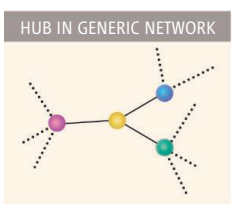
BIOCHEMISTRY

Proteins in a Small World

Layering three-dimensional structural information on top of protein interaction networks reveals two different kinds of highly connected proteins or "hubs."

Todd O. Yeates and Morgan Beeby

Networks of interacting components have been put forward as models for understanding systems as diverse as food webs (*1*), the topology of the Internet, the social ties of guests at a cocktail party, and the collaboration networks of hip-hop acts in popular music (*2*). In the area of systems biology, networks of interacting proteins have been explored as models for understanding cellular processes (*3*–*6*). Many of these networks exhibit a "small world" property (*7*), meaning that a connection can be established between any two elements of the network by following only a small number of links. The small world property of some networks is due in part to the presence of hubs, which are elements connected to many other elements in the network. In protein interaction networks, hubs play an important role, and their properties have been investigated (*8*). On page 1938 of this issue, Kim *et al.* (*9*) add a new chapter to the analysis of hubs in protein interaction networks, clarifying some previously murky issues about how they operate in the cell and how they may have evolved.



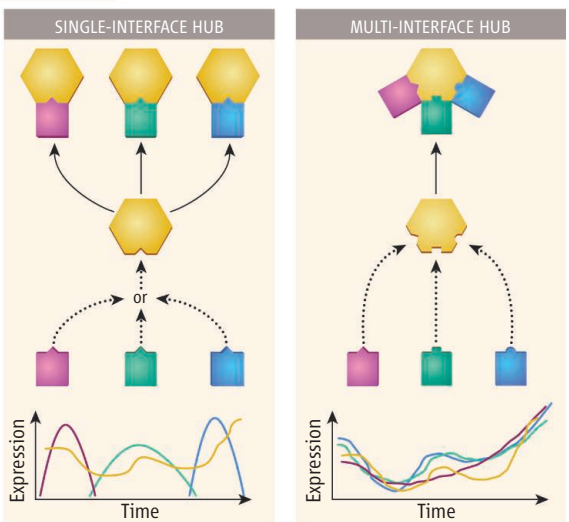
In some protein networks, a link can denote a variety of different kinds of relationships between two proteins: direct physical interaction, correlated expression of proteins in the cell, performance of successive steps in a metabolic pathway, and so on. Kim *et al.* focus on direct physical interactions between proteins, but they take a further step by using the database of known three-dimensional structures to make inferences about what parts of the protein surfaces are involved in the various interactions.

This simple but elegant advance adds considerable information to the network view. With the added information it becomes possible to determine which of the multiple interactions or connections that are made to a given protein can occur simultaneously, and which are mutually exclusive due to overlapping binding surfaces. The thrust of their work is that they are able to distinguish between two different kinds of hubs in protein interaction networks (see the figure). One type, referred to as a single-interface hub, makes interactions to numerous other proteins using just one binding surface; these interactions are mutually exclusive. The other type, a multi-interface hub, makes simultaneous interactions to multiple other proteins using multiple distinct binding surfaces. The distinction between the different kinds of hubs makes it possible to ask new questions and to see new features in protein networks. It also provides a way to begin to factor out spatial versus temporal complexities in networks, which are otherwise convoluted in generic protein interaction networks.

After discriminating between the two kinds of protein hubs, the authors are able to clarify a number of issues that had been examined previously. There has been some disagreement over whether hub proteins evolve at a slower rate than more peripheral proteins (*10, 11*). Kim *et al.* clarify the topic by showing that the rate of mutation of a hub is constrained by the amount of the protein surface involved in interactions with other proteins, not simply by the number of proteins with which the hub interacts. Thus, multi-interface hubs evolve more slowly than typical cellular proteins, whereas single-interface hubs generally do not. They also note that, compared to single-interface hubs, multi-interface hubs show more highly correlated cellular expression with their interacting partners. This parallels the recent observation of a division of hubs into "date" hubs, which tend to be coexpressed with only one binding partner at a time, and "party" hubs that are co-transcribed alongside multiple partners (*12*). The interacting partners of multi-interface hubs are more often involved in stable multisubunit complexes, whereas interactions to single-interface hubs are apparently more transient and temporally variable.

