# Protein complexes take the bait

Anuj Kumar and Michael Snyder

Many cellular functions are carried out by proteins that are bound together in complexes. In two new large-scale studies, labelled proteins are used as 'bait' to capture and identify those complexes.

To appropriate a quote from John Donne, "no protein is an island entire of itself" — or at least, very few proteins are. Most seem to function within complicated cellular pathways, interacting with other proteins either in pairs or as components of larger complexes. A comprehensive understanding of these interactions will be needed before we can appreciate the mechanisms by which cellular pathways function and interlink. On pages 141 and 180 of this issue, Gavin *et al.*[1] and Ho *et al.*[2] describe significant advances towards this goal. Each group has characterized hundreds of distinct multiprotein complexes in the budding yeast *Saccharomyces cerevisiae*, using approaches in which individual proteins are tagged and used to pull down associated proteins, which are then analysed by mass spectrometry.

These studies[1,2] exemplify an emerging paradigm in protein biology: the systematic analysis of an organism's complete complement of proteins (its 'proteome'). Protein interactions on a proteome-wide scale have already been analysed in several ways. In a pair of landmark papers, Uetz *et al.*[3] and Ito *et al.*[4] adapted the yeast 'two-hybrid' assay — a means of assessing whether two single proteins interact — into a high-throughput method of mapping pair-wise protein interactions on a large scale. The authors collectively identified over 4,000 protein–protein interactions in *S. cerevisiae*. Our own group[5] has developed a microarray technology in which purified, active proteins from almost the entire yeast proteome are printed onto a microscope slide at high density, such that thousands of protein interactions (and other protein functions) can be assayed simultaneously.

Gavin *et al.*[1] and Ho *et al.*[2] take a different approach — one that is particularly effective at identifying protein complexes that contain three or more components. Large-scale efforts to characterize protein complexes are generally rate-limited by the need for a nearly pure preparation of each complex. In the new studies[1,2], protein complexes were purified as follows (Fig. 1). First, the authors attached tags to hundreds of different proteins (to create 'bait' proteins). They then introduced DNA encoding these bait proteins into yeast cells, allowing the modified proteins to be expressed in the cells and to form physiological complexes with other proteins. Then, using the tag, each bait protein was pulled out, often fishing out the entire complex with it (hence the term 'bait'). The proteins extracted with the tagged bait were identified using standard mass-spectrometry methods.

Applying this approach on a proteome-wide scale, Gavin *et al.*[1] have identified 1,440 distinct proteins within 232 multiprotein complexes in yeast. As 91% of these complexes contain at least one protein of previously unknown function, the study provides a wealth of new information on 231 previously uncharacterized yeast proteins, and on a further 113 proteins to which the authors ascribe a previously unknown cellular role. Furthermore, Gavin *et al.* find that most of these complexes have a component in common with at least one other multiprotein assembly, suggesting a means of coordinating cellular functions into a higher-order network of interacting protein complexes.

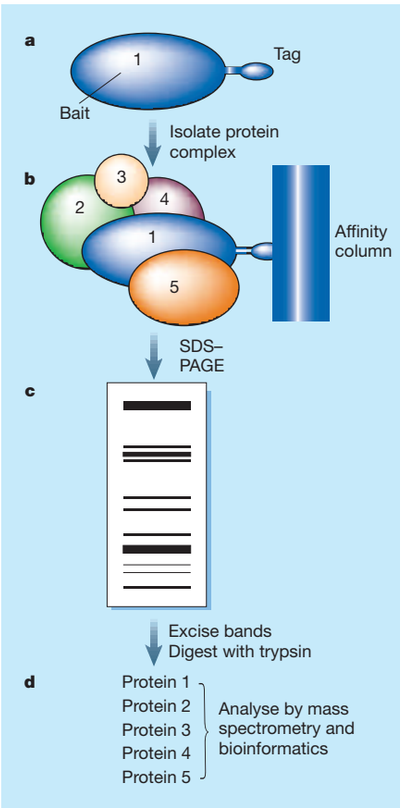


Figure 1 **Analysing protein interactions. In the 'co-precipitation/mass spectrometry' approach used by Gavin *et al.*[1] and Ho *et al.*[2], an 'affinity tag' is first attached to a target protein (the 'bait';** a). b, **Bait proteins are systematically precipitated, along with any associated proteins, on an 'affinity column'.** c, **Purified protein complexes are resolved by one-dimensional SDS–PAGE, a technique that involves running an electric charge through the complexes on a gel, so that proteins become separated according to mass.** d, **Proteins are excised from the gel, digested with the enzyme trypsin, and analysed by mass spectrometry. Database-search algorithms (bioinformatics) are then used to identify specific proteins from their mass spectra.**

