



The Fast Track to Valuable Drug Targets

Helios Biosciences White Paper from Biofutur 275 (March 2007) pp. 40-44

« System Biology Challenges »

www.heliosbiosciences.com

HELIOS BIOSCIENCES
8 RUE DE GENERAL SARRAIL
94010 CRETEIL-CEDEX
FRANCE

The Target Designer

Defining therapeutic targets with dynamic models

Progress in life science research has shown that most diseases involve highly complex biological phenomena. The study of these diseases therefore requires the development of models reflecting this complexity. Helios Biosciences has developed such mathematical models for the generation of new, more effective treatment strategies.

Most morbidity and mortality in industrialised countries is associated with multifactorial diseases for which current treatment is insufficient, for example cancers, cardiovascular diseases, neurological diseases and autoimmune diseases. The identification and validation of novel, more effective treatment strategies are therefore required.

Large-scale description of living organisms has been possible for the last 10 or 15 years, and this in turn makes it possible to study these diseases. The descriptive approaches involved generate essentially two types of data:

- 1) Quantification of biomolecules through “expression screening”, as carried out with DNA microarrays containing sequences representing all the mRNAs produced in a given cell or tissue, and
- 2) Identification of interactions between these biomolecules. The compilation of these interactions can be used to reconstitute signalling (intracellular and intercellular) and metabolic (intracellular) networks.

This descriptive capacity has led to an increase in the number of potential

therapeutic targets of weak value such that the failure rate for new target development has increased and therapeutic progress has been limited. Of the 881 million dollars currently required for the successful development of a new drug, 665 million dollars are due to failures, 42% (275 million dollars) of which are due to erroneous therapeutic target selection (1). This situation is the consequence, in large part, of insufficient development of analysis and integration strategies suitable for the massive amounts of data generated by new technologies for studying living organisms.

A dynamic model is a mathematical model that can be used for simulations *in silico*. It reproduces certain aspects of the behaviour, dynamic in nature, of the elements of the system modelled. However, the laws regulating elementary chemical processes are generally unknown, even qualitatively (permissive, additive, co-operative effects etc.) and, even when such laws can be proposed, their kinetic constants are often difficult to determine.

Our knowledge of the networks of interactions between molecules also remains incomplete. Modelling, in the generally understood sense of creating a faithful representation of reality, can therefore only be applied to a few well-delimited and well-described physiological phenomena involving no more than a few tens of genes or proteins. Large-scale observations and treatment failures have clearly demonstrated the complexity of the mechanisms involved: in general more than 100 genes or proteins, together with all the intermediate products and their modifications, are involved in a given cellular physiological function. Research in human physiopathology, one of the main aims of which is to identify new therapeutic targets, must therefore take this complexity into account.

Strategies for the large-scale modelling of human physiopathology

In the absence of detailed information about biological processes, the need to integrate extensive networks has led us to modelling strategies that do not describe processes in terms of elementary chemical reactions. We have developed a “higher-level” modelling strategy, inspired by the logic of signal propagation. Data are effectively available at this level.

Large-scale modelling strategies are now seen as credible only if they include an adaptive or data mining approach. These approaches are designed to obtain knowledge from large amounts of data, structured to various extents, using automatic methods sometimes called learning methods. Generally speaking, they involve the construction of a “flexible” model, in our case a dynamic system.

The available data include data concerning molecular interactions (annotated according to the activating or inhibitory nature of the actions) and high-throughput expression screening data in the form of kinetic information. The network of interactions is formalised as a signed directed graph. Biological signals are propagated within this network not only in one direction, but rather, they may also pass through many feedback loops. Furthermore, the number of connections at each node is highly heterogeneous.

