2011 International Conference on Bioscience, Biochemistry and Bioinformatics
IPCBEE vol.5 (2011) © (2011) IACSIT Press, Singapore

On the Relation between Structure and Biological Function in Transcriptional Networks and ncRNA-Mediated Interactions

J.C. Nacher
Department of Complex And Intelligent Systems
Future University Hakodate,
Hokkaido, Hakodate, Japan
nacher@fun.ac.jp

N. Araki
Graduate School of Systems Information Science
Future University-Hakodate,
Hokkaido, Hakodate, Japan
g2109002@fun.ac.jp

***Abstract*— We investigate the distribution of molecular functionalities in transcriptional networks and non coding RNA-mediated interactions by using complex network methods and optimization algorithms. When network nodes belong to two distinct functional classes, the networked structure is analyzed by means of two parameters. These parameters define a phase diagram whose boundary is computed using optimization methods. The results reveal that molecular functions occupy specific locations in the networked structure as well as in the phase diagram. The findings show that both transcriptional and ncRNA-interactions exhibit characteristic couplings between network and functionality that deviate from random expectations, supporting growing evidences for a ncRNA regulatory capability at a system level.**

Keywords-Non-coding RNA, transcriptional networks, complex networks, optimization algorithms

I. INTRODUCTION

It is of common knowledge that genetic information flows from DNA to proteins by means of mRNAs molecules. As a consequence, it is possible to consider a one-to-one correspondence between genes and proteins. While this affirmation seems to be accurate for simple prokaryotes, recent studies have observed that the proportion of protein-coding genes decreases as a function of developmental complexity [1]. In particular, the ratio of non-protein coding genes rises to 98.5% in humans. The finding that the transcription of non-coding RNA (ncRNA) in higher organism is so abundant raises the question on its cellular functionality and suggests that current knowledge on genetic information processing might be largely incomplete [1-3]. If the latter hypothesis was correct, it could imply that RNA molecules were able to evolve and to adapt to transcriptional programs of higher eukaryotes.

NcRNA represents a functional RNA molecule that is not translated into a protein. Non-coding RNA comprises introns in protein-coding genes as well as other transcripts that do not seem to encode proteins. Some classifications include transfer RNA (tRNA) and ribosomal RNA (rRNA), as well as snoRNAs, microRNAs, siRNAs and piRNAs as non-coding RNA genes [3].

In this work, we investigate the coupling between networked structure and functionality in transcriptional networks and ncNRA-mediated interactions. As a model organism we used *S. cerevisiae* whose transcriptional and non-coding mediated interactions are available. We first collected data for the transcriptional regulatory interactions defined by transcriptional factors (TFs) that regulate target genes [4]. Functional interactions between non-coding RNAs (ncRNAs) and proteins related bio-macrolecules (PRMs) were also collected for *S. cerevisiae* organism and the ncRNA-mediated network was constructed [5].

By using information about functional classes, the coupling between network structure and functional distribution could be investigated. Classification information places each molecule into one of several possible functional classes, depending on whether it plays (1) or does not (2) some specific molecular function. Two parameters, *dyadicity* D and *heterophilicity* H, can measure the distribution of node functionalities in the network as shown in [6, 7]. Recently, this framework was successfully used to investigate the functional characteristics of mobile services and protein interaction networks [6]. Our study provides answers to questions about whether TFs and ncRNA with similar functions have higher probability to link genes with similar functionality. The results show that both transcriptional and ncRNA-interactions deviate from random couplings between network and functionality supporting growing evidences for ncRNA regulatory capability.

II. NETWORK CONSTRUCTION AND FUNCTIONAL CLASSES

The datasets for constructing the transcriptional regulatory network and the ncRNA-protein interactions were downloaded from databases. In particular, the data for the transcriptional network was obtained from the Gerstein Laboratory [8] and the ncRNA dataset from NPInter Database [5]. We define these networks as bipartite graphs, where each link represents a functional interaction between a TF (ncRNA) and a gene target or protein-related molecule. The dataset contained 121 TFs and 896 genes for transcriptional network. On the other hand for ncRNA-mediated interactions we collected 25 ncRNAs and 105 proteins.

The functional classes were obtained from different databases. NONCODE and MIPS Functional Catalogue Databases were used to download functional information for functional classes of TFs-genes and ncRNAs-proteins [9].

We then identified classes and subclasses common to both systems as shown in Table I and II.

