Evolution thinks modular

Alessandro Vespignani

Groups of interacting proteins define functional modules that govern a cell's activity. A new study suggests that specific interaction motifs and their constituents are highly conserved across species, identifying potential functional modules used in the evolutionary process.

Network analysis is increasingly recognized as a powerful approach to understanding biological organization and the function of cellular components and may also help us to understand the principles driving the evolution of living organisms. Indeed, most genes and proteins do not have a function on their own; rather, their role is realized through a complex web of interactions with other proteins, genes and biomolecules.

This perspective, largely fostered by the recent abundance of high-throughput experiments and the identification of entire genome sequences and gene coexpression patterns, has led to an intensification of research focusing on the architecture of biological networks. On page 176, Wuchty *et al.*[1] move the analysis of the network framework to a deeper level by offering quantitative evidence for the existence of topological motifs of interacting proteins that are conserved to a high degree during evolution. These results provide the first hint that evolution preserves modules that define specific biological functions.

Network science

Large complex networks arise in a vast number of natural and artificial systems[2,3]. Ecosystems consist of species whose interdependency can be mapped in intricate food webs. Social systems are best represented by diagrams describing various interactions among individuals. The Internet and the World Wide Web (WWW) are prototypical examples of self-organized networks emerging in the technological world. The living cell is no exception; its organization and function are the outcome of a complex web of interactions among genes, proteins and other molecules.

In the late 1990s large network maps began being developed in different disciplines. Mapping of the WWW and the physical Internet offered the first chance to study the topology of large complex networks. Other maps followed, describing many networks of practical interest in social science, critical infrastructures and biology. In the latter case, novel experimental techniques allowed collection of data on metabolic pathways, gene coexpression patterns and protein interactions in a given proteome (**Fig. 1**), making possible the systematic study of a variety of cellular networks[4,5].

The analysis of these maps has led to a better understanding of the architecture of these networks. Indeed, measurements pointed out that the topology of complex networks is far from random, as was previously believed. Particularly important was the realization that many networks are characterized by the statistical abundance of 'hubs', or nodes with a large number of connections to other elements.

The presence of hubs implies a certain degree of hierarchy, as most networks have a core group of highly connected elements that are generally more relevant in the system architecture. Another important feature of hierarchical networks is the presence of discrete groups of interconnected elements, referred to as topological modules or motifs. Modules can be repeated at different hierarchical levels and interconnected by the hubs of the system. How the hierarchical and modular structure of networks can be reconciled with the other properties of networks is one of the exciting fields of research in network analysis and is producing interesting results[2,3,6].

The building blocks of evolution

Biological networks regulating the cell's activity possess the same topological fea-

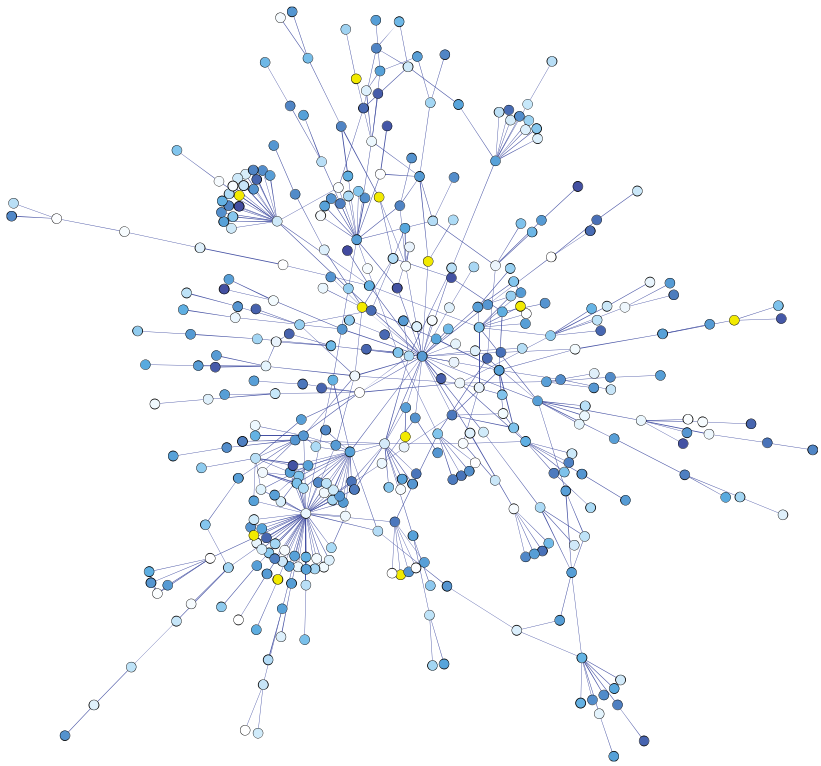


Figure 1 A protein-protein interaction network. Image courtesy of Gary Bader (Memorial Sloan-Kettering Cancer Center).

Alessandro Vespignani is at the Laboratoire de Physique Theorique, Batiment 210, Universite de Paris-Sud, 91405 ORSAY Cedex, France.
e-mail: Alessandro.Vespignani@th.u-psud.fr

tures of complex networks. Moreover, biological networks can be broken down into groups of interacting molecules or modules, each responsible for different cellular functions[7,8]. This is evident in the protein-protein interaction network of the yeast, *Saccharomyces cerevisiae*, obtained from two-hybrid experiments, which has a clear hierarchical and modular architecture.

