Bulletin of Mathematical Biology (2002) **64**, 133–145
doi:10.1006/bulm.2001.0269
Available online at http://www.idealibrary.com on IDEAL®

Generalized Analysis of Experimental Data for Interrelated Biological Measurements

ZVI KAM
Department of Molecular Cell Biology,
Weizmann Institute of Science,
Rehovot 76100,
Israel
E-mail: zvi.kam@weizmann.ac.il

Important biological mechanisms, such as signal transduction and gene expression, are mediated by numerous interacting multifunctional molecules, whose expression and activation are tightly regulated in space and time in response to stimuli. In order to describe the network of functional inter-relationships that govern such mechanisms, we use simple algorithms to interpret multiple variable measurements, identify the prominent participants, evaluate their interactions and obtain a 'functional fingerprint' of cell behaviour. Dynamic measurements of responses yield hierarchical information about causal relations in the underlying pathway. As a proof of principles we apply this approach to phosphorylation assays in protein gels, probing hormone and insulin signalling.

1. INTRODUCTION

Living systems are being explored with a broad spectrum of methods whose capabilities range from the level of individual molecules to that of intact organisms. With better understanding of the molecules, more attention is directed to their role in the functioning of whole systems. The emerging complexity of biological mechanisms makes the extension from components to modules rather nontrivial. This is attributed to two faces of the same coin: first, to the large number of participating molecules and second, to the large number of interactions between them (Hartwell *et al*., 1999). In this work we make use of the network of interactions to identify the molecules that respond to a stimulus, and characterize the connectivity between them.

Searching for the molecules involved in a process, new and old techniques deal with exhaustive probing of the whole biological context, measuring multivariable patterns of proteins and their modifications (Celis *et al*., 1996; Østergaard *et al*., 1997) or of gene expression levels (Arkin *et al*., 1997; DeRisi *et al*., 1997; Spellman *et al*., 1998; Wen *et al*., 1998; Alon *et al*., 1999; Eisen *et al*., 1998). Analysis of such patterns during development or following stimulation reveals many induced changes. Thorough characterization of these changes attempt to assign

0092-8240/02/010133 + 13 $35.00/0

'cause' and 'effect' relations which sometimes appear obvious (e.g., absence of down-stream events after blocking the activity of a component, for example, by mutational analysis), but become difficult when the changes 'distribute' over many molecules, and overlap for different processes. The distributed response is due to multiple inducing and inhibiting interactions and feedback loops that emerged from duplication of genes that form networks of cross-talking molecules. These interactions have been modified throughout evolution so that the whole network responds quantitatively to a growing repertoire of environmental signals, in a context and cell-type specific manner (Bray, 1990; Alberts *et al*., 1994; Bray and Lay, 1994; Bray, 1995). Branched networks are 'robust', as evidenced by their persisting outputs under changing environmental parameters, and the many cases of mild phenotypes in transgenic animals with altered or deleted pathway components. However, one can often locate 'degenerate' network structures, such as regulatory 'checkpoints', where interactions are channelled through a single component (a decision node) that integrates information from several sources and is solely responsible for the decision to activate or inhibit signal progression. Processes mediated by such nodes are extremely sensitive to mutations and are often associated with severe abnormalities such as cancer (Fearon and Vogelstein, 1990). Characterization of the connectivity architecture in cellular networks is therefore important for understanding cellular mechanisms and for developing approaches to control or modulate them.

In this work we propose a simple yet general approach for a preliminary analysis of network architectures using experimental measurements of changes associated with a biological process. A classical source of such data is protein gels. Tyrosine phosphorylation of proteins is a good reporter of changes in signalling pathway component activities, and was analysed in the examples here. The examples deal with steady state and with time-dependent data. The results indicate what are the most relevant variables in the studied process and what are the potentially novel components. Connectivity is evaluated, so that suitable variables could be considered in hypotheses of models and molecular mechanisms, and might be the focus for the design of specific experiments. The results are also expected to serve as multiparameter functional fingerprints of complex cellular states. Genomic and proteomic measurements employing large-scale probing techniques are potentially very attractive data banks for this kind of preliminary exploratory analysis.

