

nature Structural biology

march 2001
vol. 9 no. 3

molecular form & function



Proteomics in a small world

As we edge into the post-genomic era, considerable efforts are being expended identifying the various interactions of all proteins encoded in a genome — the proteome. Two recent papers^{1,2} concentrating on the proteome in the budding yeast *Saccharomyces cerevisiae* offer examples. But aside from their cataloging function, what do such studies tell us about the organization of the proteome and is it something that structural biologist have known for some time?

By introducing and expressing tagged proteins as 'baits' inside yeast cells, Gavin *et al.*¹ and Ho *et al.*² isolated through affinity chromatography all other proteins either directly bound or peripherally associated with them. The identities of the proteins captured in this way were then characterized using mass spectrometry. Together, these two studies recorded over 10,000 different protein-protein interactions and identified several hundred multiprotein complexes.

While such data are impressive, they should be interpreted with caution. Many known associations of proteins were not detected in these studies, and many of the observed interactions may prove to have no physiological significance. Indeed the most hopelessly optimistic estimate would be that no more than a third of all the interactions occurring in the proteome have been uncovered. Nonetheless, these studies have almost certainly sketched out the approximate shape of the web of interactions within the proteome.

Gavin *et al.* found that ~17% of the proteins studied made no detectable interactions with any other proteins, while the rest formed complexes containing anywhere from two to eighty three proteins. Over 90% of these complexes had at least one component that was not found in other complexes; however, around half of the proteins identified were found in more than one complex, and a few proteins were highly promiscuous, turning up in very many complexes. Although not explicitly checked, it would appear that the proteome forms a 'small-world' or scale-free network of interactions.

Such networks have been identified in many non-biological situations. The highly interconnected Internet is probably one of the most obvious examples. Another example, the readily traceable relationships among seemingly unrelated people, forms the basis of a well-known play: "Six degrees of separation"³. In biology, the interwoven web of metabolic pathways has recently been shown to have a scale-free character⁴.

Two properties of scale-free networks — robustness and adaptability — may be particularly advantageous to biology. The network is robust because some of its components can be altered or lost with little consequence for the overall behavior of the net-

work. Individual proteins can be mutated or completely lost, and yet, with the exception of a few crucial examples, the organism can more or less continue to function normally. Importantly, this robustness facilitates evolution as it allows a range of small variations or mutations to exist, some of which, with changing environmental pressures, will have a selective advantage.

Scale-free networks can also be highly modular. It may seem reasonable to expect that the metabolic networks from complex organisms, such as humans, would be larger than those from simple organisms, such as bacteria. This turns out not to be the case; the diameters of metabolic networks (approximated by the average shortest distance between any two points in the network) from organisms as different as archae and eukaryotes are remarkably similar⁴. Complexity is achieved more by slotting new routes into the heart of the network — and using as many existing connections as possible — than by bolting new components onto the outside.

To a structural biologist, modularity is nothing new. Many proteins consist of a number of domains. It often appears that evolution has occurred by using these domains like building blocks — shuffling and recombining domains from existing proteins to form new ones whose approximate function can often be deduced simply from the sort of domains they contain.

The use of common protein domains to form multidomain proteins shows striking similarities with the promiscuous behaviour of proteins in multiprotein complexes as documented by Gavin *et al.*¹. When investigating the 518 enzymes involved in metabolizing small molecules in *Escherichia coli*, Teichmann *et al.*^{5,6} found that these proteins were constructed from 213 different types of domains. While half the proteins have a single domain, less than a quarter of the domain types were found in these single domain proteins. Significantly, a small number of domain types were so prevalent that they made up 7% of all domains.

Proteomics is only beginning to reveal the complex context in which proteins play different parts, depending on their partners in multiprotein complexes. The emerging view is of a highly connected small world. A structural biologist should recognize this as a familiar phenomenon, projected at a different scale.

1. Gavin, A.-C. *et al.* *Nature* **415**, 141–147 (2002).
2. Ho, Y. *et al.* *Nature* **415**, 180–183 (2002).
3. Guare, J. *Six Degrees of Separation: A Play* (Vintage, New York; 1990).
4. Jeong, H., Tombor, B., Albert, R., Oltvai, Z.N. & Barabási, A.-L. *Nature* **407**, 651–654 (2000).
5. Teichmann, S.A. *et al.* *J. Mol. Biol.* **311**, 693–708 (2001).
6. Teichmann, S.A. *et al.* *Trends Biotechnol.* **19**, 482–486 (2001).