III. DYAD-RELATED PARAMETERS

A. Statistical functions

Each molecule (node) can be characterized by a functional class that takes only two values 1 or 0. Number of molecules with a specific function 1 (0) in the cell are called n_1 (n_0). Therefore, the total number of nodes will be denoted by $N=n_1+n_0$. The number of links with its two end nodes 1-1 or 1-0 and denoted by m_{11} and m_{10}, respectively. If function 1 is distributed randomly in a network of size N, the expected values of m_{11} and m_{10} read as:

$$\overline{m_{11}} = \binom{n_1}{2} \times p = \frac{n_1(n_1-1)}{2} p \qquad (1)$$

$$\overline{m_{10}} = \binom{n_1}{1}\binom{n_0}{1} \times p = n_1(N-n_1)p \qquad (2)$$

with $p=2M/N(N-1)$ and M is the total number of interactions (links) in the network [10]. Computations of m_{11} and m_{10} in real networks highlight deviations from the randomly expected distributions giving insights into the coupling between structure and functionality.

B. Dyadicity and heterophilicity

Deviations from random values can be measured by means of the *dyadicity* D and *heterophilicity* parameters H [6]:

$$D = \frac{m_{11}}{\overline{m_{11}}} \quad \text{and} \quad H = \frac{m_{10}}{\overline{m_{10}}} \qquad (3)$$

High values of D parameter indicates a network where nodes with a given function 1 are connected more strongly among themselves than expected by random distribution of functionalities. D >1 shows a dyadic behavior. In contrast, D <1 indicates an anti-dyadic coupling. Equivalently, we can consider the case when nodes with a given property 1 have a low density of links connecting nodes with function 0 than expected by random chance. In such a case, H > 1 shows a heterophilic behavior and H < 1 indicates a heterophobic configuration. Phase diagrams of possible values m_{11} and m_{10} (D, H) can be plotted showing the possible configurations for a given set of functionalities and network structures. Values for (D, H) cannot be arbitrary and are constrained by networked structure. Some values can admit distinct network configurations. This is known as degeneracy factor. (See [6] for further details).

IV. RESULTS

By using the functional classes information collected from databases, we assigned a given function 1 to each node, depending on whether it plays (1) or nodes not (0) that specific function. We applied this method to both transcriptional regulatory networks and ncRNA-mediated interactions for the same organism *S. cerevisiae*. We then computed the parameters (D, H) and each network and each available functional class. Finally, the results obtained for both networks were compared. As an example, Fig. 1a shows the ncRNA-mediated network with the mapping of functional classes considered in this study (see Table I).

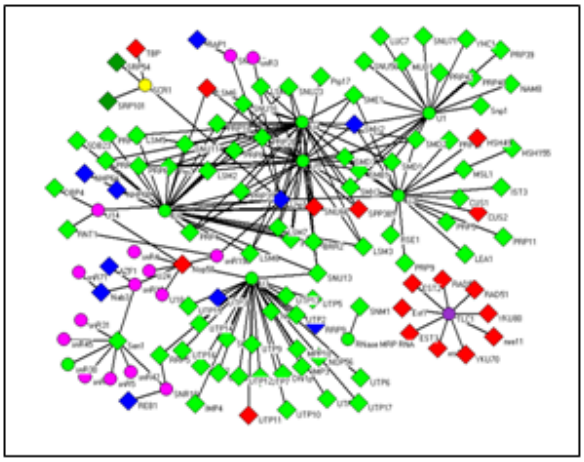

Figure 1. (a) ncRNA-protein network. Each color denotes a functional class

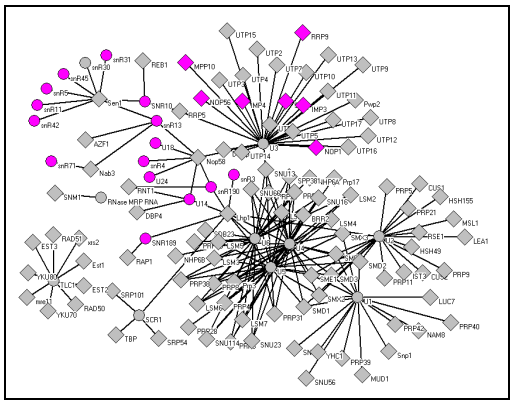

Figure 1. (b) RNA modification functional class

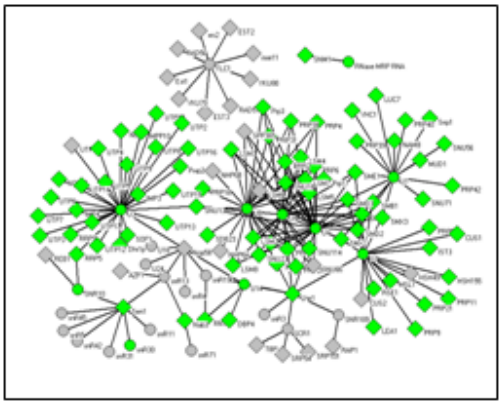

Figure 2. (c) RNA processing functional class

A. *Transcriptional regulatory network*

Among all the possible configurations, the results show that the parameters D and H exhibit a defined heterophobic behavior. Only one functional class shows heterophilic and dyadic pattern. (See Fig. 2 and Table II). The result for a randomly distributed pattern of functional classes in the network is also shown in Fig. 2.