2. METHOD OF ANALYSIS

Given measured changes in molecular components of a biologically responding system, how can one learn about the underlying mechanism? We recall the often made analogy between signalling pathways and man-made control circuits (Brent, 2000). When electronic circuits are hidden in a 'black box' one can identify the constituent components and their connectivity by introducing specific inputs

('stimuli'), measuring the outputs ('responses'), assuming linearity and applying linear algebra to solve the transformation between input and output signals (Bode, 1955; Kuo, 1966; Skilling, 1974). The present work is based on the related method of neural networks (NNs), a field originally inspired by brain research, and expanded to a powerful model-fitting tool. Recently, various mathematical methods have been utilized to investigate multivariable biological measurements, such as temporal pattern correlation (Arkin *et al*., 1997; DeRisi *et al*., 1997; Spellman *et al*., 1998), clustering (Michaels *et al*., 1998; Wen *et al*., 1998; Eisen *et al*., 1998; Iyer *et al*., 1999), Boolean networks (Somogyi and Sniegoski, 1996; Akutsu *et al*., 1999) and neural network techniques. NNs have served to fit detailed models to known molecular mechanisms and paradigms such as in bacterial chemotaxis (Bray *et al*., 1993; Alon *et al*., 1999), immune response (Marille *et al*., 1996) and development (Burstein, 1995; Marnellos and Mjolsess, 1998). Alternatively, the network architecture can be determined from the actual measurements ['reverse engineering', see Arkin *et al*. (1997), Liang *et al*. (1998), Akutsu *et al*. (1999), Weaver *et al*. (1999), Plouraboué *et al*. (1992), and Thieffry *et al*. (1998)], as is the purpose of the present study. We shortly outline the formalism, and its application here.

A neural network implements a matrix transformation:

$$O_j^{(k)} = f\left\{\sum_{i=1}^{N} W_{ji} I_i^{(k)} + B_j\right\}, \qquad (k = 1, \ldots, L;\ j = 1, \ldots, M) \qquad (1)$$

where the matrix of weights, W_{ji}, operates on sets of input variables, $I_i^{(k)}$, [$I_i^{(k)}$ is the jth component of the kth input vector] and computes sets of output variables, $O_j^{(k)}$ (Beale and Jackson, 1990; Demuth and Beale, 1993). The number of variables in the input and output sets are N and M, and the number of such sets is L. The function f is a monotonically increasing 'transfer function', relating the inducing and responding signals, and B_j are 'bias' parameters that quantitate threshold and basal activities. The experimental variables can be conveniently organized to visualize their inter-relationships (evaluated by the NN weights) as shown in Fig. 1 (bias not shown). NNs are 'trained' to produce prescribed 'target' outputs when 'presented' with inputs ('learning'). This is achieved by starting from an initial guess for the biases and weights and modifying them iteratively by a 'learning rule'. The commonly applied Widrow–Hoff adaptive linear network (ADALINE) learning rule (Widrow and Stern, 1985) iteratively computes changes in the weights and biases:

$$\delta W_{ji} = \sum_{k=1}^{L} [E_j^{(k)} - O(k)_j^{(k)}] I(k)_i \qquad \delta B_j = \sum_{k=1}^{L} [E_j^{(k)} - O_j^{(k)}]. \qquad (2)$$

Here $E_j^{(k)}$ are the experimentally measured data sets, and $O_j^{(k)}$ are the outputs computed from equation (1). For L larger than N and M, the weights can be computed