In their new study, Wuchty *et al.*[1] obtain quantitative evidence that proteins belonging to specific topological motifs in the protein interaction networks seem to be highly conserved across species during evolution. Interacting proteins can be assigned to defined motifs, such as triangles, squares and pentagons, which describe potential functional modules. By identifying the list of conserved proteins across five eukaryotes, the authors find that proteins belonging to topological motifs in the protein interaction networks of *S. cerevisiae* are conserved with higher probability than those not present in such motifs. Furthermore, the more internal connections in the motif, the more highly conserved it is. These results suggest that conserved proteins are more likely to be parts of topological motifs and that these motifs could identify the functional modules used as building blocks in the evolutionary process. This conclusion is further supported by the authors' observation that specific motifs are more likely to be associated with certain cellular functions.

The results of Wuchty *et al.*[1] should be confirmed by investigating conservation of motifs in higher organisms. Such experiments would require new high-throughput techniques capable of reconstructing the protein interaction networks of higher organisms. Although this issue and the concerns regarding the reliability of two-hybrid experiments require further consideration, the work of Wuchty *et al.*[1] represents the first step in connecting the topological architecture of biological networks with its evolution and function. This work should have an immediate impact on evolutionary models aimed at identifying the structure of the protein interaction network based on gene duplication-divergence mechanisms[9,10]. Furthermore, the relevance of specific motifs and local configuration analysis in identifying cellular function modules may enrich local and global algorithms for protein function assignment based on the protein interaction network[11,12].

With this paper, network analysis seems to have entered a new stage in which the general theoretical framework developed so far is beginning to provide answers to specific and detailed questions about complex biological processes.

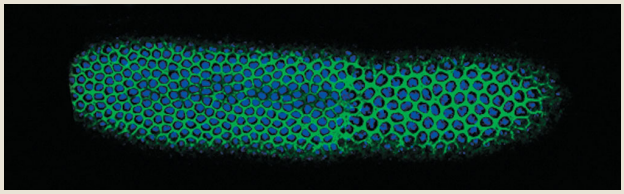
1. Wuchty, S. *et al. Nat. Genet.* **35**, 176–179 (2003).
2. Barabasi, A.-L. *Linked* (Perseus, Cambridge, Massachusetts, 2002).
3. Dorogovtsev, S.N. & Mendes, J.F.F. *Evolution of Networks* (Oxford University Press, Oxford, 2003).
4. Hodgman, T.C. *Bioinformatics* **16**, 10–15 (2000).
5. Oltvai, Z.N. & Barabasi, A.-L. *Science* **298**, 763–764 (2002).
6. Ravasz, E. *et al. Science* **297**, 1551–1555 (2002).
7. Hartwell, L.H. *et al. Nature* **402**, C47–C52 (1999).
8. Milo, R. *et al. Science* **298**, 824–827 (2002).
9. Vazquez, A. *et al. Complexus* **1**, 38–42 (2003).
10. Pastor-Satorras, R. *et al. J. Theor. Biol.* **222**, 199–210 (2003).
11. Hishigaki, H. *et al. Yeast* **18**, 523–531 (2001).
12. Vazquez, A. *et al. Nat. Biotechnol.* **21**, 697–700 (2003).

Frühstart at the midblastula transition

In 1982, Newport and Kirschner defined the midblastula transition (MBT), a key developmental event in *Xenopus laevis* and other animal embryos. The *X. laevis* embryo undergoes 12 rapid divisions after fertilization, with the thirteenth division initiating a new developmental program characterized by zygotic gene transcription and blastomere motility. Only fitful progress has been made in defining the molecules that regulate the MBT, but Jörg Grosshans and colleagues have now identified a protein that may have an important role (*Dev. Cell* **5**, 285–294; 2003).

Previous work showed that although the pre-MBT *X. laevis* embryo is fully competent to initiate zygotic transcription, the rapid cell cycles of the first 12 cleavages don't allow enough time for such transcription to occur. Several groups have contributed to a model proposing that the increasing nucleocytoplasmic ratio in the blastomeres of the early embryo depletes factors that are required for DNA replication or cell cycle progression, thus lengthening interphase and allowing time for zygotic transcription. Such a model has also been suggested for the equivalent of the MBT in the *Drosophila melanogaster* embryo, in which zygotic transcription accompanies a checkpoint-dependent elongation of cell cycles 11–13. The fly embryo, which consists of shared cytoplasm up to this point, then pauses at cleavage cell cycle 14 and cellularizes. A central question that remains is the nature of the additional factors that link nucleocytoplasmic ratio with control of the cell cycle.

Grosshans *et al.* show that the cytoplasmic protein frühstart (meaning 'false start') has all the expected properties of a linchpin



of the MBT in the fly embryo. Frühstart (frs) was originally identified as a mitotic inhibitor that delays division in cells of the ventral furrow during gastrulation. Noting that *frs* is first expressed coincident with the pause in the mitosis at cycle 14, the authors asked whether it might also delay mitosis at the MBT. As pictured here, injection of *frs* mRNA into the posterior end of the embryo (at right) during cycles 10–12 results in large patches where there are fewer cell divisions. Taking a closer look at *frs*-null embryos, they observed that a small fraction of them have patches of higher nuclear density, which the authors suggest is due to an extra cleavage before cellularization.

Is expression of *frs* regulated by the nucleocytoplasmic ratio? Expression of the gene, which peaks shortly after cycle 13, is delayed in haploid embryos. This and other lines of evidence lead the authors to argue that *frs* transcription is probably a direct readout of this ratio. Future work will focus on the identification of *cis*-acting regulatory elements that control expression of *frs* in response to the amount of DNA in the embryo as a whole.

Alan Packer