Model of Interactions in Biology and application to heterogeneous network in yeast

Serge Smidtas

Genoscope and CNRS UMR 8030

91057 Evry Cedex, France

E-mail: sergi@sergi5.com

Anastasia Yartseva (auteur pour correspondance)

IBISC - Université d'Evry Val d'Essonne, Tour Evry 2,

523 place des Terrasses de l'Agora, F-91000 Evry, France, and

ISI Foundation, Viale S. Severo 65, I-10133 Torino - Italy

E-mail: aiartsev@lami.univ-evry.fr

Tel +33 1 60 87 38 12, Fax +33 1 44 32 03 19

Vincent Schächter

Genoscope and CNRS UMR 8030

91057 Evry Cedex, France.

E-mail: vs@genoscope.cns.fr

François Képès

Epigenomics Project, and ATelier de Genomique Cognitive, CNRS UMR 8071, Genopole,

523 Terrasses de l'Agora, 91000 Evry, France.

E-mail: Francois.Kepes@genopole.cnrs.fr

June 15, 2006

Abstract

A major challenge for bioinformatics and theoretical biology is to build and analyse a unified model of biological knowledge resulting from high throughput experiment data. Former work analyzed heterogeneous data (protein-protein interactions, genetic regulation, metabolism, synexpression) by modelling them by graphs. These models are unable to represent the qualitative dynamics of the reactions or to model the n-ary interactions. Here, MIB, the Model of Interactions in Biology, a bipartite model of biological networks, is introduced, and its use for topological analysis of the heterogeneous network is presented. Heterogeneous loops and links between synexpression pattern and underlying molecular mechanisms are proposed.

Keywords : Formal model; Biological network; Heterogeneous data

Resumé

Modèle de réseaux d'Interactions Biologiques. Un défi important pour la bioinformatique et la biologie théorique est de construire un modèle unifié qui intégre de nombreuses connaissances biologiques issues notamment d'expériences haut débit, et qui permet leur analyse. Des travaux antérieurs ont analysé des données hétérogènes (interactions protéiques, régulation génétique, métabolisme, synexpression) en les modélisant par des graphes. Toutefois ces modèles ne sont capables, ni de représenter la dynamique qualitative des réactions biochimiques, ni de modéliser les interactions n-aires. Un modèle bipartite des réseaux hétérogènes MIB (Modèle d'Interactions Biologiques), est présenté et illustré par les résultats d'analyses des boucles régulatoires hétérogènes, ainsi que des mécanismes moléculaires sous-jacents à la synexpression des gènes.

Mots-clés : Modèle formel; Réseau biologique; Données hétérogènes

1 Introduction

The last few years have seen the advent of high-throughput technologies to analyze various properties of the transcriptome and proteome of several organisms. The congruency of these different data sources, or lack thereof, can shed light on the mechanisms that govern cellular function. A central challenge

for bioinformatics research is to develop a unified framework for combining the multiple sources of functional genomics information thus obtaining a robust and integrated view of the underlying biological phenomena.

Since the complete DNA sequence of *S. cerevisiae* became available in 1996 [9], a variety of large-scale, high-throughput experimental studies have provided partial, potentially complementary insights into the structure of the yeast regulatory network and, indirectly, into its dynamics.

A major challenge of the post genomic research is to understand how cellular phenomena arise from the interaction of genes, proteins and metabolites. Investigations into the structure of these molecular interaction networks include studies on their global topological properties[33, 32], such as connectivity distribution [1] or scale-free nature [3] have been performed. The local properties such as clustering proteins within the network into functional subnets using combinations of attributes and local connectivity properties to uncover a higher level of network organization [1, 23, 16, 26, 30] were also studied on each homogeneous networks separately.

Several studies [11, 34, 26] have already tried to aggregate many types of data , mostly extending the approach of [24] based on the research of under or over-expressed static graph motifs only in order to understand the topological properties of biological graphs.

In previous work, gene expression data in *Saccharomyces cerevisiae* have already been combined with a Gene Ontology-derived predictions [26] and phenotypic experiments [2]. Recent studies assembled an integrated *S. cerevisiae* network in which nodes represent genes (or their protein products) and differently coloured links represent five types of biological interaction: protein-protein interaction, genetic interaction, transcriptional regulation, sequence homology, and expression correlation [11, 34].

However, most of these studies rely on the graph-theoretic approach which fails to represent n-ary relations between biological objects, for example in metabolic networks or complexes, as well as qualitative dynamics of the interaction, for example, the distinction between activation and inhibition, production and consumption.

In this work, we present a bipartite graph model of heterogeneous biological network that comprises directed transcriptional regulation, protein-protein interaction, the complexes, the metabolic networks, synthetic lethality experiments and micro-array expression results.

This type of models allows searching for complex heterogeneous network motifs with qualitative

dynamics and biologically relevant properties.

Based on this model, the *S. cerevisiae* data set was represented as a global database including the aforementioned data types.