An understanding of this high-order organization will undoubtedly offer insight into corresponding networks in other organisms, as most yeast complexes have counterparts in more complex species (one reason why researchers are interested in this unicellular organism). Gavin and colleagues illustrate this point by purifying and analysing three equivalent multiprotein complexes from yeast and human cells: the Arp2/3 complex, a component of the cellular 'skeleton'; the Ccr4–Not1 complex, which is found in the nucleus; and the TRAPP complex, which is involved in transport from one intracellular compartment (the endoplasmic reticulum) to another (the Golgi). In each case, the authors retrieved human and yeast complexes that were similar, if not identical, in composition.

Using the same general approach, Ho *et al.*[2] constructed an initial set of 725 yeast bait proteins, from which they identified 3,617 interactions involving 1,578 different proteins. They describe interaction networks assembled around the protein kinase Kss1 — a known component of pathways involved in mating and filamentous growth — and complexes associated with the cyclin-dependent kinase Cdc28 and the gene-transcription factors Fkh1 and Fkh2. In addition, Ho and colleagues used 86 bait proteins that are implicated in the DNA-damage response, allowing them to delineate much of the yeast damage-response network. In particular, they reveal many regulators and targets of the protein kinase Dun1, and a possible role for the DNA-repair protein Rad7 in processes of targeted protein degradation.

The approach taken by Gavin *et al.* and Ho *et al.* is clearly powerful, but it does have

drawbacks. Both groups find a significant number of false-positive interactions, while failing to identify many known associations. Gavin *et al.* estimate that 30% of the interactions they detect may be spurious, as inferred from duplicate analyses of 13 purified complexes. Conversely, they failed to detect any interacting partners for Bmh2 (ref. 6), a regulatory protein that has previously been shown to interact with a number of other proteins, including Ste20 (involved, for example, in yeast mating)[7], and Msn2 and Msn4 (stress-responsive transcription factors)[8]. Ho *et al.*, meanwhile, did not detect nucleotide excision repair factor-2, a tight complex[9] that contains the well-characterized DNA-repair proteins Rad4 and Rad23. So, as in most large-scale studies, these results are imperfect. It will be essential to integrate data from many different sources to obtain an accurate understanding of protein networks.

Proteomic studies such as these[1,2] have generated a huge volume of exciting data. Yet — setting aside the problem of false positives and negatives — there is much still to be learned before we have a comprehensive knowledge of functional pathways within even a model organism such as yeast. To understand the magnitude of the task, consider the yeast proteome. Assuming that each protein interacts with an average of five partners — a reasonable estimate drawn from experience and preliminary two-hybrid results — the yeast proteome should encompass some 30,000 protein interactions, many of which change during the life cycle of the organism. So far, protein microarray analyses and studies like those of Gavin *et al.* and Ho *et al.* have collectively identified, at most, 11,000 different protein associations (and probably fewer, considering the potential overlap between data sets). Although feasible, the characterization of all remaining interactions will almost certainly be labour intensive. But the resulting data will be more than worth the effort. ■

*Anuj Kumar and Michael Snyder are in the Department of Molecular, Cellular and Developmental Biology, Yale University, New Haven, Connecticut 06520-8103, USA.*
*e-mails: anuj.kumar@yale.edu*
*michael.snyder@yale.edu*

1. Gavin, A.-C. *et al. Nature* **415,** 141–147 (2002).
2. Ho, Y. *et al. Nature* **415,** 180–183 (2002).
3. Uetz, P. *et al. Nature* **403,** 623–627 (2000).
4. Ito, T. *et al. Proc. Natl Acad. Sci. USA* **98,** 4569–4574 (2001).
5. Zhu, H. *et al. Science* **293,** 2101–2105 (2001).
6. Gelperin, D. *et al. Proc. Natl Acad. Sci. USA* **92,** 11539–11543 (1995).
7. Roberts, R. L. *et al. Cell* **89,** 1055–1065 (1997).
8. Beck, T. & Hall, M. N. *Nature* **402,** 689–692 (1999).
9. Guzder, S. N., Sung, P., Prakash, L. & Prakash, S. *J. Biol. Chem.* **273,** 31541–31546 (1998).