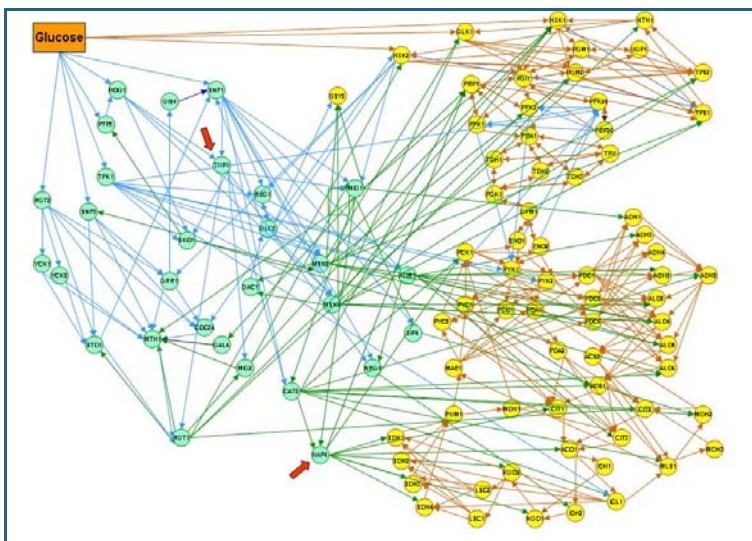


Figure 1: The glucose sensor system

Green circles: genes/proteins propagating the glucose signal; yellow circles: genes/proteins metabolising glucose. Blue arrows: regulation at the protein level; green arrows: transcriptional regulation; orange arrows: metabolic reactions.

The direction of the arrows indicates the direction of signal propagation or metabolic flux. The genes/proteins are identified using their standard name, as in the *Saccharomyces Genome Database*.

The flexibility is reflected by the presence of parameters to be determined by optimisation methods from some of the experimental data: this is the learning step. Other experimental data are used to evaluate the prediction capacity of the model thus obtained: this is the extrapolation step. These approaches aim at the resolution of specific, concrete problems rather than at the mimetic reproduction of the process modelled itself. They are designed more to predict than to explain, they are qualitative rather than quantitative, they focus on the design of decision support tools capable of producing credible hypotheses — hypotheses concerning potential therapeutic targets in our case.

In the absence of precise information about the dynamics of the network, we describe the dynamics of the system by means of a system of differential equations inspired by standard models of signal transduction. This type of model includes a linear degradation term and a “creation” term corresponding to the sigmoid transformation of the total regulatory effect.

The adaptive aspect of our strategy lies in the “learning” of the constants of the model — determination of the set of parameters minimising the distance between the simulated kinetics of the expression of the molecules of the network on the one hand, and temporal experimental data on the other. Kinetic learning does not in itself impose constraints on the dynamic behaviour to be learnt; however, the incomplete nature of the interaction networks and their formalisation, and the insufficient accuracy of expression data suggest that we should limit ourselves to the simplest types of behaviour — steady states, which are the lasting regimes of functioning of living organisms, as opposed to other dynamic regimes such as periodic, quasi periodic and chaotic behaviours. In this context, the kinetics represent transient states that the system learns to follow between stable states.

Under the hypothesis that the network represents the principal mechanisms operating in these transitions and that these stable states provide an exhaustive description of the “geometry” of all stable states, at least in a limited region of the state space, the model parameterised in this manner should be able to classify manipulations (typically inactivation or overexpression of genes) by simulations according to their capacity to direct the system to one of these states, and thus identify the key elements required to induce or to prevent a given type of response.

The glucose sensor system in yeast

We applied this classification strategy to the signal transduction network of the glucose sensor system of the model organism *Saccharomyces cerevisiae* (baker’s and brewing yeast), with the aim of identifying which genes in this network are essential for good growth in the absence of glucose. Yeast cultured in the presence of high extracellular concentrations of glucose break down this glucose, by glycolysis, to generate energy (in the form of ATP) and ethanol (by fermentation).

This consumption of glucose decreases the extracellular glucose concentration, leading to the so-called “diauxic shift”, in which the directions of synthesis and metabolic degradation are modified (and partly reversed) so that the yeast can use the ethanol it has synthesised to produce energy and glucose (gluconeogenesis). This transition is required for growth in the absence of glucose, and allows the yeast to use other carbon sources. It is directed by specific networks for glucose detection, intracellular signalling and metabolism, and the topography of these networks has been extensively described (3). They interact and regulate each other, forming the glucose sensor system (GSS).