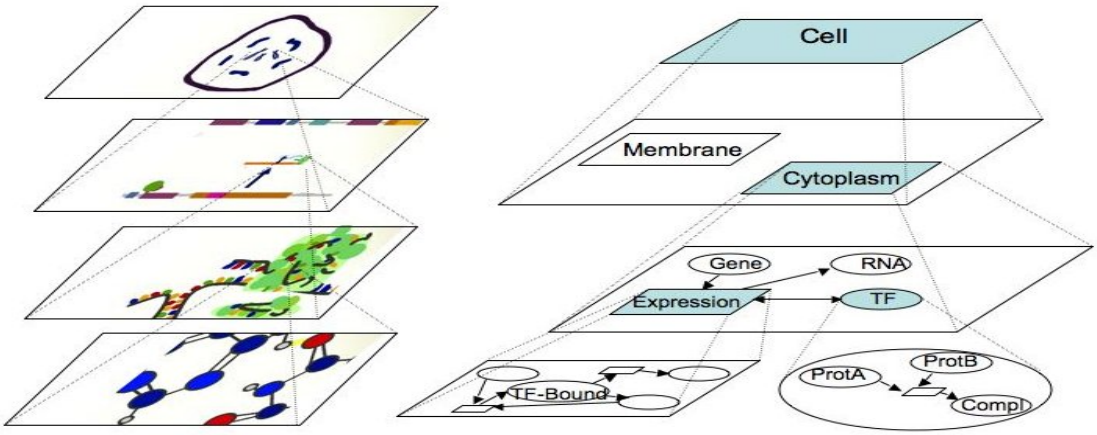

Figure 1: Representation of biological systems seen as a set of chemical reactions. The top layer represents the most general view of the hierarchy. The bottom layer is the most detailed view of the system structure. Two intermediate layers are presented, showing the topological or functional structure of the system. On the left side (top down), the artistic view of a cell with chromosomes is shown, followed by the gene regulatory network scheme, the translation and ribosomal machinery layer and interacting molecules and atoms layer. On the right side (top down) the artistic view of the biological system is modelled in MIB . The first layer box represents the cell that contains membrane and cytoplasm (second layer). Zooming out the cytoplasm (third layer), gene expression, involving a transcriptional factor, is represented. At the bottom layer, the transcriptional factor is magnified into a complex made of two proteins, and gene expression is symbolized by the transient TF/DNA complex.

2 The MIB model

The main model constructing principle that we used is made to apprehend the organization of the complex system that constitutes the cell with its distributed control (see Figure 1). Here we proposed a qualitative modelling framework, Model of Interactions in Biology (MIB), a bipartite graph model of heterogeneous biological network. MIB is designed to fill the gap between, on the one hand, existing techniques for quantitative modelling of biological systems [27, 7, 20, 5], and, on the other hand, techniques for analysis of the network structure mostly based on graph theory [32, 33, 3]. Our approach is largely inspired by the Structured Analysis and Design Technique [22].

A biological system can be seen as an emergent[18] phenomenon of the chemical reactions set, including protein-protein interaction (PPI) and transcriptional regulation interactions (TRI). This set may

be modelled by a composite reactions network and it should satisfy the following constraints:

- to include information about chemical species and chemical reactions of the biological system;
- to consider biological interactions which are not binary, like in the case of a complex of several proteins;
- to distinguish between undirected and directed (positive or negative) interactions of species;
- the representation should be simple enough to allow the study of global structural properties of the network and the search for sub networks in the composite network.

Thus, the set of biochemical reactions composing the biological system is represented in MIB as a network which comprises nodes, either *entities* (chemical species) or *transformations* (chemical reactions), and links between nodes, divided in four *roles*: *consumed, produced, activates, inhibits*. The same chemical species may have different properties and participate in different reactions depending on intracellular localization. In this case such a species may be represented by more than one entity in the MIB model. The next paragraph presents the formal definition of the MIB model.

Definition 1 (MIB model) *The MIB network N is a tuple $(\{X, Y\}, E)$ where*

- *X is a set of* entities *$x = (n, l, t)$ where n is a* name*, l is a* localization*, and t is a* type *of the entity;*
- *Y is a set of* transformations *$y = (n, s, t)$ where n is a* name*, s is a* speed *(kinetic rate) and t is a* type *(e.g., inversible or not, protein-protein or DNA-protein etc.) of the transformation;*
- *E is a set of* links *(x, y, r) or (y, x, r) where $x \in X$ is an entity and $y \in Y$ is a transformation and r is one of four possible* roles *(production, consumption, activation and inhibition) of an entity x in a transformation y.*

Kinetic rates can be dependant of biological context. The above definition does not make any restrictions on it.

The MIB network $(\{X, Y\}, E)$ can be represented graphically as a bipartite graph (as shown in Figure 2) where elliptic nodes represent entities X and rectangular ones represent transformations Y. Nodes are labelled with the attributes of related entities and transformations. Edges of this graph represent links E between an entity and a transformation. There are four arrow types to express four possible roles

of an entity in a transformation: production ($\square \rightarrow \bigcirc$) or consumption ($\bigcirc \rightarrow \square$) of an entity by a transformation and activation ($\bigcirc \leftrightarrow \square$) or inhibition ($\bigcirc \dashv \square$) of a transformation by an entity.

