World Academy of Science, Engineering and Technology
International Journal of Chemical and Molecular Engineering
Vol:3, No:4, 2009

Modeling Stress-Induced Regulatory Cascades with Artificial Neural Networks

Maria E. Manioudaki and Panayiota Poirazi

Abstract**—**Yeast cells live in a constantly changing environment that requires the continuous adaptation of their genomic program in order to sustain their homeostasis, survive and proliferate. Due to the advancement of high throughput technologies, there is currently a large amount of data such as gene expression, gene deletion and protein-protein interactions for *S. Cerevisiae* under various environmental conditions. Mining these datasets requires efficient computational methods capable of integrating different types of data, identifying inter-relations between different components and inferring functional groups or 'modules' that shape intracellular processes. This study uses computational methods to delineate some of the mechanisms used by yeast cells to respond to environmental changes. The GRAM algorithm is first used to integrate gene expression data and ChIP-chip data in order to find modules of co-expressed and co-regulated genes as well as the transcription factors (TFs) that regulate these modules. Since transcription factors are themselves transcriptionally regulated, a three-layer regulatory cascade consisting of the TF-regulators, the TFs and the regulated modules is subsequently considered. This three-layer cascade is then modeled quantitatively using artificial neural networks (ANNs) where the input layer corresponds to the expression of the up-stream transcription factors (TF-regulators) and the output layer corresponds to the expression of genes within each module. This work shows that (a) the expression of at least 33 genes over time and for different stress conditions is well predicted by the expression of the top layer transcription factors, including cases in which the effect of up-stream regulators is shifted in time and (b) identifies at least 6 novel regulatory interactions that were not previously associated with stress-induced changes in gene expression. These findings suggest that the combination of gene expression and protein-DNA interaction data with artificial neural networks can successfully model biological pathways and capture quantitative dependencies between distant regulators and downstream genes.

Keywords**—**gene modules, artificial neural networks, yeast, stress

M. Manioudaki is with the Department of Chemistry, University of Crete and the Institute of Molecular Biology and Biotechnology (IMBB), Foundation for Research and Technology-Hellas (FORTH), Heraklion, Crete, Greece. Email: mmanioud@imbb.forth.gr

P. Poirazi (corresponding author) is with the Institute of Molecular Biology and Biotechnology (IMBB), Foundation for Research and Technology-Hellas (FORTH), Heraklion, Crete, Greece. Tel: +30 2810 3911 39, Fax: +30 2810 391101, Email: poirazi@imbb.forth.gr

This work was funded by the action 8.3.1 (Reinforcement Pro-gram of Human Research Manpower) of the operational program "competitiveness" of the Greek General Secretariat for Research and Technology (PENED 2003-03EΔ842) and partly by a Marie Curie Outgoing Fellowship (PIOF-GA-2008-219622) of the European Commission.

I. Introduction

IN contrast to the twentieth-century perception in biology, whose basic approach was to isolate each part of a system of interest and characterize the individual members with great detail, current biological research has reached a different era. Nowadays the focus has moved from the individual to the whole, trying to get a maybe blur, but more complete picture of the system under study. The advent of the so-called '-omics' technologies and the development of new computational approaches able to cope with the massive amount of data available, constitute the systems biology approach where the components of a biological system, i.e. the cell, are not studied in isolation but as parts of a large complicated network. The processes inside this network are carried out by the organized interactions of functionally coherent groups of biomolecules, the so-called *functional modules*, entities that are now recognized as the basic structural unit in any biological system [1]. Whether modularity is the effect of natural selection or whether it arose because of biased mutational mechanisms, is still an open question [2], but modularity is ubiquitous and has been observed in many biological networks, including protein-protein interaction [3-5], metabolic [6-9] and transcriptional regulatory networks [9-11].

The concept of modularity simplifies the study of biological systems. The reconstruction of gene networks from gene expression data requires the identification of every interaction between the participating genes. Despite the large number of available data, there is an even larger number of parameters that need to be estimated in complex networks and methods based on Boolean models, Bayesian networks, differential equations or hybrids of those can only be restricted to small subsystems [12]. Parameter optimization is often successful at the level of gene module analysis, where genes in the same regulatory module for example are expected to be co-expressed and thus regulated in a coordinated way by the same set of regulators [11, 13, 14]. The commonality of their properties allows the consideration of each module as a single node in the network. Therefore, inferring the network can be reduced to a problem with two subtasks: identification of the modules and identification of the regulators responsible for the concerted action [12].