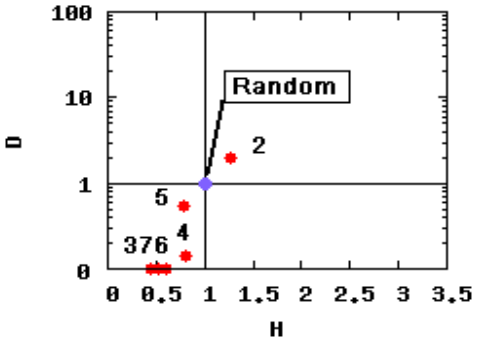

Figure 3. Distribution D-H for TFs-gene network

B. *NcRNA-mediated interactions*

The analysis of ncRNA-mediated interactions revealed that all analyzed functional are heterophobic (see Fig. 3 and Table I). This finding is similar to the one observed for transcriptional networks. These classes included DNA transcription, RNA modification, RNA processing, RNA translation, RNA translocation and protein transport. In addition, RNA processing and protein transport exhibit a dyadic structure. Furthermore, it is worth noticing that similar heterophobic patterns were observed for the analyzed of protein-protein interaction network of S. cerevisiae [6]. For protein interactions all classes were heterophobic with only one antidyadic exception. The corresponding random configuration is shown in Fig. 3, highlighting the statistical significant deviations of the coupling between structure and functionality in the ncRNA network. Network configurations with specific RNA modification and processing functionalities are shown in Fig. 1b-c, respectively.

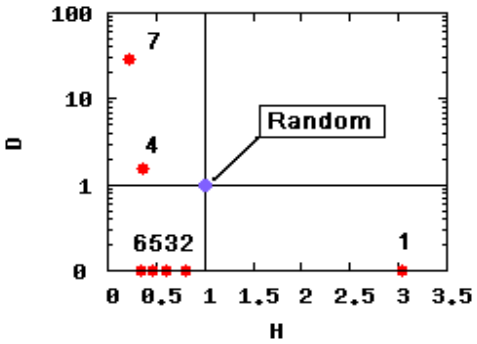

Figure 4. Distribution D-H for ncRNA-protein network

C. *Phase boundaries analysis for networks*

Although it is possible to construct the phase diagram for (D, H) parameter using exhaustive computation of all possible ways of placing n_1 nodes in a network, it is certainly more efficient to use heuristic algorithms to compute these phase diagrams for larger networks. Here we show the results obtained by computing the phase diagram (D, H) using Ant Colony Optimization (ACO) [11] and a version of Fiduccia-Mattheyses (F-M) algorithm [12]. The results shown in Fig. 4 indicates that the computation of both heuristic algorithm leads to similar results (both diagrams show similar shapes), however, in most cases F-M algorithm was able to find faster and more efficiently maximum and minimum possible values for certain configurations, identifying a larger area for the phase space. We briefly summarize ACO implementation as follows: (1) Network configuration is initialized with all nodes set to n_0. The network can be represented by means of its adjacency matrix. A pheromone matrix is also defined using a complete graph with edges set to 0.001. (2) Each ant randomly selects one node and the selected node changes its membership to n_1. Then, ants also select an adjacent n_0 node using pheromone preferential edge selection. Step (2) is iterated until n_1 desired number of nodes is generated. The process is repeated for each ant. (3) For all ants each m_{11} and m_{10} are computed. Best (maximum or minimum) values are stored. (4) Pheromone matrix is updated. Elitist and pheromone decays were also considered. Steps 1 to 4 are iterated 100 times.

We also implemented a version of F-M algorithm (FM1) and compared with the one used in [6] (FM2). We briefly describe our implementation as follows: The network is configured with a required number of n_1 nodes. (1) For each n_1 node, we select n_0 nodes and permute its membership. (2) Configurations m_{11} and m_{10} are computed before and after the permutation and the change is evaluated. (3) The pair of nodes n_0-n_1 with optimal change in m_{11} and m_{10} is selected and its membership is changed. (4) Using this configuration the global m_{11} and m_{10} values for the whole network are computed and maximum or minimum values are updated. (5) Steps (1-4) are repeated for the next n_1 node ($n_0 \cdot n_1$ times). The process (1-5) is iterated 100 times. The results showing FM1 implementations are displayed in Fig. 5a-b for RNA

modification functional class for ncRNA and TFs networks, respectively. Figure shows similar phase boundaries although both networks have different sizes.