In the following paragraphs, two examples of MIB model of common biochemical reactions will be presented. The first example is catalytic. The second is stoechiometric.

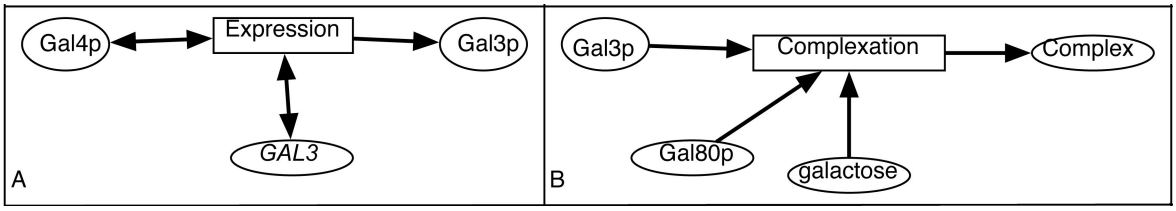

Figure 2: Examples of representation of a biological system. **A.** In yeast, Gal4p is the transcriptional factor that regulates the GAL3 gene. **B.** Gal3p, Gal80p and galactose constitute a complex.

Example 1 - Transcriptional regulation. One of the important properties of the reaction *transcriptional regulation* is that the participating species are not consumed. (This type of reaction can be also called *gene expression regulation*). This type of reaction (the expression of Gal3 protein) is shown in Figure 2 A. The $GAL3$ gene and transcriptional factor Gal4p are needed for the reaction (activates it) but are not consumed [25].

More generally speaking, the *information transfer reaction* represents the production of a biological macromolecule using the informational template (DNA for transcription or RNA for translation reaction). The template is not consumed in such a reaction.

Example 2 - Association reaction. In Figure 2 B the complexation of the Gal3 and the Gal80 proteins and galactose is represented [25]. This is the example of a chemical reaction that can not be represented with a simple graph because it involves three different entities. It may be labelled with the kinetic rate. The association reactions are generally reversible, and the corresponding reverse transformation could also exist and encoded in a distinct reaction.

The *topology* of the MIB or its parts can be described by *motifs*, thus characterizing the number of reactions, species and roles of the species in the system.

Definition 2 (Motif of MIB and its occurence) *A* motif M *on MIB is a tuple* $\{(X_M, Y_M), E_M\}$ *where:*

- X_M *is a set of entities;*
- Y_M *is a set of transformations;*
- E_M *is a set of links between entities and transformations of the motif.*

An occurrence *of a motif M in the MIB model $N = \{(X_N, Y_N), E_N\}$ is a sub network $O = \{(X_O \subset X_N, Y_O \subset Y_N), E_O \subset E_N\}$ and two bijections $B_X : X_O \rightarrow X_M$ and $B_Y : Y_O \rightarrow Y_M$ can be established between nodes of both graphs such that if $x_M = B_X(x_O), l_{x_M} \in l_{x_O}, t_{x_M} \in t_{x_O}$ and $y_M = B_Y(y_O), s_{y_O} \in s_{y_M}, t_{y_O} \in t_{y_M}$ then $\forall (x_M, y_M, r_M) \in E_M \ \exists r'_M : \exists (x_O, y_O, r'_M) \in E_O \wedge \exists (x_M, y_M, r'_M) \in E_M$.*