Clustering algorithms applied to gene expression data have been extensively used to identify modules of genes that are co-regulated across experimental conditions [15, 16]. Other methods include biclustering algorithms where data from microarray experiments are organized in matrices and genes

can be grouped in potentially overlapping modules [11, 17]. Microarray data have also been modeled as linear mixtures of latent variables [18] while projection methods are used to decompose the data into mutually independent groups and place genes into non-exclusive modules [19, 20].

Module inference provides new insights about the organization of gene activities for different intracellular processes. However, it offers little or no information about the type of regulation (e.g. positive or negative) exerted by various regulators onto the members of a given module under different conditions. This limitation was first confronted by Segal et al [21] who used a probabilistic model on gene expression data to identify the modules, their regulators and the conditions under which the regulation takes place. Since then, several other studies incorporated heterogeneous experimental data to infer the module networks, such as motif information [22-24], chromatin immunoprecipitation (ChiP-chip) [25-27] and protein interaction data [28, 29].

ChiP-chip data provide direct information about the binding (direct physical interaction) of a regulator (transcription factor) to the corresponding gene. Transcription factors regulate the expression of a gene by binding to specific locations at the promoter of the gene and activating or repressing its transcription. Genes that are bound by the same regulators (cause) and display similar expression (result) can be considered as participants in the same biological process and assigned to the same module. Studies that integrate binding and gene expression information result in the identification of module networks, where genes are classified in modules that are regulated by a common set of transcription factors. Several approaches have been developed to infer such networks including MA-networker [26], COGRIM [30], ReMoDiscovery [31], SAMBA [32] and GRAM [33]. The latter is a heuristic method that identifies module networks by first searching for a core set of genes tightly co-regulated using a stringent criterion for the transcription factor binding. The group of genes is then expanded using both a relaxed binding criterion and an expression profile that is similar to the mean expression of the initial group.

Most of the abovementioned studies have been performed on *S. Cerevisiae,* due to the simplicity of its genome and the availability of large amounts of data. A plethora of expression studies are currently available concerning the response of yeast to different environmental changes [34-38], underlying the importance of this mechanism. Particularly for unicellular organisms such as yeast, maintaining their internal homeostasis in a constantly changing surrounding is of vital importance [39, 40]. In response to drastic and stressful changes, yeast cells reorganize their expression program in order to arrest normal cellular processes and activate specific pathways so they can adjust to the novel conditions. During hyperosmotic shock stress, for example, cells arrest normal growth, accumulate glycerol internally by closing Fps1p channel and trigger the high-osmolarity glycerol pathway resulting in regulation of gene expression [41].

The perception of how biological networks contribute to specific intracellular processes will improve further if the regulatory interactions are investigated at more than one level. It is well known that transcription factors are themselves transcriptionally regulated, as part of complicated regulatory cascades that exist in the cell. In this work the regulatory module-hypothesis is expanded into three layer cascades: the level of the gene modules that show the concertized action of genes, the level of the regulators (TFs) of the modules and the level of the regulators of these regulators (TF-regulators). The models capture both the structure and the regulatory function of these cascades using artificial neural networks (ANNs) where the input layer corresponds to the expression of the up-stream transcription factors (TF-regulators) and the output layer corresponds to the expression of genes within each module. Since the expression profile of a transcription factor over time is not necessarily indicative of the timing of its activity - possibly due to various post-transcriptional and post-translational modifications [42, 43]- our models also consider simplified time delays.

II. Methods

A. Data acquisition and preprocessing

Microarray gene expression data from yeast cells in response to different environmental stresses [34] were downloaded from http://genome-www.stanford.edu/yeast_stress/. The dataset comprised of measurements for the expression of 6152 yeast genes for 19 different conditions over several time points (173 experiments in total) as well as over-expression and knockout experiments. The stress conditions included 1) heat shock from 25^0C to 37^0C, 2) heat shock from various temperatures to 37^0C, 3) steady-state temperature growth, 4) temperature shift from 37^0C to 25^0C, 5) mild heat shock at variable osmolarity, 6) response of mutant cells to heat shock, 7) hydrogen peroxide treatment, 8) response of mutant cells to H_2O_2 exposure, 9) menadione exposure, 10) diamide treatment, 11) DTT exposure, 12) hyper-osmotic shock, 13) hypo-osmotic shock, 14) amino acid starvation, 15) nitrogen source depletion, 16) diauxic shift, 17) stationary phase, 18) steady-state growth on alternative carbon sources, 19) steady-state growth at constant temperatures, 20) over-expression studies, 21) knockout experiments for several time points many of which were performed in duplicates or triplicates. For the case of replicated experiments the average expression value was used in our analysis. In an attempt to focus on responses specific to certain stress conditions, over-expression and knockout experiments were excluded and the rest of the experiments were divided in four main categories as shown in Table I.

