TECHVIEW: PROTEIN BIOCHEMISTRY

Protein Interaction Methods—Toward an Endgame

Andrew R. Mendelsohn and Roger Brent

Thousands of new protein-protein interactions have been identified through the use of the two-hybrid system, shedding much light on molecular and cellular biology. In the next decade, we can hope to see large catalogs of protein interactions, predictive models of those interactions, and at least some ability to follow networks of those interactions in real time.

Here, we review newer methods that may get us closer to this goal. These include descendants of conventional two-hybrid methods, other methods that rely on reconstitution of biochemical function in vivo, fluorescence resonance energy transfer methods, protein mass spectrometry, and evanescent wave methods. Taken together, these methods will help reveal not only the partners of particular proteins, but how tightly the interacting proteins touch one another, which surfaces they use to make contact, and where and when in living cells those contacts occur. With these methods, entire networks of interacting proteins can be analyzed.

Modified Two-Hybrid Systems

Two-hybrid methods and their descendants (*1*) can be used in a number of ways. One of these descendants enables the charting of genetic networks and depends on interaction mating (*2*). By mating "baits" expressed in haploid yeast of one mating type with "preys" expressed in another, diploid exconjugants are created that can be interrogated to determine whether the proteins interact. This microbiological technique facilitates scaling up of the detection of individual binary protein-protein interactions. Mating has been used to survey protein-protein interactions for bacteriophage T7 (*3*) and for the *Saccharomyces cerevisiae* RNA splicing machinery (*4*), and is being used to map interactions among the proteins encoded by the *S. cerevisiae* genome (*5*).

Two-hybrid systems that can identify disruption of protein interactions ("reverse two-hybrid systems") have been implemented in three ways. In one, the positive selectable reporter is replaced by a counterselectable reporter, such as URA3, LYS2, or CYH2 (*6*). Vidal *et al.* used this approach to isolate point mutants of the positive cell-cycle regulator E2F1 that do not interact with DP1 (*6*). In another system, two reporters are coupled such that interacting proteins induce expression of a DNA-binding protein, Tn10 tetracycline repressor (TetR), which then represses transcription of a second TetRop-HIS3 gene, so that only cells in which the original protein interaction is disrupted grow in the absence of histidine (*7*). Shih *et al.* used this approach to identify CREB mutants that no longer bind CREB-binding protein (*7*). In a third approach, Geyer *et al.* used peptide aptamers that potentially disrupted protein interactions in cells engineered to express two interacting proteins at low concentrations, and identified disrupted interactions by diminution of the positive signal (*8*).

The last 2 years have seen the advent of "two-bait systems" in which different baits are bound to DNA upstream of different reporters. These systems have been used to identify proteins that interact with different domains of a protein (Snf1) (*9*), different alleles of a protein (Ras) (*10*), and to identify mutant proteins that differentially bind to two known interactors of a wild-type protein (Ste5) (*11*). Combining data from these systems with data from conventional two-hybrid systems likely can help dissect topologies of multimeric protein complexes (*10*). Two-bait systems, like other systems in which a third protein is expressed, can detect interactions that depend on bridging or modification by a third protein (*12*). More baroque yeast systems would allow simultaneous analysis of more than three proteins.

Other Reconstitution Methods

Transcription activation is not the only biochemical activity that can be reconstituted by protein interaction. For example, Aronheim *et al.* (*13*) described systems in which interacting protein partners reconstitute guanine exchange factor (GEF) or Ras by bringing together the catalytic domain with the required membrane localization domain, which then complements yeast carrying a temperature-sensitive mutation in yeast GEF (Cdc25ts). Although these systems are well suited for assaying interactions between cytoplasmic and membrane-proximal proteins, as well as proteins that activate transcription (*13*, *14*), they can be deployed only in appropriately engineered yeast cells and are prone to false positive signals, such as those due to reversion of the Cdc25ts allele (*13, 14*).

Another system more generally illustrates that interacting protein fragments can sometimes reconstitute a split protein's function. Varshavsky and co-workers (*15*) showed that a mutated NH_2-terminal fragment of ubiquitin and a COOH-terminal fragment fused to a bait moiety (*15*) could, upon interaction, reconstitute ubiquitin, whose cleavage from the COOH-terminal fragment could be detected by protein immunoblotting. For *S. cerevisiae*, Stagljar *et al.* have described a transcription-based selectable version of this "ubiquitin split protein sensor" (*16*) in which, upon interaction, ubiquitin reconstitution, and cleavage, a transcription factor activator is released to activate a nuclear localized reporter.