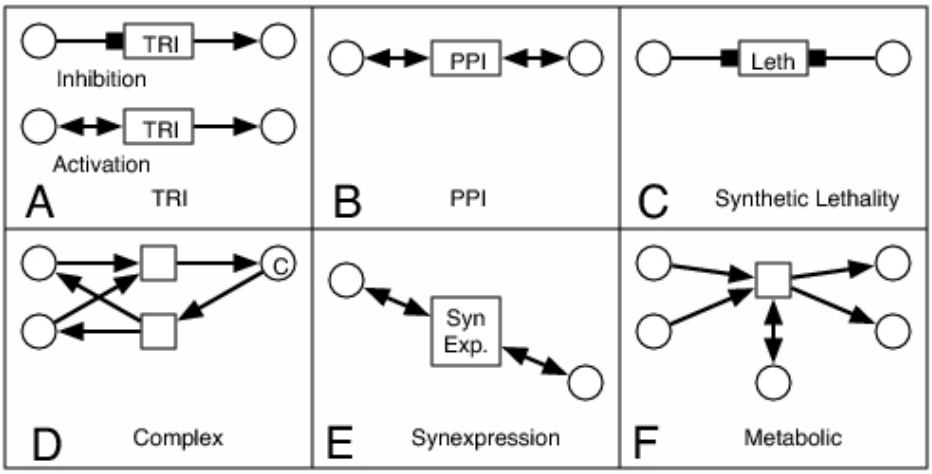

Figure 3: Motifs used for biological data representation in MIB . **A.** Two motifs representing *TRIs*: inhibition (top) and activation (bottom) of the production of the entity (macromolecule) (right) by another entity (Transcription Factor) (left). **B.** A motif representing *physical interaction*: two entities activate a transformation (PPI). **C.** The *synthetic lethality* is represented by a motif with two entities inhibiting a transformation "Leth" (for *lethality phenotype*). **D.** A motif representing *association* transformation (top) which consumes two entities and produces a complex C. The reverse transformation (*dissociation*) is represented in the bottom of the panel. **E.** The *synexpression* of a couple of entities is represented by a motif with two transformations in which they are produced (top) and consumed (bottom) together. **F.** A motif representing a *metabolic reaction*. Two entities are consumed by a transformation, one entity activates it and two entities are produced.

A motif can have several occurrences in the network, in which case they are distinguished by their labels. Figure 3 represents the MIB motifs used to represent every type of biological data included into the database. Motif A illustrates a transcriptional factor that inhibits (or activates) the expression of a protein. Reactions involving two proteins that form a complex were represented by motifs D, and PPIs by motif B. Two more transformations represent indirect and even unknown mecanisms: Synexpression data (correlated expression of a couple of proteins) are represented by motif E, and synthetic lethality by motif C. So long distance and short distance interactions can be mixed during the analysis as we studied for synexpression and it's molecular mechanism (Figure 5).

Finally, a metabolic reaction catalysed by an enzyme is illustrated by motif F where two reactants are consumed, two other molecules are produced and one enzyme is needed by the transformation.

3 Application to heterogeneous network of *S. cerevisiae*

Modelled data, coming from various sources, were integrated in the *Biological Interaction Browser* (BIB) (http://www.genoscope.cns.fr/ biopathways/bib/). We integrated the following data sets: protein-protein interaction (PPI) data, generated using high-throughput variants of the yeast two-hybrid method to identify binary interactions [15, 28] or using techniques to isolate multi-protein complexes based on mass-spectrometry such as HMS-PCI [12], TAP [8] and compilation from the literature [13]. The data include also direct transcriptional interactions (TRI) compiled from the literature [10] and from ChIP-Chip experiments [17]. The synexpression results come from microarrays experiments [31] representing pairs of genes with a correlated expression. The synthetic lethality results [31] represent pairs of yeast genes whose joint disruption is lethal. Finally, the metabolic network data were taken from [14]. The complete network contains 6513 proteins, 1440 complexes, 2 phenotypes. The interactions include 7455 cases of DNA-protein interactions, 8531 protein-protein interactions, 16496 synexpressions, 886 synthetic lethality cases. Feedback loops and synexpression patterns were searched in this entire heterogeneous network.

Feedback Loops Feedback loops are a basic example of a static motif from which dynamical properties such as homeostasis and differentiation can be inferred. The dynamical behaviour of regulatory loops has been studied by several authors using a variety of techniques [5], mostly in the context of transcriptional networks and abstract networks of regulatory influences. Here, we searched for the first time for feedback loops that include both TRI and PPI.

Before studying heterogeneous motifs, TRI-only loops were searched. 108 TRI-only feedback loops were found in the entire network, with lengths ranging from 2 to 10 (see Table I, columns 1 and 2).

Then, one TRI at a time was replaced by a PPI. Figure 4 shows feedback loops each comprising four entities (circles) and the following sets of transformations (squares): TRI only (A), 3 TRIs + 1PPI (B) and 2TRIs + 2 PPIs (C). For example, the motif (B) illustrates a feedback loop made of 4 entities, one PPI and three TRIs. All TRIs are oriented in the same direction and can represent either an activation (double arrows) or an inhibition (squared arrows).

We compared the number of TRI-only loops with the number of loops where a TRI had been replaced by a PPI (Table I, columns 2 and 3). Depending on the loop size, 3-50 times more loops with one PPI were found. If two non-adjacent TRIs are replaced by two PPIs, the number of loops increases up to 3

Loop size	TRIs + 0 PPI	TRIs + 1 PPI	TRIs + 2 PPIs
2	5	17	-
3	4	32	-
4	5	71	125
5	4	144	529
6	9	222	1372
7	6	390	3140
8	12	740	8464
9	22	1197	14863
10	41	1987	30444

Table I: Number of feedback loops as a function of loop size (column 1): loops including only TRIs (column 2), TRIs and one PPI (column 3), TRIs and two PPIs that are not adjacent (column 3).