A dataset containing genome-wide location analysis for the binding of 106 transcriptional regulators to promoter sequences across the genome [44] was downloaded from http://jura.wi.mit.edu/cgi-bin/young_public/navframe.cgi?s=17&f=downloaddata. In the respective study [7], the authors used a myc epitope tag for each transcription factor and performed a genome-wide location analysis using microarrays to detect, through hybridization, those promoter regions of the genome that were enriched in epitope tags after chromatin immunoprecipitation

experiments. Binding data are represented as confidence values (P values) for each microarray spot.

TABLE I
STRESS CONDITIONS ORGANIZED INTO FOUR CATEGORIES

Category A (heat shock)	Category B (starvation)	Category C1	Category C2
Heat shock from 25^0C to 37^0C	Amino acid starvation	Hydrogen peroxide treatment	Hydrogen peroxide treatment
Heat shock from various temperatures to 37^0C	Nitrogen source depletion	Menadione exposure	Menadione exposure
Steady-state temperature growth	Diauxic shift	Diamide treatment	Diamide treatment
Temperature shift from 37^0C to 25^0C	Stationary phase	DTT exposure	DTT exposure
Mild heat shock at variable osmolarity	Steady-state growth on alternative carbon sources		Hyper-osmotic shock
Steady-state growth at constant temperatures			Hypo-osmotic shock

B. Identification of gene modules

For each of the four data categories, the GRAM algorithm [33] was used to infer modules of genes which share a common expression pattern as well as a common set of regulators. The algorithm first uses the protein-DNA binding data to group genes according to their common regulators, with a stringent criterion for this binding. Then, gene modules are refined to contain only those genes with similar expression profiles across all tested conditions. Finally, modules are expanded by the incorporation of genes with a similar expression pattern bound by a relaxed value.

C. Modeling regulatory networks

Since transcription factors (TFs) are nodes in a complex regulatory cascade, they are themselves transcriptionally regulated. For GRAM-inferred gene modules that were regulated by at least two transcription factors, the YEASTRACT database was used (http://www.yeastract.com/) to find proteins that regulate these transcription factors. In addition to bibliographical support, a protein was assumed to transcriptionally regulate the gene coding for a specific transcription factor only if it had at least one binding site upstream the promoter of that gene. Binding site data were retrieved from the SGD database (http://www.yeastgenome.org/).

The above interaction relationships were used to build three layer cascades in which the output layer comprised of a gene in a module derived from GRAM, the middle layer contained the regulators (TFs) of that gene and the top layer the regulators (TFs) of the middle-layer regulators. The structure of these cascades was then used to constrain the connectivity in three-layer Artificial Neural Network models (ANNs). Examples of such networks are shown in Figures 1 and 2 in the *Results* section. The activation functions for all ANN models were sigmoidal-logarithmic in the middle-layer nodes and linear in the output layer.

The ANN models were developed in order to investigate whether the expression of upstream regulators (top-layer TFs) can predict the expression of genes two steps down in the cascade (output-layer). To answer this question, for each ANN model representing a respective biological cascade, the expression profile of the top-layer regulators was used as input to the model and the expression of the down-stream gene was assessed as the output. The middle-layer regulators were only used to constrain the network connectivity and they served as hidden nodes, i.e. their expression was *not* provided to the model. ANN models were developed and trained within the Matlab environment using the Neural Network Toolbox with the back-propagation algorithm. For each model, training was done using 50% of the experimental conditions and the remaining 50% was used for validation and testing. Validation during training was done using 25% of the data in order to avoid over-fitting while the rest 25% was used as a test set. This procedure was repeated 100 times and the training/validation/test data sets were randomly selected for each repetition. The correlation coefficient between the model predictions and the desired output for the test set was estimated for each run. The model's performance was assessed as the average correlation coefficient taken over the test set for 100 runs and was considered good for values higher than 0.70, except for the C1 condition in which the threshold was lowered to 0.65.

To capture the influence of possible time delays between the top-layer regulators and the target genes all time-course experiments in the expression dataset were re-analyzed. 74 networks with poor performance were re-trained using delayed expression data whereby, the expression of the target gene was shifted one or two time steps ahead. The networks' performance was assessed as described previously. The analysis was restricted to a maximum of two time delay steps due to the limited temporal resolution used to sample the yeast expression data.

D. Statistical significance

To assess the statistical significance of the models' prediction accuracy, all ANNs were also trained using randomly shuffled data in the expression profile of the output gene.