Kim *et al.* also argue that the distinction between the two kinds of hubs bears on issues of network growth and evolution. Networks with hubs and small-world properties can arise by preferential attachment of new nodes to nodes that already participate in multiple interactions (*13*). This kind of preferential attachment is generally consistent with evolution by gene duplication. However, Kim *et al.* note that this seems to apply only to single-interface hubs. Two homologous proteins that arose within the same organism by gene duplication would tend to bind to the same surface of a hub protein; the data support this contention. The gene duplication and preferential attachment mechanism therefore appears to provide a route mainly to single-interface hubs. This leaves questions about the evolution of hubs of the multi-interface type.

A particularly interesting and perhaps unanticipated finding of the analysis concerns the degree to which hub proteins are essential to cell function. This topic has been examined before (*8*). Kim *et al.* find that, compared to single-interface hubs, multi-interface hubs are twice as likely to be essential. This suggests the idea that multi-interface hubs are somehow more highly integrated into the network of the cell. This makes some sense in view of the persistence of the interactions involved. Any given interaction to a single-interface hub will tend to be transient to a degree, in that it may be



Hub connections. (**Top**) In this simplified network diagram, a hub protein (yellow) connects three other proteins. There may be two kinds of hubs: (**left**) a single-interface hub or (**right**) a multi-interface hub. Idealized gene expression levels are indicated on graphs to reflect the generally low correlation between expression levels for single-interface hubs and the high correlation for multi-interface hubs.

The authors are in the Department of Chemistry and Biochemistry, University of California Los Angeles, Los Angeles, CA 90024–1569, USA. E-mail: yeates@mbi.ucla.edu

replaced at another time by interaction to a different partner. In contrast, interactions to multi-interface hubs may be persistent, and might therefore reflect necessity.

The idea that multi-interface hubs are more highly integrated into cellular networks may also affect the issue of horizontal gene transfer as an evolutionary mechanism. It has been argued that the more integrated a protein is into cellular organization, the less likely it is that a horizontally transferred gene will displace it (*14*). It will be interesting to see if there is a notable difference in horizontal gene transfer tendencies between single-interface and multi-interface hubs.

After years of fruitful work in systems biology, network analysis, bioinformatics, and structural genomics, cross-fertilization of these inherently related perspectives is beginning to take place (*15*). The work of Kim *et al.* shows the shift toward increased integration of multiple perspectives. Future progress in understanding cellular networks will require more complete data sets describing the underlying interactions. A knowledge of which proteins are interacting in the yeast cell is approaching some degree of completion, but only a fraction of those interactions can presently be mapped onto protein surfaces in three dimensions, which is required for the analysis developed by Kim *et al.* The insights drawn by their methods of analysis will be further strengthened as more structural data become available.

References

1. R. J. Williams, E. L. Berlow, J. A. Dunne, A. L. Barabasi, N. D. Martinez, *Proc. Natl. Acad. Sci. U.S.A* **99**, 12913 (2002).
2. R. D. Smith, *J. Stat. Mech.* **P02006**, 10.1088/1742-5468/2006/02/P02006 (2006).
3. A. L. Barabasi, Z. N. Oltvai, *Nat. Rev. Genet.* **5**, 101 (2004).
4. E. M. Marcotte, M. Pellegrini, M. J. Thompson, T. O. Yeates, D. Eisenberg, *Nature* **402**, 83 (1999).
5. L. Salwinski *et al.*, *Nucleic Acids Res.* **32**, D449 (2004).
6. R. Milo *et al.*, *Science* **298**, 824 (2002).
7. D. J. Watts, S. H. Strogatz, *Nature* **393**, 440 (1998).
8. H. Jeong, S. P. Mason, A. L. Barabasi, Z. N. Oltvai, *Nature* **411**, 41 (2001).
9. P. M. Kim, L. J. Lu, Y. Xia, M. B. Gerstein, *Science* **314**, 1938 (2006).
10. I. K. Jordan, Y. I. Wolf, E. V. Koonin, *BMC Evol. Biol.* **3**, 1 (2003).
11. H. B. Fraser, A. E. Hirsh, L. M. Steinmetz, C. Scharfe, M. W. Feldman, *Science* **296**, 750 (2002).
12. J. D. Han *et al.*, *Nature* **430**, 88 (2004).
13. G. P. Karev, Y. I. Wolf, F. S. Berezovskaya, E. V. Koonin, *BMC Evol. Biol.* **4**, 32 (2004).
14. C. R. Woese, *Microbiol. Mol. Biol. Rev.* **68**, 173 (2004).
15. S. A. Levin, *PLoS Biol*. **4**, e300 (2006).