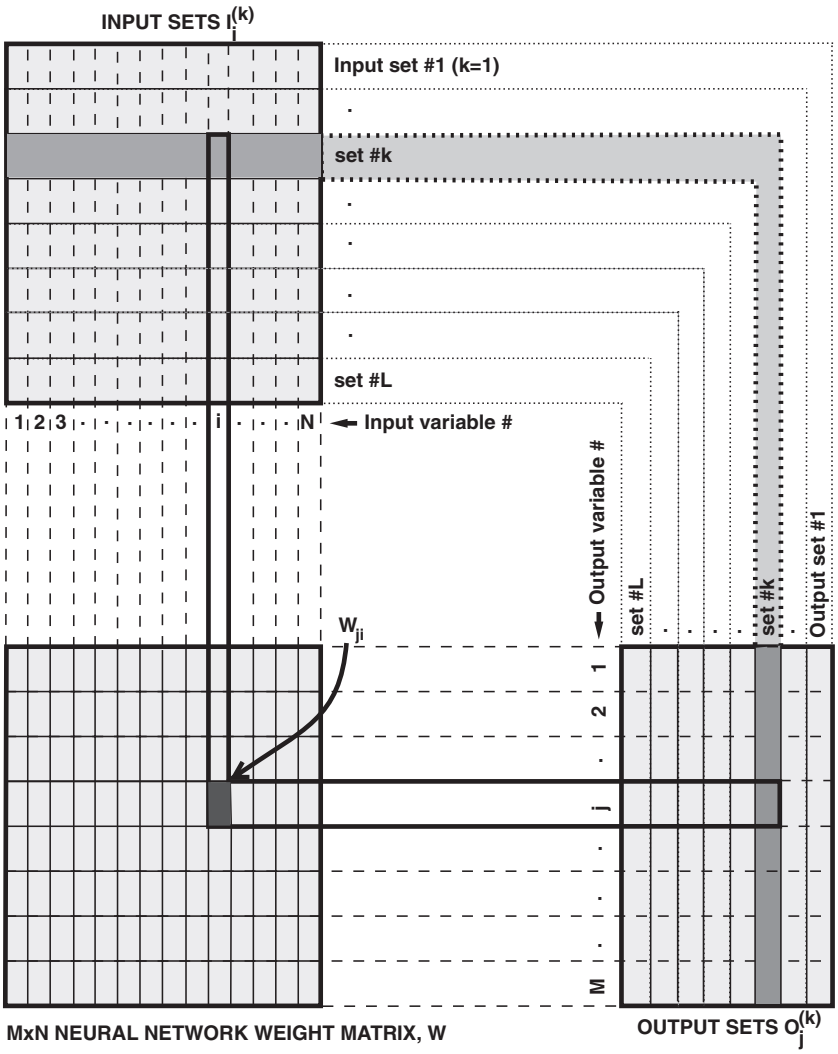

Figure 1. Multivariable data sets (input and output) and a matrix of neural network weights (bias B_j is not shown). In order to visually relate the weights to the variables, the input and output vector sets are shown as rows and columns, respectively, such that the weight W_{ji} (black square) presents the 'interaction' between an input variable in column i and an output variable in row j. When the neural network is presented with the input set number k (dark grey), it computes the corresponding output set k.

directly by generalized matrix inversion or least square fit. For small L, there is no unique solution, and the trained network parameters depend on the initial guess. Yet, the average of many solutions that start from random guesses decreases the values of fluctuating weights that are free to wonder unconstrained by the data. This process may mimic stochastic cell behaviour, which integrates fluctuating responses of individual molecular events (McAdams and Arkin, 1999). For a linear transfer function such 'minimal' weight solution can be uniquely calculated

using singular value decomposition (Press *et al*., 1992), avoiding large and opposite weights that mutually compensate each other's contribution. These weights may be considered a first order perturbation approximation for the nonlinear case. Uniqueness is not guaranteed for highly nonlinear transfer functions (Krauth and Mézard, 1987), but iterations starting at the solution for a linear transfer function converge to weights also characterized by small variation. Brute-force averaging is then required for the nonlinear case. *The key point is that the largest weights emerging from such a method of averaging NN solutions indicate linked input and output variables that dominate the overall behaviour, while small weights imply variables that may not contribute significantly*.

When presented with yet uncharacterized input sets ('recognition'), the NN outputs can be used to identify input patterns ('fingerprinting'). The definition of input and output variables is extremely versatile; they need not correspond to stimuli and responses and can include variables of different types. In fact the same variables measured in different experiments or times can be used as both inputs and outputs (see later).

The majority of NN applications split the task of processing data into consecutive operations by several weight matrices on 'hidden layers' of variables. The power of multilayer networks is associated with this hidden flexibility. The present application aims to deduce the network architecture from the weights, and therefore uses a single weight matrix that directly connects only measured input and output variables. For this simple case the arrangement of the large weights corresponds directly to 'skeletal' network architectures of typical biological pathways as depicted for idealized cases in Fig. 2. In real-life, a more diffuse picture is expected due to limited experimental precision, unknown response functions, and simplifying assumptions. Nevertheless, as shown below, the largest weights can be indicative of the network dominant connectivity, and serve to sketch a skeleton as a preliminary description of the mechanisms underlying a pathway.

3. EXAMPLES

Two examples of multivariable analyses of experiments are provided here, both using tyrosine residues phosphorylation (Py) data, measured from protein two- and one-dimensional gels. The first data set is a single time point analysed to create a fingerprint for a cellular state, and the second data set is a dynamic analysis of cell response with a reconstruction of a network skeleton.

The first example analyses the changes in Py of 60 proteins induced by epidermal growth factor (EGF) taken from Tables 1 and 2 of Romano *et al*. (1994). This study evaluated Py in two-dimensional (2D) gels for two cell lines, one expressing the EGF receptor and the other a chimeric EGF receptor with erbB-2 intracellular kinase (as the wild-type erbB-2 does not have a known ligand). Two different phosphorylation assays were used: radioactive ^{32}P labelling following immuno-