III. RESULTS

Using the GRAM algorithm, the genes involved in the four stress categories were grouped into co-expression modules in which the genes were regulated by common transcription factors. For each category, a number of 87-100 modules were identified, each of which was regulated by 1-3 transcription factors. As shown in Table II, the majority of the modules had only one regulator. The number of genes in modules with one regulator was nearly double compared to

modules with two or three regulators, suggesting that genes in larger modules are more likely to be involved in more than one process. Since the goal of this study is to investigate whether non-linear ANNs can quantitatively model a single regulatory cascade, the analysis was limited to the 89 modules with at least two regulators.

TABLE II
STATISTICS OF THE MODULES IDENTIFIED BY GRAM

	Category A	Category B	Category C1	Category C2
Number of modules	99	100	89	87
Number of modules with one regulator	73	76	69	68
Number of modules with >2 & < 10 regulators	26	24	20	19
Average number of genes in modules with one regulator	24.6	22.63	15.84	15.45
Average number of genes in modules with >2 regulator	10.5	11.58	8	8.47

For each transcription factor that regulated these 89 modules the regulators (also TFs) that regulate these transcription factors were identified bibliographically. Modules for which there was insufficient information about upstream regulators (i.e. the TF regulators were unknown or the TFs did not have a binding-site motif in their promoter region) or with a very large number of regulators (>10) per TF were excluded from the analysis. This filtering resulted in 26 modules which were in turn used to built three-layer regulatory cascades. Note that each module corresponded to a single structural cascade as genes in the module were regulated by the same TFs. However, for each gene in each module a different neural network was trained, validated and tested as described in the *Methods* section. A total of 94 ANNs was simulated, corresponding to genes in the 26 modules distributed over all four categories. 18 ANN models had a correlation coefficient $r > 0.7$, among which three had a correlation coefficient $r > 0.8$. For category C1 the threshold was lowered to 0.65 and thus two more ANNs were identified. A representative example for such a network is shown in Figure 1 for gene YLR179C (r=0.723).

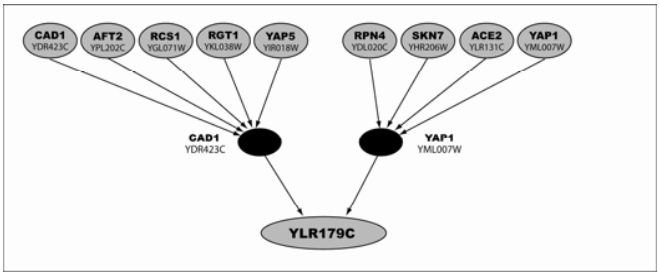

Fig. 1. The three-layer regulatory cascade for gene YLR179C. The structure of the cascade represents the connectivity of the respective ANN model, which had a correlation coefficient of 0.723±0.100.

Since transcriptional regulation is a process that could require a significant amount of time (tens of minutes) before results are seen in the expression of a target gene, the next step was to consider such time-delay cases for the 74 ANNs with poor performance. Specifically, the expression profile of target genes (output-layer) was shifted forward by one or two time-steps compared to the expression of top-layer TFs for each condition and the ANNs were retrained. For a single time-shift, 7 ANN models were identified corresponding to 4 different modules that had $r > 0.7$. For a two-step shift, 6 more ANNs were found corresponding to 3 modules that had $r > 0.65$. A representative example for an one-time step network is shown in Figure 2 for gene YCL059C (r=0.712). It should be noted that the trend in correlation coefficient values was module-dependent, as one would expect since genes within a module have similar expression profiles. Thus, no significant differences in r values were observed between ANNs corresponding to genes belonging to the same module. Overall, for a total of 26 modules tested, all of the 33 genes contained in 19 of these modules were successfully modeled using ANNs achieving correlation coefficients > 0.7 (in categories A, B and C2) or > 0.65 (in category C1).

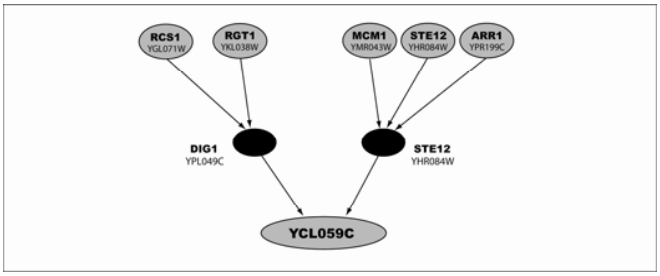

Fig. 2. The three-layer regulatory cascade for gene YCL059C. The respective ANN model incorporates a one-time step shift in the expression of the target gene. The model had a correlation coefficient of 0.712±0.162

Finally, it is important to point out that both of the genes shown in Figures 1 and 2 (and others successfully modeled by ANNs) are regulated directly or/and indirectly by TFs that have been implicated in the yeast stress response. Moreover, the expression of both of these genes has been shown to change in response to environmental stimuli. Specifically, YLR179C is up-regulated in conditions of

oxidative stress [45] while YCL059C is altered in conditions including heat shock [46] and oxidative stress [46, 47].