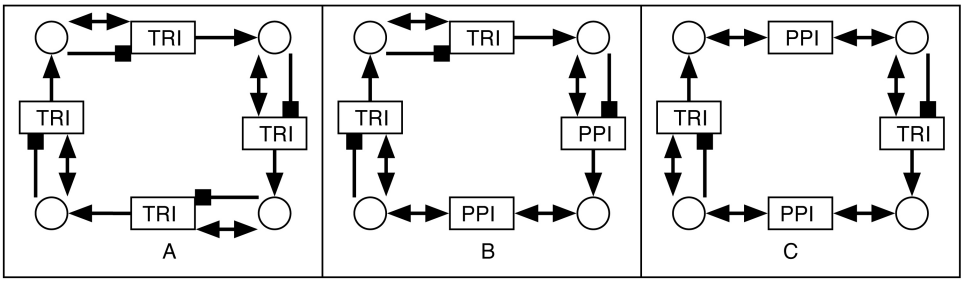

Figure 4: Feedback-loop motifs made of TRIs only (A), with one PPI (B) or with 2 PPIs (C). Each motif contains 4 transformations (rectangular shapes), 4 entities (circles), and possible roles of entities in transformations are represented by arcs.

orders of magnitude, depending on the loop size (Table I, columns 2 and 4). Thus, adding a second PPI in a motif that already included one PPI increases the number of matching subnets from 2 to 15 times.

Micro-arrays Synexpression may involve various underlying molecular mechanisms, thus being a biological result at an intermediate level between molecular physical mechanisms and phenotypes (see Figure 1). To evaluate correlation between the molecular knowledge integrated in the BIB and synexpression data, we searched for possible mechanisms accounting for each synexpressed couple of genes.

We used BIB to find the correlation between the micro-array data on the synexpression of gene pairs, and the biochemical reactions in which these two genes participate. Thus, a molecular mechanism underlying the synexpression of two genes, based on the PPI and TRI graphs, could be proposed. These molecular mechanisms, symbolized by candidate motifs, are presented in Figure 5, together with the number of observed occurrences of each motif type. To determine which motifs are under- or over-represented, the ratio of motif occurrences with and without synexpression was calculated for six candidate mechanisms

(the last column in Figure 5).

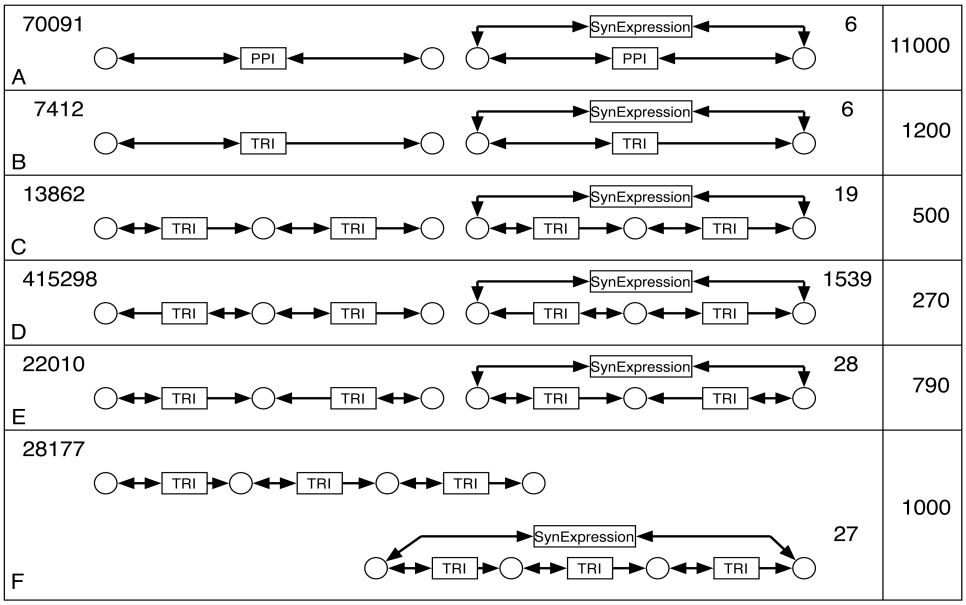

Figure 5: Correlation between synexpression data and underlying biochemical mechanisms. Six motifs were proposed to be candidates for the synexpression mechanisms (A-F, left). For each motif, the number of occurrences in BIB database is indicated on the side. The motifs combining the regulatory mechanism and the synexpression data (A-F, right) were searched, and the number of encountered occurrences of such subnets is indicated. The last column shows the ratio between occurrences of each motif without or with synexpression condition.

We looked for modules comprising one gene that regulates the transcription of another gene (Figure 5, B, left) and where the two genes are synexpressed (Figure 5, B, right). Six occurrences of such a module were found with synexpression, and 7412 occurrences were observed without synexpression, which makes the difference of 1200 times. A more complex motif would include one (Figure 5, C, right) or two (Figure 5, F, right) additional genes between the two initial ones. Such motifs were found 19 and 27 times, respectively, with a ratio of 500 and 1000 times less compared to the same motif without synexpression.

A different candidate motif that accounts for synexpression of two genes could involve a third gene that regulates these two genes (Figure 5, D, right). This motif is found 1539 times in yeast, 270 times less than without synexpression constraint. It is interesting to see that the inverse situation, when two synexpressed genes regulate a third one (Figure 5, E, right) is much less frequent (28 cases, 790 times less than without synexpression). As for the synexpression motif A, it was strongly underrepresented (6 cases, 11000 times underrepresented), meaning that synexpressed genes are seldom participating in a PPI.