10.1126/science.1137400

EVOLUTION

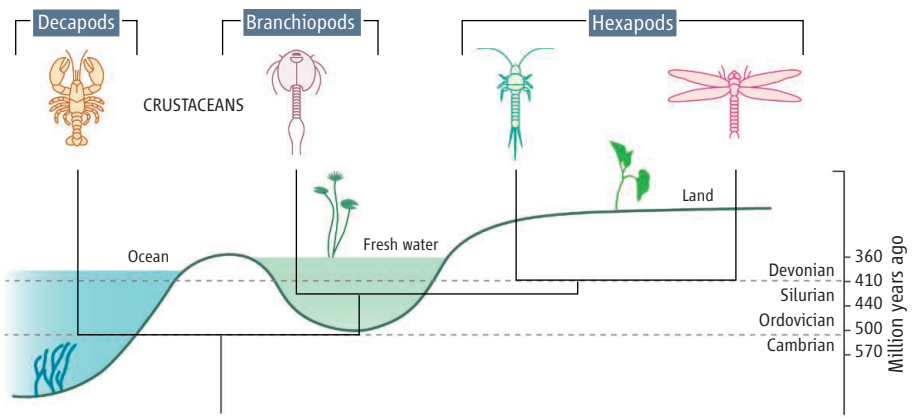
The Origin of Insects

Recent molecular evidence suggests that crustaceans may have successfully invaded land as insects.

Henrik Glenner, Philip Francis Thomsen, Martin Bay Hebsgaard, Martin Vinther Sørensen, Eske Willerslev

Although hexapods—those arthropods having six legs, including insects—are the most diverse group of contemporary animals in terms of biological niches and number of species, their origin is highly debated. A key problem is the almost complete absence of fossils that connect hexapods to the other major arthropod subphyla, namely Crustacea, Myriapoda (such as centipedes and millipedes), and Chelicerata (such as scorpions and spiders). Over the years, hexapods (insects, springtails, proturnas, and diplurans) have been phylogenetically linked to all of these major arthropod taxa (*1*).

Traditionally, hexapods and the multilegged myriapods have been united in a group named Atelocerata on the basis of morphological similarities between their tracheal respiration systems and head appendages. However, recent evidence from phylogenetic analyses of molecular sequence data from a variety of genes, as well as from newer morphological studies, points to a relationship between hexapods and crustaceans (*2*–*9*), a grouping commonly referred to as Pancrustacea. Furthermore, studies on neurological development in the major arthropod groups have pointed out similarities between the myriapods and chelicerates (*10*). Hence, pancrustacean monophyly seems to be gaining more support. So, what does this view tell us about the possible origin of hexapods?



Hexapod evolution. The last common ancestor of hexapods and crustaceans (branchiopods, specifically) may have originated in freshwater during the Late Silurian, giving rise to extant freshwater dwelling branchiopods (fairy shrimps, water fleas, and tadpole shrimps) and insects. This hypothesis accounts for the missing fossil record of branchiopods and hexapods before the Devonian.

The crustaceans are recorded at least as far back as the Upper Cambrian, about 511 million years ago (*11*), where they are found in marine sediments (see the figure). However, except for the debated *Devonohexapodus bocksbergensis* specimen (*12*, *13*), all hexapod remains are found only in freshwater or terrestrial strata no earlier than the Devonian, around 410 million years ago (*14*). This leaves a gap of 100 million years to the earliest crustaceans. The common explanation has been that earlier traces of hexapods have been erased from the fossil record and that hexapods, like other major groups of terrestrial animals, have closely related ancestors to be found in the marine environment.

The recent morphological and molecular-based studies suggest an alternative interpretation—that hexapods originated within the crustaceans rather than as a sister group (*15*–*20*). Although the morphological studies mainly favor a close phylogenetic connection between

The authors are at the Centre for Ancient Genetics, Department of Evolutionary Biology, Biological Institute, University of Copenhagen, Universitetsparken 15, DK-2100 Copenhagen, Denmark. E-mail: ewillerslev@bi.ku.dk