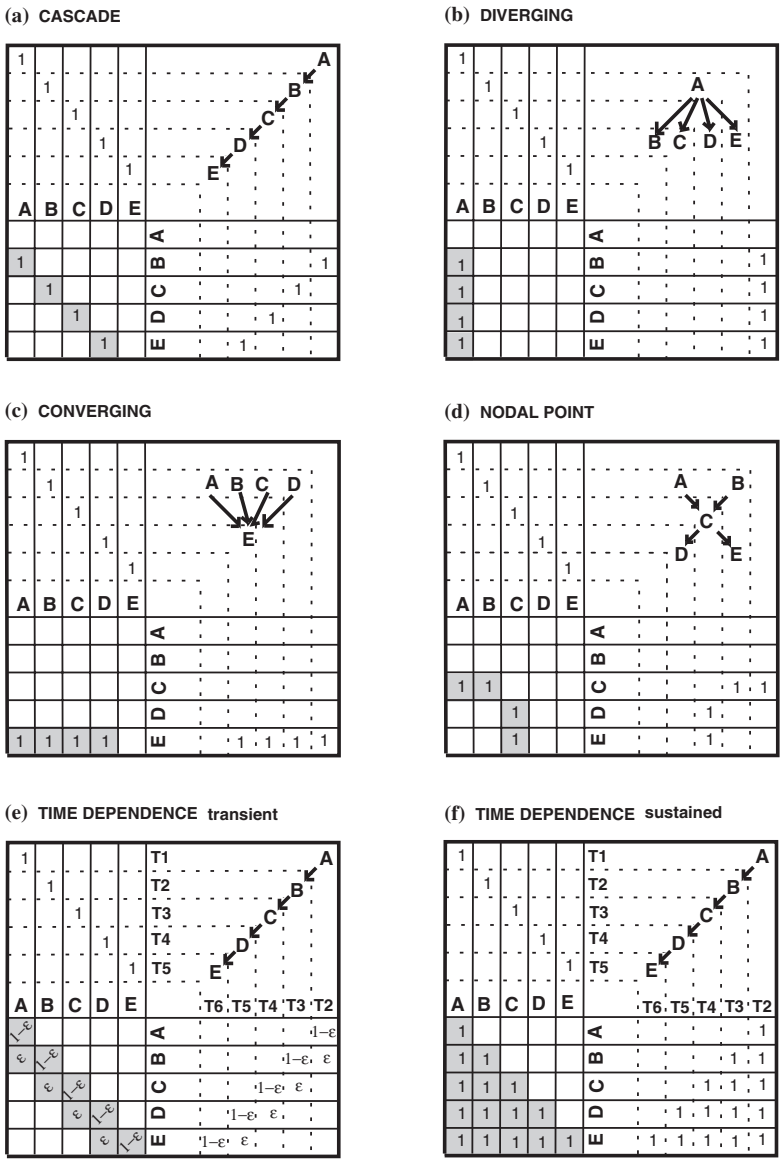

Figure 2. Weight matrices for skeletal network architectures with examples of corresponding idealized data sets. Five variables (A, B, C, D, E), which comprise both the input and output data sets, quantify the state of activation of five components. Thus, the output set (a_0, b_0, c_0, d_0, e_0) denotes the measurements following activation according to the values of the input set (a_1, b_1, c_1, d_1, e_1). (a)–(c) Cascade, diverging and converging pathways with nonzero weights along an off-diagonal, a column, and a row, respectively (empty cells are zero, or negligibly small weights); (d) nodal point, which is the meeting between converging and diverging pathways, depicted as the cross between a row and a column of large weights; (e), (f) dynamic data, where each output set serves as the next time input set, and the weights evaluate correlated causal changes; (e) transient response, induced through a cascade of events. The off-diagonal weight, ε, denotes the fractional change in consecutive times (ε^2 neglected, under the assumption that consecutive times are close together); (f) sustained activation of a cascade (which corresponds to a triangular matrix).

precipitation with anti-Py antibodies, and gel blot staining with anti-Py antibodies. This gives altogether four input data sets (top panel, Fig. 3). Romano *et al*. (1994) sorted the protein peaks into two groups of EGF-induced Py (protein spots labelled by numbers) or erbB-2-induced Py (labelled by letters). Elevated levels of phosphorylation of three protein (spots H, J and Z) were correlated with the mitogenic phenotype expressing the erbB-2 chimera (Romano *et al*., 1994). Our single output variable was mitogenic induction (from mitotic index) which was 100 times higher for the cell line expressing the chimera. The resulting neural network weights using the ADALINE learning rule (Widrow and Stern, 1985) are shown in the bottom panel, Fig. 3. Qualitatively similar results are obtained using back-propagation with sigmoidal transfer function (Beale and Jackson, 1990). Most of the largest weights are also obtained when the data sets for the radioactive or the chromogenic assays are analysed separately (standard deviation of weight differences is 65%). The EGF-induced and erbB-2-induced phosphorylations are clearly depicted by the negative (blue) and positive (red/black) regions in the weights.

The trained NN clearly identified these three variables [see the labelled largest weights in red and black, Fig. 3(b)]. The NN analysis identified additional components (B, K and S) that had large weights. Protein spots 3, 9, and 17, and some other weaker spots strongly correlated negatively with mitogenic induction (dark blue), which was not readily apparent from direct examination of the data. The negative weights may suggest inhibitory molecular components in the induction of mitogenic response, such as proteins containing YXXM motif that interfere with SHC activation (Prigent and Gullick, 1994).