Oceanography

# Bubbling under

Chris German

The study of hydrothermal vents is a young and fertile discipline. The latest findings, and the enticing prospects offered by new technology, came in for discussion at two meetings held late last year.

Deep-sea hydrothermal activity was discovered only in 1977, on the sea floor near the Galapagos Islands, so it is little wonder that the phenomenon continues to spring surprises. Hydrothermal vents arise where cold sea water interacts with freshly formed, hot ocean crust along chains of submarine volcanoes, termed mid-ocean ridges, which run for more than 50,000 km across the world's ocean basins. Two different meetings on hydrothermal systems were held late last year, in San Francisco and Berlin*, attesting to the vigour of research on the topic.

The event in San Francisco was convened in honour of the geochemist John Edmond, who died last year aged 57. Edmond was the first to recognize the full importance of hydrothermal activity. He used the first observations from near the Galapagos Islands to make prescient estimates about the extremely high temperatures (around 350 °C) that hydrothermal fluids can reach, and discussed their significance in terms of chemical fluxes to the oceans ("These," he remarked laconically in a 1979 paper[1], "are large.") Fittingly, at San Francisco, it was two of Edmond's former students who described the most exciting new results.

Beneath the permanent ice cover of the Arctic Ocean, evidence has been found of extensive hydrothermal activity along the world's slowest-spreading and least volcanically active plate-tectonic boundary, the Gakkel Ridge (H. Edmonds, Univ. Texas). This discovery is as notable for astrobiologists as for others, in that it will provide a natural testing ground for techniques that could be used to explore Jupiter's ice-bound satellite Europa. Meanwhile, preliminary investigations of some of the first vent fluids to be collected from the Indian ocean — from one of only two known sites, both identified within the past year or so — have shown that these samples are more chemically enriched than similar fluids studied previously (K. Von Damm, Univ. New Hampshire).

Soon after seafloor venting was first discovered, it was proposed that hydrothermal activity may be geographically rare and occur only along the volcanically active East Pacific Rise. But the news about Gakkel Ridge adds to a mounting body of evidence that hydrothermal activity can occur no matter how low the rate at which magma is supplied from the underlying mantle. As recently as ten years ago, it was commonly thought that hydrothermal fluids should all have a similar composition that would remain invariant for decades. On the evidence of Von Damm's report, however, it seems that we are still some way from understanding what the full flux from hydrothermal activity to the oceans may be. Clearly, the world is not nearly as straightforward as we once thought.

But why study these hydrothermal systems? Surely, even if they are dramatic (Fig. 1), they are of only marginal significance and appeal? The answer has several facets, not least of which is that hydrothermal activity is part of the way in which the Earth works. Such activity is intimately linked to plate tectonics and occurs almost exclusively at ocean ridges, yet to Earth scientists it represents a significant cooling process for the planet as a whole. For oceanographers, hydrothermal fluxes of chemicals entering the oceans (notably iron and manganese, but also methane, hydrogen and hydrogen sulphide) are comparable to those from rivers, and help to buffer the chemical balance of the oceans and atmosphere. The best estimates suggest that the entire volume of the oceans is processed through high-temperature vents every ten million years, and through deep-sea hydrothermal plumes every few thousand years. The latter timescale is similar to that of the mixing time for the modern-day deep ocean, and shorter than that of glacial–interglacial oscillations.

Hydrothermal vents offer plenty for life scientists, too: more than 400 new species have so far been found at vent sites, representing the discovery of one new species every two weeks, on average, throughout the past quarter-century. These organisms range from anaerobic microbes all the way to spectacular red-plumed tubeworms, which can reach several metres in length. In between are organisms such as the blind shrimps of the Mid-Atlantic Ridge, now also known to be present in the Indian Ocean, which are only a few centimetres long but swarm in millions around individual vent sites. Hydrothermal vent microbes seem to sit at the very base of the 'tree of life', and provide insights into both the origins of life here on Earth and its possible existence elsewhere in the Solar System.

**Pushing the Envelope: A Tribute to the Career and Accomplishments of John M. Edmond*, American Geophysical Union meeting, San Francisco, 11–15 December 2001; *Energy and Mass Transfer in Marine Hydrothermal Systems*, 89th Dahlem Workshop, Berlin, 15–19 October 2001.