For further analysis of the link between synexpression phenotype and the physical interaction network structure , we analyzed the shortest path length distribution between synexpressed genes compared to that

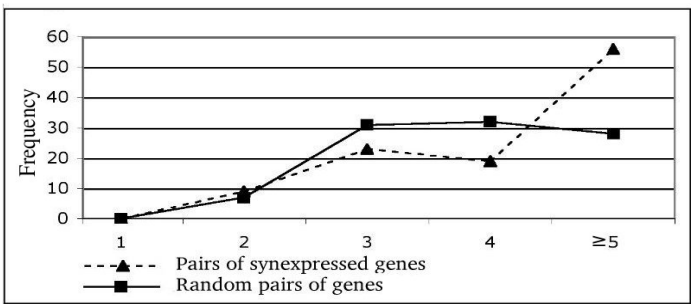

Figure 6: Shortest path length distribution between all synexpressed pairs of proteins (dashed line) versus all possible pairs of proteins (plain line). The shortest paths of length one to four have been searched. The value of five on the x-axis indicates that no shorter path than five has been found.

of any pair of genes. The results are shown in Figure 6. There is little difference between the two distributions, except for long paths ($\geq$5 steps). The average path length between two synexpressed genes is significantly different from that between random pairs of genes for long paths only, in contrast with previous results [2].

4 Discussion

Most studies involving heterogeneous networks thus far have focused either on network topology, either local or global. However, most important biological processes such as signal transduction, cell-fate regulation, transcription and translation involve more than four but much fewer than hundreds of proteins. MIB is slightly more complex than a simple graph representation, but has greater expressiveness. One of the greate advantages of these approach is that this model enables various static and dynamic analysis. It directly represents n-ary relations that are essential for the representation of complexes and of metabolic reactions. The added expressiveness is also related to the assumption that each modelled transformation occurring in the biological system may be broken down into elementary parts [21]. Our model is more abstract than the one proposed in [29], so we can deal with different types of biological objects and processes uniformly. MIB enables the semi-automatic translation in other modelling formalisms such as, for example, Petri Nets, Ordinary Differential Equations, or Pi-calculus (Yartseva et al. in prep.). The BIB tool adapts some of the algorithms available for graphs (e.g. motif search) to the case of bipartite graphs. It can be used to analyse how various data types complement each other in the full heterogeneous network. As most biologically interesting features concern the dynamics of biological functions implemented by molecules, reactions or pathways, biologically meaningful queries are better expressed at the

level of functions and the objects that support these functions. A simple graph representation does not allow this type of query formulation. Figure 7 provides an example of how the MIB formalism allows to search for instances of a function, independently of the precise ”implementation” of this function in a cell. Both subnetworks at the bottom of Figure 7 can fulfill the specified dynamics depicted by the motif at the top. The subnetwork on the left is implemented by TRIs only, and the one on the right by one TRI, one metabolic reaction (transport) and one physical interaction (binding).

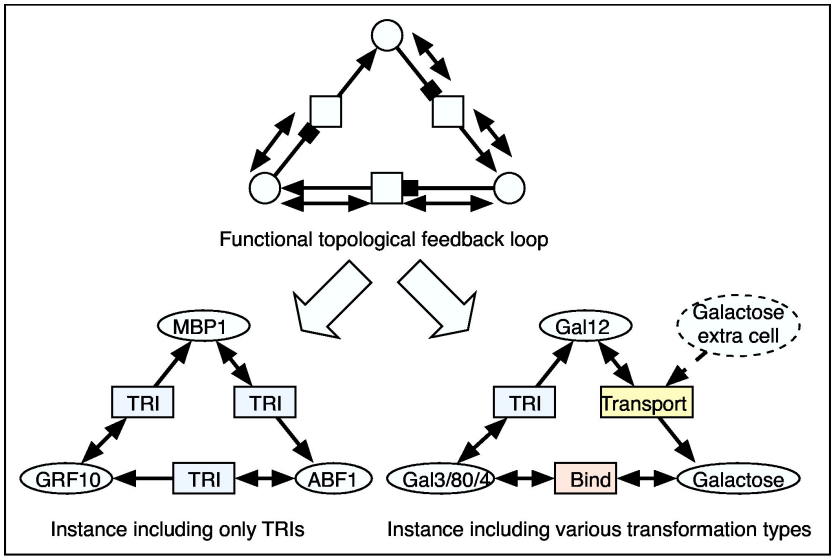

Figure 7: A MIB motif (with a specified dynamics; top) allows searching for both TRI only subnets (left) and mixed TRI/Metabolism/PPI subnets (right). This is illustrated here with a feedback loop.

TRI only feedback loops have already been studied [10]. In the present study, we searched for such loops in larger data sets, and therefore we found more loops in the larger size range. We also provide a new perspective on these feedback loops studies by relaxing previous constraints [24] to allow PPI anywhere in the loops. Some of the modules found are well-known such as the Ste12-Fus3 feedback circuit [4], others are unknown.