TABLE I. FUNCTIONAL CATEGORIES FOR NCRNA-PROTEIN NETWORK

ID	ncRNA-protein network			
	Functional category(out, in node)	*D*	*H*	
1	DNA_stability (1, 0)	…	3.0311	Hilic
2	DNA_transcription (0, 10)	…	0.7965	Hbic
3	RNA_modification (15, 7)	…	0.6034	Hbic
4	RNA_processing (10, 81)	1.5278	0.3672	D, Hbic
5	RNA_translation (0, 6)	…	0.4670	Hbic
6	RNA_translocation (0, 2)	…	0.3394	Hbic
7	Protein _transport (1, 2)	28.9637	0.2281	D, Hbic
Dyadic: D, Anti-dyadic: Ad, Heterophobic: Hbic, Heterophilic: Hilic				

TABLE II. FUNCTIONAL CATEGORIES FOR TRANSCRIPTIONAL NETWORK

ID	TFs-gene network			
	Functional category(out, in node)	*D*	*H*	
1	DNA_stability (0, 0)	…	...	
2	DNA_transcription (120, 323)	2.0032	1.2581	D, Hbic
3	RNA_modification (0, 27)	…	0.4416	Hbic
4	RNA_processing (4, 192)	0.1425	0.7975	Ad, Hbic
5	RNA_translation (1, 295)	0.5546	0.7786	Ad, Hbic
6	RNA_translocation (1, 130)	0.0000	0.6050	Ad, Hbic
7	Protein _transport (0, 67)	…	0.5278	D, Hbic
Dyadic: D, Anti-dyadic: Ad, Heterophobic: Hbic, Heterophilic: Hilic				

V. DISCUSSION AND CONCLUSION

Several findings have been derived from this analysis. First, we were able to explore for the first time the coupling between structure and functionality in both transcriptional and ncRNA-protein interactions networks. Functional classes analyzed were classified as heterophobic, showing that most ncRNA molecules with a given function 1 tend to have more links with proteins with function 0 than expected by random. Similar heterophobic structures were found in the analysis for transcriptional networks. TFs also tend to have more links with genes with distinct or multiple functionalities than expected by random chance. This suggests a similar link in the coupling between network structure and functionality in transcriptional and ncRNA-mediated interactions. Second, both transcriptional and ncRNA networks show distributions of functionalities that significantly deviate from their randomly expected values. The finding that ncRNA structures are not random raises the question about its possible functionality. The present study supports, in our view, growing evidences about the regulatory capability of ncRNA molecule [1,13]. Recent works show that ncRNAs molecules could be implicated in processes related to regulation, cell differentiation and tumorigenesis. Some studies have also linked ncRNA molecules to other complex diseases like coronary disorders and diabetes [2]. Correlations between microRNA repression and protein interaction using expression data have also been reported [14]. Our node characteristics analysis supports earlier evidences about its possible regulatory functionality and the hypothesis that non-coding regulation could explain the huge ratio of non-coding to protein coding DNA observed in higher eukaryotes. This analysis provides new insights into the non-coding interactions and offers knowledge about the coupling between its structural organization and functionality.

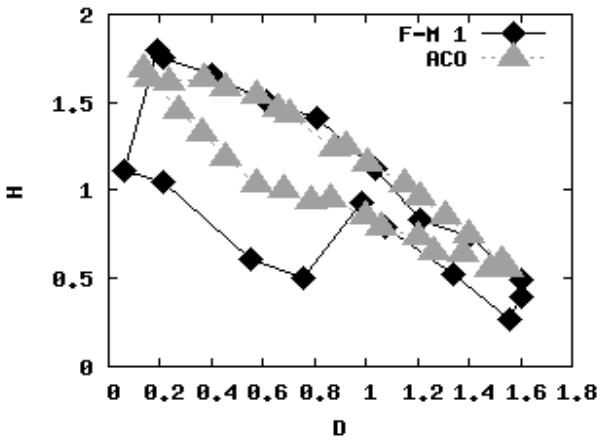

Figure 4. Phase boundaries for the ncRNA-protein network corresponding to the RNA processing functional class (n1=91). Results are shown for ACO and FM1 algorithms.

