37. Saucerman JJ, Brunton LL, Michailova AP, McCulloch AD: **Modeling beta-adrenergic control of cardiac myocyte contractility *in silico***. *J Biol Chem* 2003, **278**:47997-48003.

38. • Hautaniemi S, Kharait S, Iwabu A, Wells A, Lauffenburger DA: **Modeling of signal-response cascades using decision tree analysis**. *Bioinformatics* 2005, **21**:2027-2035.
This study introduces decision tree analysis as a modeling approach for signaling networks. Figure 1 in the review here is an unconstrained example of a decision tree.

39. •• Janes KA, Albeck JG, Gaudet S, Sorger PK, Lauffenburger DA, Yaffe MB: **A systems model of signaling identifies a molecular basis set for cytokine-induced apoptosis**. *Science* 2005, **310**:1646-1653.
This study uses partial least squares modeling to predict cellular responses from measurements of the upstream signaling network. The model identifies two groupings of intracellular signals — a stress-apoptosis grouping and a survival grouping — that constitute fundamental dimensions within the apoptotic signaling network.

40. Kalir S, Alon U: **Using a quantitative blueprint to reprogram the dynamics of the flagella gene network**. *Cell* 2004, **117**:713-720.

41. Weinberger LS, Burnett JC, Toettcher JE, Arkin AP, Schaffer DV: **Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity**. *Cell* 2005, **122**:169-182.

42. Lok L, Brent R: **Automatic generation of cellular reaction networks with Moleculizer 1.0**. *Nat Biotechnol* 2005, **23**:131-136.

43. Johnson SA, Hunter T: **Kinomics: methods for deciphering the kinome**. *Nat Methods* 2005, **2**:17-25.

44. Papin JA, Palsson BO: **The JAK-STAT signaling network in the human B-cell: an extreme signaling pathway analysis**. *Biophys J* 2004, **87**:37-46.

45. Hoffmann A, Levchenko A, Scott ML, Baltimore D: **The IkappaB-NF-kappaB signaling module: temporal control and selective gene activation**. *Science* 2002, **298**:1241-1245.

46. Lee KH, Dinner AR, Tu C, Campi G, Raychaudhuri S, Varma R, Sims TN, Burack WR, Wu H, Wang J *et al.*: **The immunological synapse balances T cell receptor signaling and degradation**. *Science* 2003, **302**:1218-1222.

47. Woolf PJ, Prudhomme W, Daheron L, Daley GQ, Lauffenburger DA: **Bayesian analysis of signaling networks governing embryonic stem cell fate decisions**. *Bioinformatics* 2005, **21**:741-753.

48. Kunkel EJ, Dea M, Ebens A, Hytopoulos E, Melrose J, Nguyen D, Ota KS, Plavec I, Wang Y, Watson SR *et al.*: **An integrative biology approach for analysis of drug action in models of human vascular inflammation**. *FASEB J* 2004, **18**:1279-1281.

49. Colman-Lerner A, Gordon A, Serra E, Chin T, Resnekov O, Endy D, Pesce CG, Brent R: **Regulated cell-to-cell variation in a cell-fate decision system**. *Nature* 2005, **437**:699-706.

50. Christopher R, Dhiman A, Fox J, Gendelman R, Haberitcher T, Kagle D, Spizz G, Khalil IG, Hill C: **Data-driven computer**

simulation of human cancer cell. *Ann N Y Acad Sci* 2004, **1020**:132-153.

51. Grigoriev A: **On the number of protein-protein interactions in the yeast proteome**. *Nucleic Acids Res* 2003, **31**:4157-4161.

52. Huh WK, Falvo JV, Gerke LC, Carroll AS, Howson RW, Weissman JS, O'Shea EK: **Global analysis of protein localization in budding yeast**. *Nature* 2003, **425**:686-691.

53. • Cheong R, Bergmann A, Werner SL, Regal J, Hoffmann A, Levchenko A: **Transient IKK activity mediates NF-kappa B temporal dynamics in response to a wide range of TNFalpha doses**. *J Biol Chem* 2005: in press.

This study used an existing NF-κB model [45] to discover that rapid IKK activation is crucial for the observed NF-κB dynamics following stimulation with varying doses of tumor necrosis factor.

54. •• Werner SL, Barken D, Hoffmann A: **Stimulus specificity of gene expression programs determined by temporal control of IKK activity**. *Science* 2005, **309**:1857-1861.

See annotation to [55••].

55. •• Covert MW, Leung TH, Gaston JE, Baltimore D: **Achieving stability of lipopolysaccharide-induced NF-kappaB activation**. *Science* 2005, **309**:1854-1857.

This and [54••] revise an NF-κB model [45] to reveal the role of autocrine tumor necrosis factor in controlling the late, sustained phase of IKK activation induced by lipopolysaccharide.