In a more general approach, a number of workers have described split enzymes. In the most powerful of these, the reconstituted enzymatic activity is the scored phenotype. Interaction-mediated reassembly of enzymatic activity has been described for *Escherichia coli* β-galactosidase (β-gal), mouse dihydrofolate reductase (DHFR), and *Bordetella pertussis* CysA adenyl cyclase, (*17–19*). The β-gal experiments bring together weakly interacting NH_2-terminal (α) and COOH-terminal (Ω) fragments of the protein via fused interacting partners (*17*). The authors used this reconstitution to detect rapamycin-dependent interaction of FK506-binding protein (FKBP12) with FRAP (FKBP-rapamycin–binding protein) in mammalian cells; they detected β-gal histochemically, assayed it biochemically, and identified and isolated β-gal–containing cells by fluorescence-activated cell sorting (FACS). Similarly, Remy and Michnick showed that fragments of mouse DHFR, $DHFR_{fragment[1,2]}$, and $DHFR_{fragment[3]}$ can be reconstituted into a complex by interacting fused partners (again, FKBP12 and FRAP), allowing the cells to grow in a nucleotide-free medium (*18*). Cells with reconstituted DHFR could be stained by a substrate analog, fluorescein-methotrexate, which allowed visualization of the protein complex by fluorescent microscopy to determine its subcellular location, and to select cells containing it by FACS. Finally, Karimova *et al.*

The authors are at the Molecular Sciences Institute, 2168 Shattuck Avenue, Berkeley, CA 94704, USA. E-mail: amendelsohn@sequence.molsci.org

(*19*) demonstrated reconstitution of adenylyl cyclase in *E. coli*. cAMP produced by this enzyme activated CAP and in turn enabled transcription from the *lac* or *mal* operons (*19*).

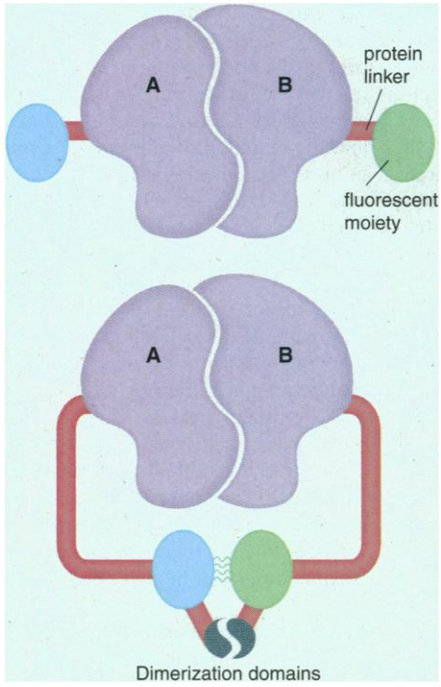
Although this second-messenger function of cAMP is confined to prokaryotes, it is easy to imagine cGMP-dependent phenotypes that could be engineered into eukaryotic cells. Analogously, the β-gal and DHFR systems, although engineered for mammalian cell culture, could be reengineered for other eukaryotes, prokaryotes, and in vitro translation experiments.

Many other enzymes are modular or can be made so. It is thus likely that other such approaches will be developed. Because transcription activation domains function over a vast range of distances and orientations, eukaryotic transcription activation is robust to much of the geometrical variation imposed by different fusion partners. It remains to be seen whether any split enzymes will manifest similar geometrical flexibility. Lack of flexibility may limit the usefulness of enzyme reconstitution for identifying interacting proteins from libraries. However, by using long and flexible linkers and fragments of enzymes that do not need initial precise positioning to reconstitute activity, it should be possible to devise general interaction detection systems (see figure). The use of more than one such system at a time should make possible the study of complex multiprotein interactions in vivo.

Fluorescence Resonance Energy Transfer

When two fluorophores with overlapping emission/absorption spectra are within ~100 Å of one another and their transition dipoles are appropriately oriented, stimulation of the higher-energy donor fluorophore excites the lower-energy acceptor fluorophore, causing it to emit photons. This phenomenon is called fluorescence resonance energy transfer (FRET) (*20, 21*).