We chose to work on yeast for two main reasons. First, data concerning the behaviour of this network during the diauxic shift are freely available for construction of the model. Secondly, qualitative descriptions are available of the adaptation to media devoid of glucose for yeast strains in which the various genes of this network have been inactivated by knock-out techniques (2). Independent test data are therefore available for validation of this strategy.

We analysed published papers to obtain a detailed picture of the GSS (figure 1). This system consists of 97 molecules linked by 353 interactions (glucose is the 98th molecule). The structure of the GSS is clearly heterogeneous in terms of its connectivity, with a large number of feedback loops. This complexity is similar to that of most signal transduction pathways in eukaryotic cells. The sign of the interaction (activation/inhibition) could be determined for 83% of these interactions.

Based on this network and quantitative data describing changes in the abundance of each of these molecules during the diauxic shift in wild-type yeasts (7 successive time points over a period of 690 minutes) (4) (5) (6) (7) (8) (9) (10), we built a dynamic model based on the strategy described above. During the learning phase, the decrease in extracellular glucose concentration alone was used to deduce the dynamics of the whole network. We assumed the stability of the initial state. This hypothesis is supported by the stability of the concentrations of the various molecules in glucose-rich medium reported by DeRisi *et al.* (4).

rich medium, that the model reproduced these results reliably enough to identify the gene for which the expression pattern was initially modified (figure 2).

Molecule identification

We then used the model to identify, by simulation, the molecules (genes/proteins) for which inhibition disturbed the diauxic shift. In other words, we sought to identify the molecules required for good growth in the absence of extracellular glucose (class 1); all other molecules would be considered inessential for this growth (class 2).

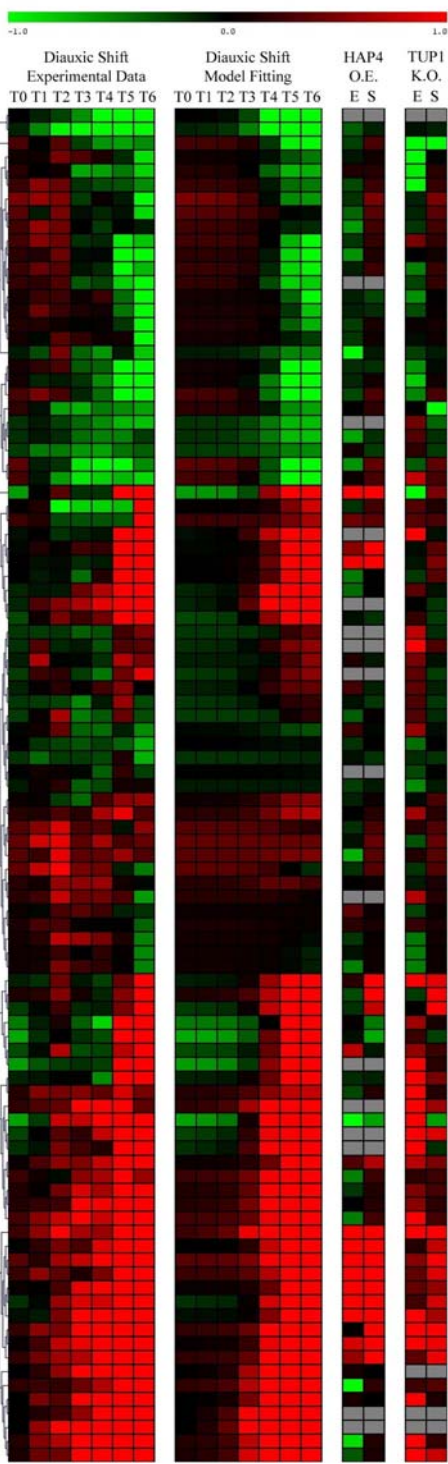


Figure 2: Learning of the glucose sensor system model and the results of simulations

The quality of the learning process can be assessed by comparing the experimental data for the diauxic shift at a series of time points and the modelling of the diauxic shift over the same time period.