The data for the EGF-stimulated cells were based on measurements at a time point well after the signalling pathway components sensed and responded to the EGF. They can serve to 'fingerprint' the resulting fast proliferating cell state (probably reflecting abundant phosphorylated proteins), but the hierarchy of activated events cannot be inferred. This type of information can be revealed from time-dependent measurements.

In the second example we examined the temporal changes of Py patterns induced by insulin and studied from one-dimensional (1D) gels kindly provided by Biener *et al*. (1996). CHO cells stably transfected with the insulin receptor (IR) cDNA expressed about $5 \cdot 10^5$ IR/cell. At various roughly logarithmic time intervals following incubation with insulin, cell extracts were analysed for Py-containing proteins using 1D SDS gel electrophoresis and immunoblotting with anti-Py antibodies. Gels were digitally scanned and analysed with quantity one software (PDI Huntington station, NY, U.S.A.). The resulting data consist of background-subtracted protein band intensities for every gel lane, which are matched for all the analysed lanes, and exported in a spread sheet directly presented as data sets for our analysis. For every time point, matched sets of 13 bands were averaged for four measurements, with and without the addition of vanadate (which enhances phosphorylation by inhibiting phosphatases). The time-dependent Py intensities were presented as inputs to the neural network. Target outputs used are the next-time

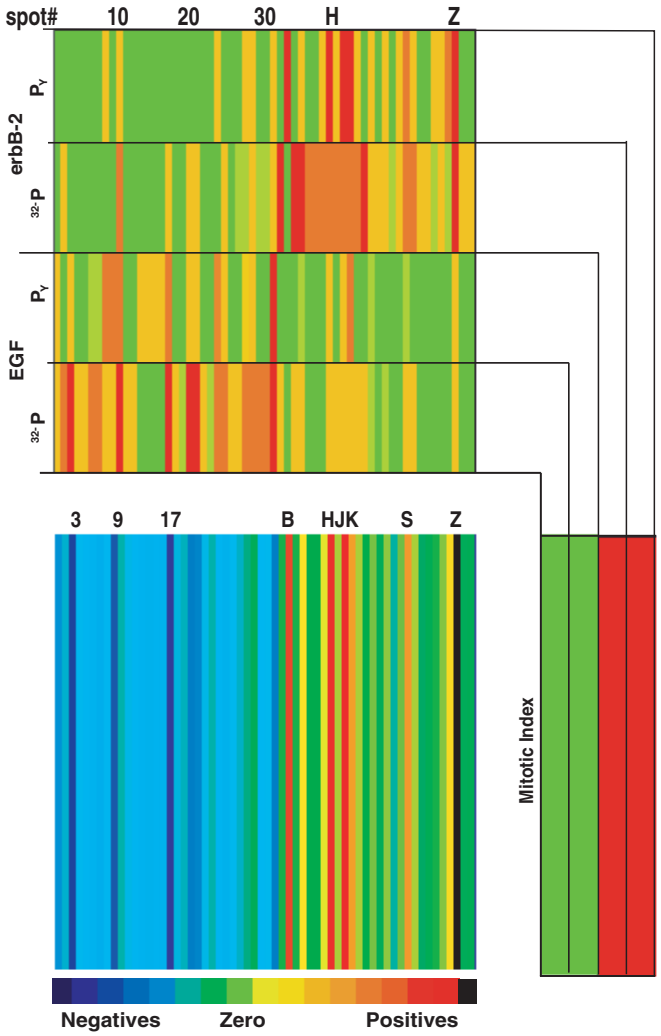

Figure 3. Analysis of EGF-induced tyrosine phosphorylation patterns mediated by two different erb receptor family members, and their mitogenic response, from Romano *et al*. (1994). **Top panel**: the four input data sets of Py levels evaluated from two assays each applied to two cell lines transfected with EGF-receptor or erbB-2 chimera. Each set consists of 60 variables that correspond to the level of Py for 60 protein spots in 2D gels; the first group (assigned as numbers 1–30) indicated EGF-receptor-induced phosphorylations, and the second group (letters A–Z, α–δ) indicated the erbB-2-induced phosphorylations. The five grades of phosphorylation levels assigned in Romano *et al*. (1994) are shown in red, orange, yellow, light green and green. The single output variable (bottom right) is mitotic index, taken after Romano *et al*.to be 1 (red) for the erbB-2, and 0.01 (green) for the EGF cell lines. **Bottom panel**: the elements of the 60×1 neural network weight matrix presenting the average of 200 training cycles, each starting from random weights and converging after about 100 iterations. The processing time on a Silicon Graphics O2 workstation was a few seconds. The overall correlation of the numbers with negative (blue) weights, and letters with positive weights is compatible with their functional grouping. Green indicates small weight values. Spots that correspond to the largest positive and negative weights are labelled.