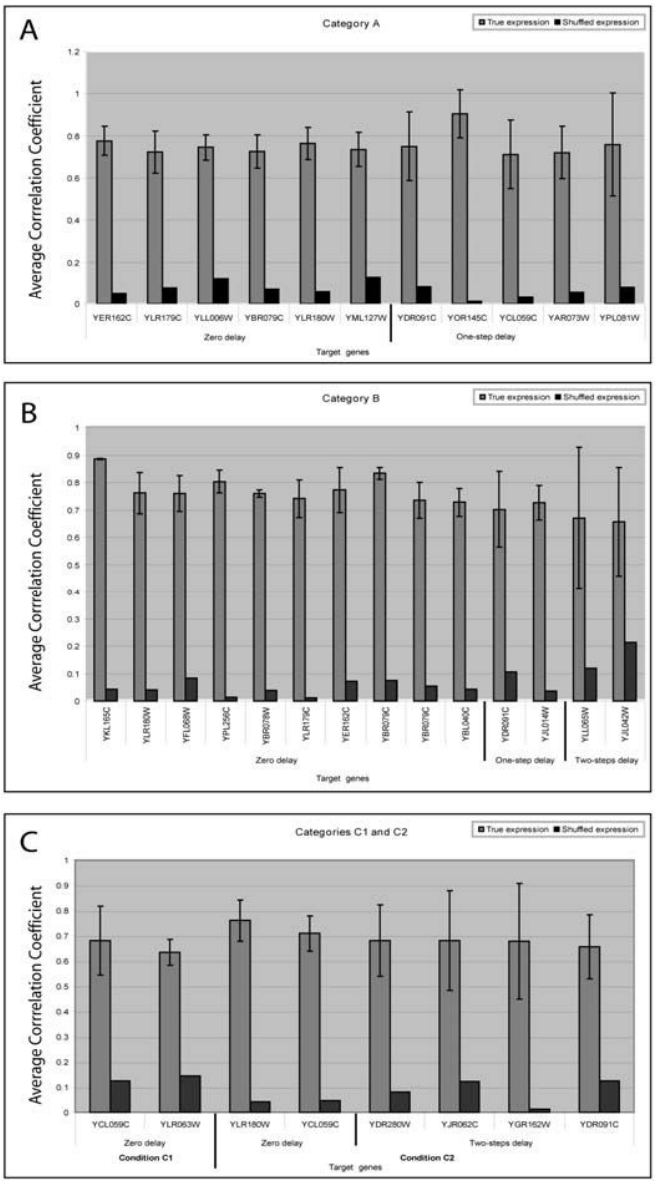

Fig. 3: Average correlation coefficients for the best ANNs using both true (dark grey) and shuffled (black) expression values for each target gene. Error bars correspond to standard deviation. (A) Condition A (B) Condition B, (C) Conditions C1 and C2

To assess the accuracy of our ANN models in predicting the expression profile of downstream genes across numerous conditions, the expression values of the target gene were randomly shuffled and the 33 (20 zero-delay + 7 one-step delay + 6 two-step delay) networks were re-trained. The correlation coefficient for the shuffled data in all networks dropped dramatically, indicating that the performance of the ANN models is far from random chance. Figure 3 illustrates the average correlation coefficients for each of the 33 ANNs organized according to their condition category, for both true and shuffled data.

In addition to the quantitative modeling of 33 three-layer regulatory cascades, the approach used in this study resulted in the identification of a set of 35 transcription factors whose interaction network is illustrated in Figure 4. As evident from the figure, 23 of these TFs have previously been associated with the yeast stress response either directly (experimentally) or indirectly (computationally). 4 transcription factors have been associated with cellular processes that could be influenced by stress conditions, such as the cell cycle while 8 transcription factors have no association with stress or other closely related processes. Given that all of these transcription factors participate in ANN models which can accurately model their effect on downstream genes under stress, it is highly likely that these TFs play a key role in the response of yeast to environmental stress.

IV. Conclusion

In this work a semi-dynamic method was introduced that models the structure of three layer regulatory cascades and predicts quantitatively the expression of genes that are differentially expressed during the stress response in *S. Cerevisiae*. The first step was to identify gene modules in which members share a similar expression profile and a common set of regulators (transcription factors) under four generalized stress categories. Then prior knowledge from existing databases was used to identify upstream transcription factors that regulate the module regulators, thus forming three-layer regulatory cascades. These cascades were modeled using artificial neural networks with constrained connectivity, as dictated by the structure of their internal interactions. The ANNs were trained using as input the expression profile of the top-layer transcription factors and as output the expression of each target gene in a module. The models' ability to predict the expression level of each target gene across numerous time points and different stress conditions was evaluated using a cross validation method.