HAP4 O.E.: overexpression of *HAP4*; TUP1 K.O.: knock-out of *TUP1*; E: experimental data; S: simulation results. Levels of gene expression/protein activity are expressed as log₂, using a colour scale.

Grey: no experimental data available.

This image was generated by TIGR MeV MultiExperiment Viewer 3.1 from the Institute for Genomic Research (13).

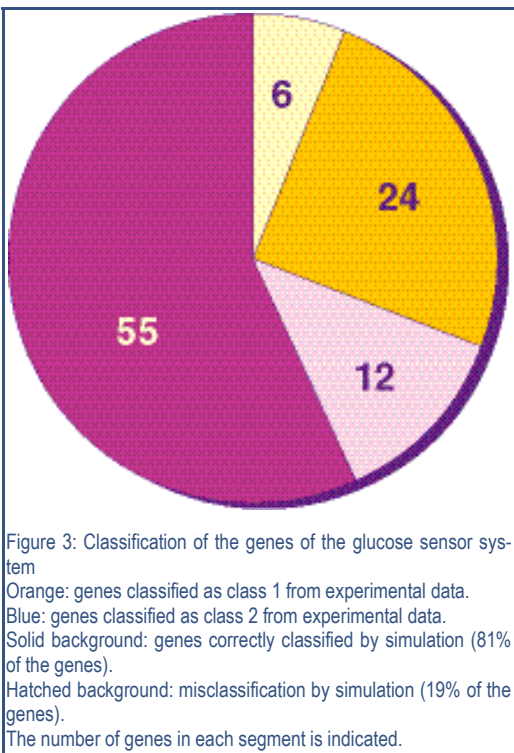
Initial perturbation

To validate the behaviour of the model, we first sought to identify, from the description of a functional state of the network, the initial perturbation responsible for that state. Published data not used in the learning phase describe the effects of *HAP4* overexpression (12) and of the knock-out of five other GSS genes (11) on the yeast transcriptome during growth in the presence of glucose. One of these knock-outs (the *TUP1* gene) and *HAP4* overexpression have significant effects on the GSS. We checked, by simulations in glucose-

Inhibition of the diauxic shift translates mathematically into a tendency of the final simulated state to resemble the initial stable state. This required 97 mutation (gene knock-out) simulations.

Available experimental data (2) indicate that 30 of the genes/proteins of the network belong to class 1. Our simulations classified 36 genes/proteins as class 1 and the other 61 as class 2. Reference data confirmed class 1 status

for 24 of these 36 molecules, but not for the other 12. Two measurements can be used to evaluate a classification method: selectivity, determined here as the proportion of class 1 genes correctly classified as class 1 by the simulations (true positives), and specificity, determined here as the proportion of class 2 genes correctly classified as class 2 by the simulations (true negatives). Our model thus had a selectivity of 80% and a specificity of 82% (figure 3).



Modelling of the Th1 lymphocyte response in man

Based on these promising results with yeast, we applied a similar strategy to the identification of new therapeutic targets for autoimmune diseases.

Cellular immunity is involved in autoimmune diseases, graft rejection, viral infections and immunity to cancers. It is largely dependent on T lymphocytes.

In this type of immunity, T lymphocytes adopt a so-called "Th1" phenotype when they encounter an antigen. We reconstituted a significant fraction of the intracellular signalling network of T lymphocytes, leading to the establishment of a network of 338 molecules and 1498 interactions. We then used expression profiling data describing the development of this Th1 lymphocyte response at seven successive time points over a 48-hour period, to calculate the parameters of the model.

Using the model developed, we carried out a first series of simulations of the inhibition of each molecule of the network, and a second series of simulations of the inhibition of each interaction of the network during the Th1 response (about 2000 simulations). In this way, we were able to identify therapeutic actions likely to inhibit the Th1 response.

The immunosuppressive targets identified included a number that have already been evaluated experimentally in published studies. Based on these targets, we estimated the validation rate for all targets at 70%. These simulations generated a restricted number of new precise, testable therapeutic hypotheses: 14 new potential targets were identified, and the key interactions of these targets, interference with which should produce the therapeutic effect desired, were also individualised (figure 4). The experimental validation of these results is underway.