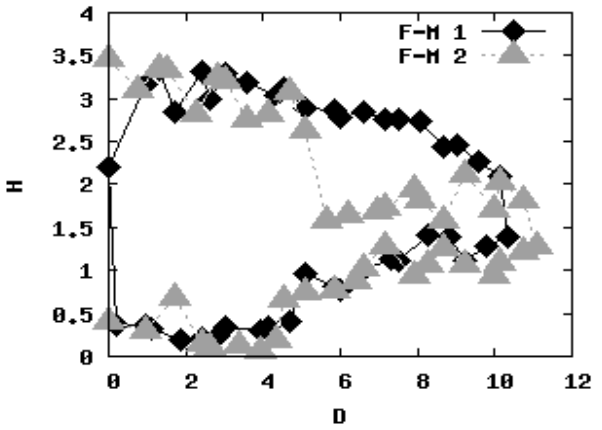

Figure 5. (a) Phase boundaries for the ncRNA-protein network corresponding to the RNA modification functional class (n1=22). Results for FM1 and FM2 algorithms.

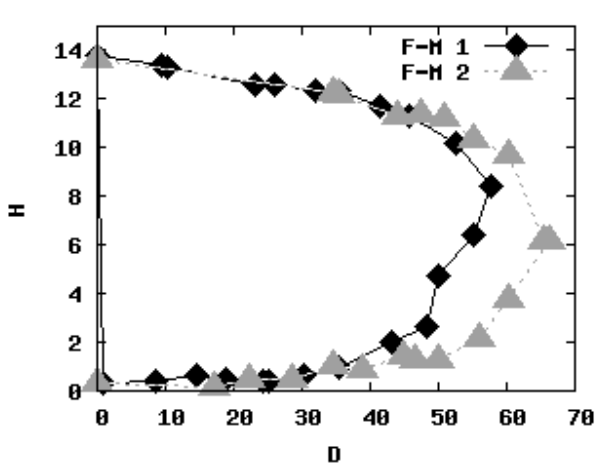

Figure 5. (b) Phase boundaries for the TFs-gene network corresponding to the RNA modification functional class (n1=27). Results for FM1 and FM2 algorithms.

ACKNOWLEDGMENT

The authors thank J. Park for fruitful discussions and suggestions.

REFERENCES

[1] J. S. Mattick, "RNA regulation: a new genetics", Nature Review Genetics, 5, pp. 316-323, 2004.

[2] P. P.Amaral, M. E. Dinger, T. R. Mercer and J. S. Mattick, " The eukaryotic geneome as an RNA machine", Science 319, pp. 1787-1789, 2008.

[3] J. S. Mattick and I. V. Makunin, "Small regulatory RNAs in mammals", Human Molecular Genetics, 14, pp. R121-R132, 2005.

[4] T. I. Lee, N. J. Rinaldi, F. Robert, D. T. Odom, Z. Bar-Joseph et al., "Transcriptional regulatory networks in Saccharomyces cerevisiae", Science 298, pp. 799-804, 2002.

[5] T. Wu, J. Wang, C. Liu, Y. Zhang and B. Shi, "NPInter: the noncoding RNAs and protein related biomacromolecules interaction database", Nucleic Acids Research 34, pp. D150-D152, 2006.

[6] J. Park and A.-L. Barabasi, "Distribution of node characeristics", Proc. Natl. Acad. Sci. USA 99, pp. 17916-17920, 2007.

[7] S. Wasserman and K. Faust, Social network analysis, Cambridge University Press: New York, 2007.

[8] N. M. Luscombe, M. M. Babu, H. Yu, M. Snyder, S. A. Teichmann et al.,"Genomic Analysis of regulatory network dynamic reveals large topological changes", Nature 431, pp. 308-312, 2004.

[9] MIPS Classification: http://mips.helmotz-muenchen.de/funcatDB/

[10] R. D. Alba, "Graph-theoretic definition of a sociometric clique", J. Math. Soc., 3, pp. 113-126, 1973.

[11] E. Bonabeau, M. Dorigo, G. Theraulaz, Swarm Intelligence: From natural to artificial systems, Oxford university Press: New York, 1999.

[12] C.M. Fiduccia and R.M. Mattheyses, "A linear-time heuristic for improving network partitions", Proc. of the 19th Conference on Design Automation (IEEE, Los Alamitos, CA), 1982, pp. 171-181, doi:0420-0098/82.

[13] Y. Shimoni, G. Friedlander, G. Hetzroni, G. Niv et al., "Regulation of gene expression by small non-coding RNAs: a quantitative view", Molecular Systems Biology, 3, pp.1-9 (2007)

[14] H. Liang and W.-H. Li, "MicroRNA regulation of human protein-protein interaction network", RNA 13, pp. 1402-1408 (2007).