Out of a total of 94 ANNs tested, 33 models were able to accurately predict the expression of down-stream genes. Among these, 13 ANN models incorporated a delay step since they achieved a significantly higher performance when the expression of the target gene was shifted one (7 models) or two (6 models) time points into the future, compared to the expression of the top-layer transcription factors. The performance of all models was significantly higher than the models' prediction accuracy on shuffled data, further supporting the validity of this approach. Moreover, the results show that this formalism can capture the modularity of the differentially expressed genes since the ANNs prediction accuracy is very similar for genes belonging to the same module.

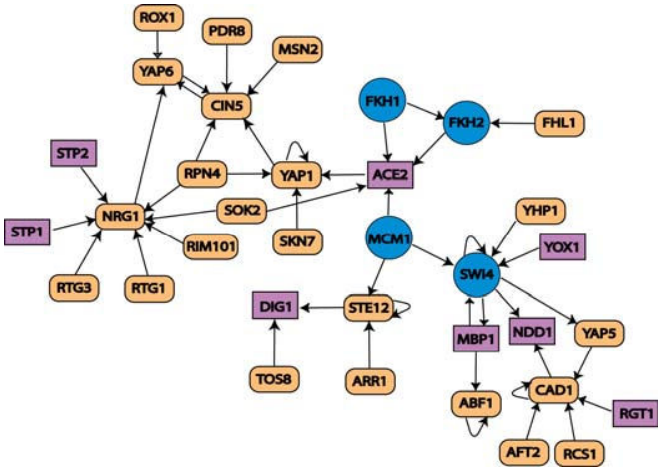

Fig. 4: The network of interactions between all transcription factors involved in the 33 successful ANNs. Arrows indicate the direction of regulation. Orange boxes represent transcription factors that have been implicated in the yeast response to environmental stresses. Cyan boxes represent transcription factors that have been implicated in processes that could be influenced by stress conditions, such as cell cycle and thus they may be involved indirectly to stress response. Grey boxes represent transcription factors which, according to our knowledge have not been implicated in stress conditions.

An important advantage of the method presented here compared to other inference techniques, lies in its semi-dynamic nature. ANN models take into consideration time-series data in order to estimate the expression profile of target genes not only for different stress conditions but also over a time course. This semi-dynamic nature of the models allows the consideration of time delays which are frequently observed in transcriptional cascades. Feed-forward ANNs have been previously used to infer the structure of the transcriptional regulatory network in yeast [50] and find interconnection among clusters of co-regulated genes [51]. Both of these studies focused on inferring the connectivity of regulatory cascades while this work builds on the connectivity map in order to provide quantitative predictions about the expression profile of downstream genes. Finally, the models not only offer advantages over the quantitative prediction of gene expression in stress-activated regulatory cascades, but can also provide new insights regarding the possible role of specific transcription factors that have not been previously associated with stress. Out of the 35 transcription factors that participate in the high performance ANN models, 27 have previously been experimentally or computationally associated with some environmental stress response. It is highly possible that the remaining 6 also play a key role in the response of yeast cells to stressful conditions

A possible explanation for the poor performance of the remaining 61 ANN models could be the over-reliance on bibliographical information for the identification of the top-layer regulators. This information is often inaccurate or incomplete since not all transcription factor-gene interactions have been extensively characterized. This may have resulted in the formation of ANN models whose structure -and thus regulatory action- is only partially correct. An additional possibility is that the effect of top-layer regulators on the expression of the target gene is differentially shifted in time, whereby different regulators are associated with different time delays. Such a feature will be investigated in future efforts.

In conclusion, this work provides a new multi-step approach for modeling both the structure and the effects on gene expression in stress-induced regulatory cascades in *S. Cerevisiae.* The proposed method could easily be applied in other organisms and other intracellular processes where the expression profile of genes over time and under different conditions is of interest.

ACKNOWLEDGMENT

The authors would like to thank George Thireos, Dimitris Tzamarias, Popi Syntichaki and George Garinis for their help in designing the simulation experiments. We would also like to thank members of our lab for useful discussions during the preparation of this manuscript.