Mathematical modelling and new therapeutic strategies

We have seen that the creation of models of large-scale biological networks (more than 100 genes/proteins) is possible by adaptive methods, provided that we accept relatively significant theoretical imperfections associated with such models. These imperfections may lead to errors in prediction. Such large-scale models can be improved by the progressive introduction of more faithful representations of biological reality, and this is possible as more accurate and extensive knowledge becomes available (due to more accurate measurement techniques, completion of networks, more precise representations of the phenomena studied including for example any permissive, co-operative or other effects).

However, if we consider such models to be decision support tools, we can already identify effects that may lead to the biological network adopting a state of interest, such as a non-pathological state. This is illustrated by the two examples cited above. Clearly, in both cases, the classifications are imperfect, due largely to the generic formalism used and which has no formal theoretical validity. Nevertheless, pertinent biological targets were identified with a satisfactory yield: selectivity and specificity were between 60 and 80% and thus were much higher than those for any other strategy for identifying therapeutic targets.

Furthermore, in the case of the Th1 lymphocyte response, it would not be possible to identify these potential therapeutic targets by experimental means. Indeed, it is currently impossible to perform experiments that reproduce all the simulations carried out. Gene inactivation technologies (siRNA etc.) remain not very effective in primary cultures of lymphocytes, and the targeted inhibition of interactions is impossible on a large scale. However, the identification of crucial interactions refines hypotheses concerning potential therapeutic targets by identifying the protein domains to be targeted and thereby opens up new avenues for the directed development of therapeutic molecule synthesis. Modelling strategies are the only feasible means of tackling this type of issue.

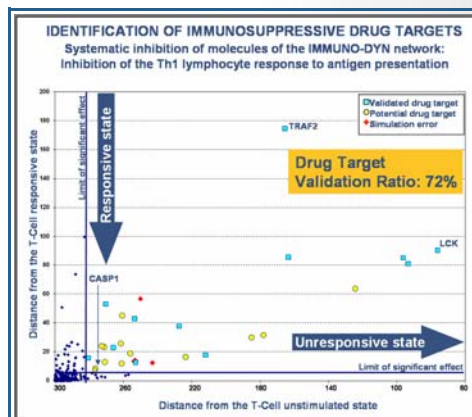


Figure 4: Example of the results of simulation using the model for the Th1 lymphocyte response

Each point represents the simulation of the inactivation of a gene in the network. The cloud of points at the bottom on the left correspond to genes for which inactivation is predicted to have no effect on the Th1 response. The points above and to the right of this cloud correspond to genes for which inactivation is predicted to inhibit the Th1 response.

More generally, this type of approach has potential applications in the exploration of all diseases in which experimentation is difficult or impossible, for example immunological, neurological and psychiatric diseases and, to a certain extent, cancers. We have begun to develop models based on similar strategies for certain cancers and neurodegenerative diseases.

References:

(1) Corr PB (2006) Drug Discovery World 7, 8-14. (2) Steinmetz LM *et al.* (2002) Nature Genetics 31, 400-4 (3) Rolland F *et al.* (2002) FEMS Yeast Research 2, 183-201. (4) DeRisi JL *et al.* (1997) Science 278, 680-6 (5) Portela P, Moreno S (2005) Cellular Signalling 18, 1072-86. (6) Wilson WA *et al.* (1996) Current Biology 6, 1426-34. (7) Kaniak A *et al.* (2004) Eukaryotic Cell 3, 221-31. (8) De Vit MJ *et al.* (1997) Molecular Biology of the Cell 8, 1603-18. (9) Treitel MA *et al.* (1998) Molecular and Cellular Biology 18, 6273-80. (10) Walther K, Schuller HJ (2001) Microbiology 147, 2037-44. (11) Hughes TR *et al.* (2000) Cell 102, 109-26. (12) Lascaris R *et al.* (2003) Genome Biology 4, R3. (13) Saeed AI *et al.* (2003) Biotechniques 34, 374-8