The analysis of synexpression data relations between 1625 pairs of genes allowed to propose for each pair a biologically relevant circuit with a parsimonious topology. This result illustrates how an interaction of higher-level order than biochemical reactions may be modelled in MIB, thus enabling the study of the whole set of yeast interactions.

We have found that the paths between synexpressed genes were longer than for random pairs of proteins (see Figure 6). We will further investigate synexpressed gene paths. However, the situation is opposite for transcriptional factors: the paths between pairs of them are shorter than between random pairs

of proteins [19]. This difference could mean that the genes which are not close in the biological interaction network need to be synexpressed in order to synchronize their biological activity. Our explanantion is in line with the results on just-in-time assembly regulation of various complexes [6].

All the interactions integrated in the model come from experimental results but the context in which a given interaction effectively takes place is not known and may vary among experiments. Therefore, the validation step consists in finding the conditions in which the modules are functional, either by calling on an expert, or if prior knowledge is unavailable, by bench experimentation, as has been done in the case of the galactose feedback-loop [25].

These preliminary studies represent a proof of concept for the MIB as a useful tool for future investigations involving regulation, protein interactions, and metabolic networks together with higher level types of interactions like synthetic lethality or synexpression.

5 Acknowledgements

We are grateful to P. Bourgine for discussions. This work was financially supported by CO3 European Project, ISI Foundation, CNRS, Genopole, Genoscope and S. Smidtas.

References

[1] G. Bader and C. Hogue. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics*, 4:2, Jan 2003.

[2] R. Balasubramanian, T. LaFramboise, D. Scholtens, and R. Gentleman. A graph-theoretic approach to testing associations between disparate sources of functional genomics data. *Bioinformatics*, 20(18):3353–62, Dec 2004.

[3] A. Barabasi and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–12, Oct 1999.

[4] L. Bardwell, J. Cook, J. Zhu-Shimoni, D. Voora, and J. Thorner. Differential regulation of transcription: repression by unactivated mitogen-activated protein kinase kss1 requires the dig1 and dig2 proteins. *Proc Natl Acad Sci U S A*, 95(26):15400–5, Dec 1998.

[5] H. de Jong. Modeling and simulation of genetic regulatory systems: a literature review. *J Comput Biol*, 9(1):67–103, 2002.

[6] U. de Lichtenberg, L. Jensen, S. Brunak, and P. Bork. Dynamic complex formation during the yeast cell cycle. *Science*, 307(5710):724–727, 2005.

[7] A. Doi, S. Fujita, H. Matsuno, M. Nagasaki, and S. Miyano. Constructing biological pathway models with hybrid functional Petri nets. *In Silico Biology*, 4(0023), 2004.

[8] A. Gavin, M. Bösche, R. Krause, P. Grandi, M. Marzioch, A. Bauer, J. Schultz, J. Rick, A. Michon, C. Cruciat, M. Remor, C. Höfert, M. Schelder, M. Brajenovic, H. Ruffner, A. Merino, K. Klein, M. Hudak, D. Dickson, T. Rudi, V. Gnau, A. Bauch, S. Bastuck, B. Huhse, C. Leutwein, M. Heurtier, R. Copley, A. Edelmann, E. Querfurth, V. Rybin, G. Drewes, M. Raida, T. Bouwmeester, P. Bork, B. Seraphin, B. Kuster, G. Neubauer, and G. Superti-Furga. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature*, 415(6868):141–7, Jan 2002.

[9] A. Goffeau, B. Barrell, H. Bussey, R. Davis, B. Dujon, H. Feldmann, F. Galibert, J. Hoheisel, C. Jacq, M. Johnston, E. Louis, H. Mewes, Y. Murakami, P. Philippsen, H. Tettelin, and S. Oliver. Life with 6000 genes. *Science*, 274(5287):546, 563–7, Oct 1996.

[10] N. Guelzim, S. Bottani, P. Bourgine, and F. Képès. Topological and causal structure of the yeast transcriptional regulatory network. *Nat Genet*, 31(1):60–3, May 2002.

[11] M. Herrgård and B. Palsson. Untangling the web of functional and physical interactions in yeast. *Journal of Biology*, 4(5), 2005.

[12] Y. Ho, A. Gruhler, A. Heilbut, G. Bader, L. Moore, S. Adams, A. Millar, P. Taylor, K. Bennett, K. Boutilier, L. Yang, C. Wolting, I. Donaldson, S. Schandorff, J. Shewnarane, M. Vo, J. Taggart, M. Goudreault, B. Muskat, C. Alfarano, D. Dewar, Z. Lin, K. Michalickova, A. Willems, H. Sassi, P. Nielsen, K. Rasmussen, J. Andersen, L. Johansen, L. Hansen, H. Jespersen, A. Podtelejnikov, E. Nielsen, J. Crawford, V. Poulsen, B. Sørensen, J. Matthiesen, R. Hendrickson, F. Gleeson, T. Pawson, M. Moran, D. Durocher, M. Mann, C. Hogue, D. Figeys, and M. Tyers. Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature*, 415(6868):180–3, Jan 2002.