REFERENCES

1. Hartwell, L.H., et al., *From molecular to modular cell biology.* Nature, 1999. **402**(6761 Suppl): p. C47-52.
2. Wagner, G.P., M. Pavlicev, and J.M. Cheverud, *The road to modularity.* Nat Rev Genet, 2007. **8**(12): p. 921-31.
3. Fernandez, A., *Molecular basis for evolving modularity in the yeast protein interaction network.* PLoS Comput Biol, 2007. **3**(11): p. e226.
4. Gursoy, A., O. Keskin, and R. Nussinov, *Topological properties of protein interaction networks from a structural perspective.* Biochem Soc Trans, 2008. **36**(Pt 6): p. 1398-403.
5. Han, J.D., *Understanding biological functions through molecular networks.* Cell Res, 2008. **18**(2): p. 224-37.
6. Ravasz, E., et al., *Hierarchical organization of modularity in metabolic networks.* Science, 2002. **297**(5586): p. 1551-5.
7. Zhao, J., et al., *Modular co-evolution of metabolic networks.* BMC Bioinformatics, 2007. **8**: p. 311.
8. Segre, D., et al., *Modular epistasis in yeast metabolism.* Nat Genet, 2005. **37**(1): p. 77-83.
9. Bruggeman, F.J., J.L. Snoep, and H.V. Westerhoff, *Control, responses and modularity of cellular regulatory networks: a control analysis perspective.* IET Syst Biol, 2008. **2**(6): p. 397-410.
10. Tanay, A., et al., *Revealing modularity and organization in the yeast molecular network by integrated analysis of highly heterogeneous genomewide data.* Proc Natl Acad Sci U S A, 2004. **101**(9): p. 2981-6.
11. Ihmels, J., et al., *Revealing modular organization in the yeast transcriptional network.* Nat Genet, 2002. **31**(4): p. 370-7.
12. Tim Van den Bulcke, K.L., Yves Van de Peer, Kathleen Marchal, *Inferring transcriptional networks by mining 'omics' data.* Current Bioinformatics, 2006. **1**.
13. Eisen, M.B., et al., *Cluster analysis and display of genome-wide expression patterns.* Proc Natl Acad Sci U S A, 1998. **95**(25): p. 14863-8.
14. Niehrs, C. and N. Pollet, *Synexpression groups in eukaryotes.* Nature, 1999. **402**(6761): p. 483-7.
15. Zhao, Y. and G. Karypis, *Data clustering in life sciences.* Mol Biotechnol, 2005. **31**(1): p. 55-80.
16. Kerr, G., et al., *Techniques for clustering gene expression data.* Comput Biol Med, 2008. **38**(3): p. 283-93.