inputs. The data sets and the resulting weight matrix are presented in Fig. 4(a). Starting at the inducing insulin stimulus and reading the largest positive and negative weights as connecting arrows between variables, one gets the pathways in Fig. 4(b). Insulin induces phosphorylation of insulin receptor (IR), its substrate (IRS), Shc, Annexin II, MAPK and other proteins. These phosphorylations are downregulated within several hours. All these known key features in the insulin-induced events are depicted, including the feedback contribution of Annexin II that downregulates the response to insulin (Biener *et al*., 1996). Two components (#4 and #7) are unknown responsive molecules. Although the hierarchy proposed here is based exclusively on temporal measurements, it is in agreement with current understanding of the insulin response (Whitehead *et al*., 2000), which has been elucidated by genetic manipulations, specific inhibitions, and other methods. This analysis is not very sensitive to elimination of weakly-connected variables (standard deviation of changes 15%). Even deletion of a variable (e.g., shc46) that strongly connects two others (IR to MAPK in this case) will be replaced by a new strong direct connection between the latter two (and changes in the other weights with standard deviation of 26%). This feature implies that even when a molecule is not directly probed by a set of measurements, its effect on other measurements can be inferred. Overlapping bands (such as IRS1/2) or bands that include multiple phosphorylated states pool their contributions to the response.

4. CONCLUSIONS

The linked behaviour of variables is expected to be preserved when moving from responsive cell lines, with high activation levels of relevant molecules, to live tissues or developing embryos with more subtle changes. If indeed linked changes preserve their relative amplitudes, the approach described here can facilitate the characterization of complex patterns of multiple RNA and protein changes in cellular processes and *in vivo* states such as proliferation, differentiation, growth arrest, or apoptosis. Such information would be extremely helpful for analyzing and characterizing both normal physiological and pathological conditions in developing and mature organisms.

It should be kept in mind that the cited examples are based exclusively on phosphorylation levels. Phosphorylation is not proportional to the levels of activation or inhibition of a molecule. Furthermore, responses to stimuli include processes that are not probed by phosphorylation changes, notably translocations between cell compartments (which modulate activity by accessibility) and expression of new components. Time-dependent cellular processes may therefore be described faithfully only by spatio-temporal differential equations (Chock and Stadtman, 1977; Marille *et al*., 1996), which depend on hard to estimate molecular rate and binding constant parameters. The deduced network weights should not be considered as a source for such quantitative molecular parameters. Yet, they can effectively serve