[13] http://mips.gsf.de/proj/yeast/catalogues/complexes/.

[14] http://www.biocyc.com.

[15] T. Ito, T. Chiba, R. Ozawa, M. Yoshida, M. Hattori, and Y. Sakaki. A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Natl Acad Sci U S A*, 98(8):4569–74, Apr 2001.

[16] H. Jeong, S. Mason, A. Barabási, and Z. Oltvai. Lethality and centrality in protein networks. *Nature*, 411(6833):41–2, May 2001.

[17] T. Lee, N. Rinaldi, F. Robert, D. Odom, Z. Bar-Joseph, G. Gerber, N. Hannett, C. Harbison, C. Thompson, I. Simon, J. Zeitlinger, E. Jennings, H. Murray, D. Gordon, B. Ren, J. Wyrick, J. Tagne, T. Volkert, E. Fraenkel, D. Gifford, and R. Young. Transcriptional regulatory networks in saccharomyces cerevisiae. *Science*, 298(5594):799–804, Oct 2002.

[18] Y. Louzoun, S. Solomon, H. Atlan, and I. Cohen. The emergence of spatial complexity in the immune system. *Physica A*, 297(1-2):242–252, 2001.

[19] T. Manke, R. Bringas, and M. Vingron. Correlating protein-DNA and protein-protein interactions. *J. Mol. Biol.*, 333(1):75–85, 2003.

[20] H. Matsuno, A. Doi, M. Nagasaki, and S. Miyano. Hybrid petri net representation of gene regulatory network. *Pac Symp Biocomput*, pages 341–52, 2000.

[21] P. Maziere, C. Granier, and F. Molina. A description scheme of biological processes based on elementary bricks of action. *J. Mol. Biol.*, 339(1):77–88, 2004.

[22] D. Ross and A. Schoman. Structured analysis for requirements definition. *IEEE Trans Softw Eng (Special issue on requirements analysis)*, 3(1):6–15, 1977.

[23] B. Schwikowski, P. Uetz, and S. Fields. A network of protein-protein interactions in yeast. *Nat Biotechnol*, 18(12):1257–61, Dec 2000.

[24] S. Shen-Orr, R. Milo, S. Mangan, and U. Alon. Network motifs in the transcriptional regulation network of escherichia coli. *Nat Genet*, 31(1):64–8, May 2002.

[25] S. Smidtas, V. Schächter, and F. Képès. The adaptive filter of the yeast galactose pathway. *J. Theor. Biol. (in press)*, 2005.

[26] B. Snel, P. Bork, and M. Huynen. The identification of functional modules from the genomic association of genes. *Proc Natl Acad Sci U S A*, 99(9):5890–5, Apr 2002.

[27] S. Troncale, D. Campard, J. Guespin, J.-P. Vannier, and F. Tahi. Modelisation of interleukin-6 system in early hematopoiesis with hybrid functional petri nets. In Genopole, editor, *Modélisation de systèmes biologiques complexes dans le contexte de la génomique, Du 4 au 8 avril 2005, Montpellier*, 2005.

[28] P. Uetz, L. Giot, G. Cagney, T. Mansfield, R. Judson, J. Knight, D. Lockshon, V. Narayan, M. Srinivasan, P. Pochart, A. Qureshi-Emili, Y. Li, B. Godwin, D. Conover, T. Kalbfleisch, G. Vijayadamodar, M. Yang, M. Johnston, S. Fields, and J. Rothberg. A comprehensive analysis of protein-protein interactions in saccharomyces cerevisiae. *Nature*, 403(6770):623–7, Feb 2000.

[29] J. van Helden, A. Naim, C. Lemer, R. Mancuso, M. Eldridge, and S. Wodak. From molecular activities and processes to biological function. *Briefings in Bioinformatics*, 2(1):81–93, 2001.

[30] A. Vazquez, A. Flammini, A. Maritan, and A. Vespignani. Global protein function prediction from protein-protein interaction networks. *Nat Biotechnol*, 21(6):697–700, Jun 2003.

[31] C. von Mering, R. Krause, B. Snel, M. Cornell, S. Oliver, S. Fields, and P. Bork. Comparative assessment of large-scale data sets of protein-protein interactions. *Nature*, 417(6887):399–403, May 2002.

[32] A. Wagner. The yeast protein interaction network evolves rapidly and contains few redundant duplicate genes. *Mol Biol Evol*, 18(7):1283–92, Jul 2001.

[33] D. Watts and S. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–2, Jun 1998.

[34] L. Zhang, O. King, S. Wong, D. S. Goldberg, A. Tong, G. Lesage, B. Andrews, H. Bussey, C. Boone, and F. Roth. Motifs, themes and thematic maps of an integrated saccharomyces cerevisiae interaction network. *Journal of Biology*, 4(6), 2005.