17. Wei, G.H., D.P. Liu, and C.C. Liang, *Charting gene regulatory networks: strategies, challenges and perspectives.* Biochem J, 2004. **381**(Pt 1): p. 1-12.
18. Li, H. and M. Zhan, *Unraveling transcriptional regulatory programs by integrative analysis of microarray and transcription factor binding data.* Bioinformatics, 2008. **24**(17): p. 1874-80.
19. Alter, O., P.O. Brown, and D. Botstein, *Singular value decomposition for genome-wide expression data processing and modeling.* Proc Natl Acad Sci U S A, 2000. **97**(18): p. 10101-6.
20. Lee, S.I. and S. Batzoglou, *Application of independent component analysis to microarrays.* Genome Biol, 2003. **4**(11): p. R76.
21. Segal, E., et al., *Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data.* Nat Genet, 2003. **34**(2): p. 166-76.
22. Lee, H.G., et al., *High-resolution analysis of condition-specific regulatory modules in Saccharomyces cerevisiae.* Genome Biol, 2008. **9**: p. R2.
23. Hu, J., H. Hu, and X. Li, *MOPAT: a graph-based method to predict recurrent cis-regulatory modules from known motifs.* Nucleic Acids Res, 2008. **36**(13): p. 4488-97.
24. Kundaje, A., et al., *Combining sequence and time series expression data to learn transcriptional modules.* IEEE/ACM Trans Comput Biol Bioinform, 2005. **2**(3): p. 194-202.
25. Imoto, S., et al., *Combining microarrays and biological knowledge for estimating gene networks via bayesian networks.* J Bioinform Comput Biol, 2004. **2**(1): p. 77-98.
26. Gao, F., B.C. Foat, and H.J. Bussemaker, *Defining transcriptional networks through integrative modeling of mRNA expression and transcription factor binding data.* BMC Bioinformatics, 2004. **5**: p. 31.
27. Xu, X., L. Wang, and D. Ding, *Learning module networks from genome-wide location and expression data.* FEBS Lett, 2004. **578**(3): p. 297-304.
28. Maraziotis, I.A., K. Dimitrakopoulou, and A. Bezerianos, *An in silico method for detecting overlapping functional modules from composite biological networks.* BMC Syst Biol, 2008. **2**: p. 93.
29. Tornow, S. and H.W. Mewes, *Functional modules by relating protein interaction networks and gene expression.* Nucleic Acids Res, 2003. **31**(21): p. 6283-9.
30. Chen, G., S.T. Jensen, and C.J. Stoeckert, Jr., *Clustering of genes into regulons using integrated modeling-COGRIM.* Genome Biol, 2007. **8**(1): p. R4.
31. Lemmens, K., et al., *Inferring transcriptional modules from ChIP-chip, motif and microarray data.* Genome Biol, 2006. **7**(5): p. R37.
32. Li, H. and W. Wang, *Dissecting the transcription networks of a cell using computational genomics.* Curr Opin Genet Dev, 2003. **13**(6): p. 611-6.
33. Bar-Joseph, Z., et al., *Computational discovery of gene modules and regulatory networks.* Nat Biotechnol, 2003. **21**(11): p. 1337-42.
34. Gasch, A.P., et al., *Genomic expression programs in the response of yeast cells to environmental changes.* Mol Biol Cell, 2000. **11**(12): p. 4241-57.
35. Causton, H.C., et al., *Remodeling of yeast genome expression in response to environmental changes.* Mol Biol Cell, 2001. **12**(2): p. 323-37.
36. Bammert, G.F. and J.M. Fostel, *Genome-wide expression patterns in Saccharomyces cerevisiae: comparison of drug treatments and genetic alterations affecting biosynthesis of ergosterol.* Antimicrob Agents Chemother, 2000. **44**(5): p. 1255-65.
37. Kwast, K.E., et al., *Genomic analyses of anaerobically induced genes in Saccharomyces cerevisiae: functional roles of Rox1 and other factors in mediating the anoxic response.* J Bacteriol, 2002. **184**(1): p. 250-65.
38. Rep, M., et al., *The transcriptional response of Saccharomyces cerevisiae to osmotic shock. Hot1p and Msn2p/Msn4p are required for the induction of subsets of high osmolarity glycerol pathway-dependent genes.* J Biol Chem, 2000. **275**(12): p. 8290-300.
39. Gasch, A.P. and M. Werner-Washburne, *The genomics of yeast responses to environmental stress and starvation.* Funct Integr Genomics, 2002. **2**(4-5): p. 181-92.
40. Attfield, P.V., *Stress tolerance: the key to effective strains of industrial baker's yeast.* Nat Biotechnol, 1997. **15**(13): p. 1351-7.
41. Mager, W.H. and M. Siderius, *Novel insights into the osmotic stress response of yeast.* FEMS Yeast Res, 2002. **2**(3): p. 251-7.
42. Tootle, T.L. and I. Rebay, *Post-translational modifications influence transcription factor activity: a view from the ETS superfamily.* Bioessays, 2005. **27**(3): p. 285-98.
43. Kel, A., et al., *Beyond microarrays: Finding key transcription factors controlling signal transduction pathways.* BMC Bioinformatics, 2006. **7 Suppl 2**: p. S13.
44. Lee, T.I., et al., *Transcriptional regulatory networks in Saccharomyces cerevisiae.* Science, 2002. **298**(5594): p. 799-804.
45. Shenton, D., et al., *Global translational responses to oxidative stress impact upon multiple levels of protein synthesis.* J Biol Chem, 2006. **281**(39): p. 29011-21.
46. Aragon, A.D., et al., *Release of extraction-resistant mRNA in stationary phase Saccharomyces cerevisiae produces a massive increase in transcript abundance in response to stress.* Genome Biol, 2006. **7**(2): p. R9.
47. Molina-Navarro, M.M., et al., *Comprehensive transcriptional analysis of the oxidative response in yeast.* J Biol Chem, 2008. **283**(26): p. 17908-18.
48. Qian, N. and T.J. Sejnowski, *Predicting the secondary structure of globular proteins using neural network models.* J Mol Biol, 1988. **202**(4): p. 865-84.
49. Bendtsen, J.D., et al., *Improved prediction of signal peptides: SignalP 3.0.* J Mol Biol, 2004. **340**(4): p. 783-95.
50. Hart, C.E., E. Mjolsness, and B.J. Wold, *Connectivity in the yeast cell cycle transcription network: inferences from neural networks.* PLoS Comput Biol, 2006. **2**(12): p. e169.
51. Huang, J., H. Shimizu, and S. Shioya, *Clustering gene expression pattern and extracting relationship in gene network based on artificial neural networks.* J Biosci Bioeng, 2003. **96**(5): p. 421-8.