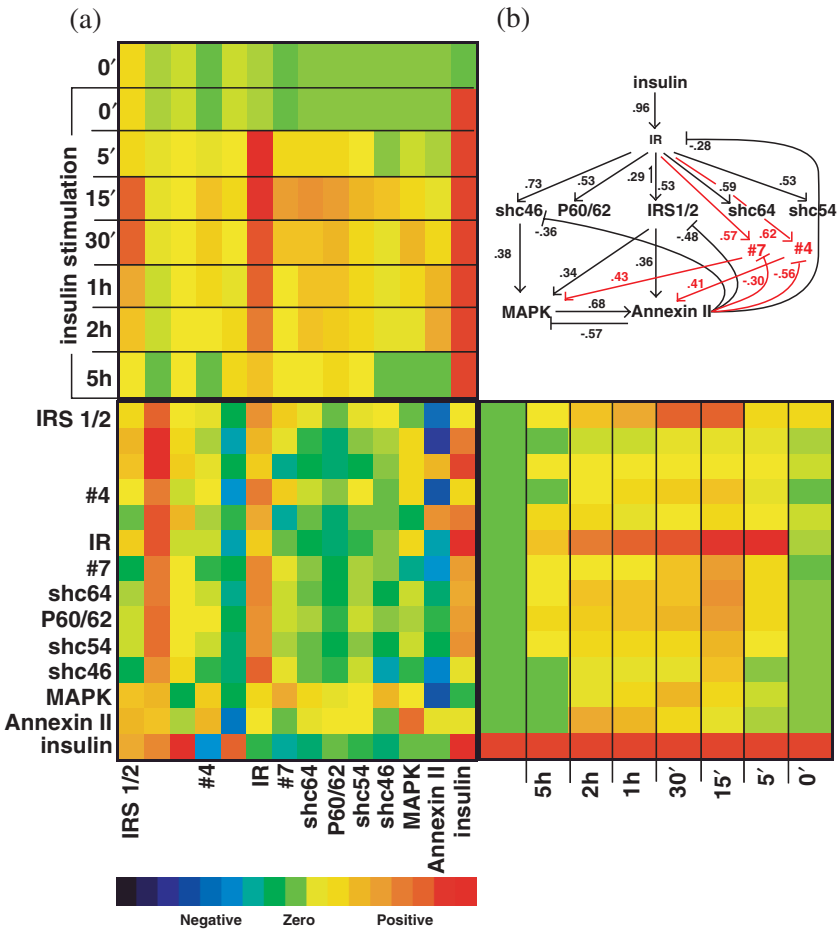

Figure 4. Analysis of the insulin signalling pathway based on the dynamics of tyrosine phosphorylation responses, from Biener *et al*. (1996). (a) The time-dependent data sets (known proteins are labelled) and the weight matrix averaged for 1000 random initial guesses are depicted according to the convention of Fig. 1. Values: green—close to zero, red—large positive, and blue—large negative. All but the last variables denote phosphorylation levels; the last variable is insulin concentration. (b) The insulin-induced network of interactions, deduced from the weight matrix in panel a. The connectivity between pairs of proteins are drawn from the list of largest positive and negative weights, starting with insulin. Inducing and inhibiting interaction weights are depicted by arrow-head and bar-head lines, respectively, with the weight values marked next to the line. Arrows connecting two proteins that are already connected strongly via a third protein are not drawn. Two unknown molecules (#4 & 7) and their arrows are drawn in red.

in preliminary examination of the data to indicate the most relevant variables and characterize the dominating rules of complex behaviour, suggesting a focus in the design of specific measurements [such as in Remy and Michnick (2001)], particularly following high-throughput experiments. The fact that, although simple, the analysis depicted all major components emphasizes the utility of this approach.

ACKNOWLEDGEMENTS

We thank Y Zick and Y Biener for providing the insulin data; E Berkovich, U Alon, A Chausovsky, J Schlessinger and Y Yarden for joint tours in signalling pathways; and B Geiger for countless discussions and suggestions. This study was supported by the Yad Aavraham Centre for Cancer Diagnosis and Therapy. ZK is the Israel Pollack Professor of Biophysics.

REFERENCES

Akutsu, T., S. Miyano and S. Kuhara (1999). Identification of genetic networks from a small number of gene expression patterns under the Boolean network model. *Pac. Symp. Biocomput.* **4**, 17–28.

Alberts, B., D. Bray, J. Lewis, M. Raff, K. Roberts and J. D. Watson (1994). *Molecular Biology of the Cell*, 3rd edn, New York: Garland, pp. 778.

Alon, U., N. Barkai, D. A. Notterman, D. Mack, K. Gish, S. Leibler and A. J. Levine (1999). Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc. Natl. Am. Sci.* **8**, 6745–6750.

Alon, U., M. G. Surette, N. Barkai and S. Leibler (1999). Robustness in bacterial chemotaxis. *Nature* **397**, 168–171.

Arkin, A., P. Shen and J. Ross (1997). A test case of correlation metric construction of a reaction pathway from measurements. *Science* **277**, 1275–1279.

Beale, R. and T. Jackson (1990). *Neural Computing: An Introduction*, New York: Adam Hilger.

Biener, Y., R. Feinstein, M. Mayak, Y. Kaburagi, T. Kadowaki and Y. Zick (1996). Annexin II is a novel player in insulin signal transduction. Possible association between annexin II phosphorylation and insulin receptor internalization. *J. Biol. Chem.* **271**, 29489–29496.

Bode, H. W. (1955). *Network Analysis and Feedback Amplifier Design*, New York: D. van Nostrand.

Bray, D. (1990). Intracellular signalling as a parallel distributed process. *J. Theor. Biol.* **143**, 215–231.

Bray, D. (1995). Protein molecules as computational elements in living cells. *Nature* **376**, 307–312.

Bray, D., R. B. Bourret and M. I. Simon (1993). Computer simulation of the phosphorylation cascade controlling bacterial chemotaxis. *Mol. Biol. Cell* **4**, 469–482.

Bray, D. and S. Lay (1994). Computer simulated evolution of a network of cell signaling molecules. *Biophys. J.* **66**, 972–977.

Brent, R. (2000). Genomic biology. *Cell* **100**, 169–183.

Burstein, Z. (1995). Protein molecules as computational elements in living cells. *J. Theor. Biol.* **174**, 1–11.

Celis, J. E. *et al*. (1996). Human 2-D PAGE databases for proteome analysis in health and disease `http:biobase.dk/cgi-bin/celis`. *FEBS Lett.* **398**, 129–134.

Chock, P. B. and E. R. Stadtman (1977). Superiority of interconvertible enzyme cascades in metabolite regulation: analysis of multicyclic systems. *Proc. Natl. Acad. Sci. USA* **74**, 2766–2770.

DeRisi, J. L., V. R. Iyer and P. O. Brown (1997). Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**, 680–686.

Demuth, H. and M. Beale (1993). MATLAB Neural Network TOOLBOX, The MATH WORKS Inc., Natick, MA.

Eisen, M. B., P. T. Spellman, P. O. Brown and D. Botstein (1998). Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci.* **95**, 14863–14868.

Fearon, E. R. and B. Vogelstein (1990). A genetic model for colorectal tumorigenesis. *Cell* **61**, 759–767.

Hartwell, L. H., J. J. Hopfield, S. Leibler and A. W. Murray (1999). From molecular to modular cell biology. *Nature* **402 (Suppl. 6761)**, c47–c52.

Iyer, V. R. *et al*. (1999). The transcriptional program in the response of human fibroblasts to serum. *Science* **283**, 83–87.

Krauth, W. and M. Mézard (1987). *J. Phys. A. (Math. Gen.)* **20**, L745–L752.

Kuo, F. F. (1966). *Network Analysis and Synthesis*, 2nd edn, New York: John Wiley Sons.

Liang, S., S. Fuhrman and R. Somogyi (1998). Reveal, a general reverse engineering algorithm for inference of genetic network architectures. *Pac. Symp. Biocomput.* **3**, 18–29.

Marille, E., D. Thieffry, O. Leo and M. Kaufman (1996). Toxicity and neuroendocrine regulation of the immune response: a model analysis. *J. Theor. Biol.* **183**, 285–305.

Marnellos, G. and E. Mjolsess (1998). A gene network approach to modeling early neurogenesis in Drosophila. *Pac. Symp. Biocomput.* **3**, 30–41.

McAdams, H. H. and A. Arkin (1999). It's a noisy business! Genetic regulation at the nanomolar scale. *Trends Gen.* **15**, 65–69.

Michaels, G. S., D. B. Carr, M. Askenazi, S. Fuhrman, X. Wen S. and R. Somogyi (1998). Cluster analysis and data visualization of large-scale gene expression data. *Pac. Symp. Biocomput.* **3**, 42–53.

Østergaard, M., H. H. Rasmussen, H. V. Nielsen, H. Vorum, T. F. Ørntoft, H. Wolf and J. E. Celis (1997). Proteome profiling of bladder squamous cell carcinomas: identification of markers that define their degree of differentiation. *Cancer Res.* **57**, 4111–4117.

Plouraboué, F., H. Atlan, G. Weisbuch and J.-P. Nadal (1992). A network model of the coupling of ion channels with secondary messenger in cell signaling. *Network* **3**, 393–406.

Press, W. H., S. A. Teukolsky, W. T. Vetterling and B. P. Plannery (1992). *Numerical Receipts*, 2nd edn, Cambridge University Press, pp. 51–62.

Prigent, S. A. and W. J. Gullick (1994). Identification of c-erbB-3 binding sites for phosphatidylinositol 3′-kinase and SHC using an EGF receptor/c-erbB-3 chimera. *EMBO J.* **13**, 2831–2841.

Remy, I. and S. W. Michnick (2001). Visualization of biochemical networks in living cells. *Proc. Natl. Acad. Sci. USA* **98**, 7678–7683.

Romano, A., W. T. Wong, M. Santoro, P. J. Wirth, S. S. Thorgeirsson and P. P. Di Fiore (1994). The high transforming potency of erbB-2 and ret is associated with phosphorylation of paxillin and a 23 kDa protein. *Oncogene* **9**, 2923–2933.

Skilling, H. H. (1974). *Electric Networks*, New York: John Wiley and Sons.

Somogyi, R. and C. A. Sniegoski (1996). Modeling the complexity of genetic networks: understanding multigenetic and pleiotropic regulation. *Complexity* **1**, 45–63.

Spellman, P. T., G. Sherlock, M. Q. Zhang, V. R. Iyer, K. Anders, M. B. Eisen, P. O. Brown, D. Botstein and B. Futcher (1998). Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. *Mol. Biol. Cell* **9**, 3273–3297.

Thieffry, D., A. M. Huerta, E. Pérez-Rueda and J. Collado-Vides (1998). From specific gene regulation to genomic networks: a global analysis of transcriptional regulation in Escherichia coli. *BioEssays* **20**, 433–440.

Weaver, D. C., C. T. Workman and G. D. Stormo (1999). Modeling regulatory networks with weight matrices. *Pac. Symp. Biocomput.* **4**, 112–123.

Wen, X., S. Fuhrman, G. S. Michaels, D. B. Car, S. Smith, J. L. Barker and R. Sologyi (1998). Large-scale temporal gene expression mapping of central nervous system development. *Proc. Natl. Acad. Sci.* **95**, 334–339.

Whitehead, J. P., S. F. Clark, B. Urso and D. E. James (2000). Signaling through the insulin receptor. *Curr. Opin. Cell Biol.* **12**, 222–228.

Widrow, B. and S. D. Stern (1985). *Adaptive Signal Processing*, New York: Prentice-Hall.

Received 1 May 2001 and accepted 10 October 2001