Because FRET falls off with the 6th power of distance, it has long been used to study proteins, mostly as a "molecular ruler" (*20*, *21*). The discovery (*22*) and cloning (*23*) of green fluorescent protein (GFP) and the development of spectral derivatives of it (*24*) made it possible to make fluorescent protein fusions inside cells. These developments allowed the use of FRET to detect protein interactions in real time in living cells. In such experiments, one protein is fused to a FRET donor, the other to a FRET acceptor. The proteins are expressed inside cells, and their interaction is monitored by fluorescence microscopy or light spectroscopy. For example, FRET between BFP-Bcl2 and GFP-Bax revealed that Bcl2 (a cell death inhibitor) and Bax (a cell death potentiator) interacted inside mitochondria (*25*). In a related development, Xu *et al.* described a FRET-based assay for activation of the death protease Caspase 3: activated protease cleaves a FRET pair linked by a protease recognition sequence (*26*). Note that FRET can, in principle, register protein-protein interaction in any cellular compartment.



Testing for protein-protein interaction. When the two fluorescent moieties are more than 100 Å apart, FRET cannot occur, even though the proteins (labeled A and B) interact (top). Fusion proteins with extremely long hinge regions that bear a moiety from a protein that homodimerizes weakly could generate a strong FRET signal upon interaction (bottom).

Currently, use of FRET is limited by the low signal-to-noise ratios of the two available mutant GFP pairs usable for FRET (*24*), by rapid photobleaching of GFP and mutant GFPs (*24*), and most importantly by the fact that FRET only works over distances up to ~100 Å, which excludes interacting fusion partners that leave the fluorescent moieties too far apart (the fluorophores are already buried 12 Å within the GFP monomer) (*24*). Although this distance problem may limit FRET's ability to compete with two-hybrid methods in genome-wide interaction assays, the other problems associated with GFP may be overcome. For example, more efficient FRET may be possible with two-photon excitation, wherein two infrared photons are absorbed within femtoseconds to excite fluorophores, including GFP, at wavelengths equal to the sum of the energies of the incident photons (*24*) or by using other optical phenomena sensitive to FRET, such as changes in the lifetime of the fluorophore (*27*), to circumvent photobleaching.

In another approach to bypass the limitations of GFP-derived fluorophores, two groups have described chemical fluorescent labeling of proteins inside cells. Shultz and co-workers used engineered UAG suppressor tRNAs designed to force incorporation of a fluorescent amino acid derivative into a protein whose coding sequence is modified to include UAG (*28*). Efforts to use new nucleotides to extend the genetic code beyond its current 64 codons (*29*) may eventually enable suspected interacting proteins to be tagged with unique codons that direct incorporation of FRET-capable fluors. However, in the near term, in vivo labeling requires chemical synthesis of fluor-charged tRNAs and their microinjection into living cells. In another interim approach, Tsien and co-workers have expressed proteins that contain a compact arsenic-binding domain, which binds a cell-permeable fluorescein-arsenic derivative (*30*).

Xu *et al.* have proposed bioluminescence resonance energy transfer (BRET) for detection of protein interactions, by using *Renilla* luciferase to transfer energy to yellow fluorescent protein, a mutant GFP (*24*). Using BRET, they showed that the cyanobacteria circadian clock protein KaiB homodimerizes (*31*). Excitation by bioluminescence eliminates photobleaching and autofluorescence associated with fluorescent excitation in FRET, but requires that the assayed cells be exposed to the luciferase substrate coelenterazine. BRET is overall less sensitive than FRET and is subject to poorly understood distance constraints (*31*).

Protein Mass Spectrometry

Many of the individual binary interactions in higher order protein complexes are likely to be too weak to be detected by two-hybrid methods. However, for organisms with known genomes, recent advances in protein mass spectrometry have vastly eased the identification of complex molecules (*32*). Proteins and tryptic peptides from these complexes can be analyzed by MALDI-TOF (*32*), sequences inferred from the mass, and the sequences compared with a database of predicted proteins encoded by the organism's genome. If mass alone cannot predict the exact sequence, fragmentation methods (nano-electrospray tandem

mass spectrometry) can be used to produce stretches of up to 16 amino acids of sequence from femtomolar amounts of protein fragments (*33*).

So far, such methods have been used to identify Caspase 8 (Flice) as an interactor with immuoprecipitated CD95 (Fas/Apol) (*34*) and to characterize the protein complement of a number of multiprotein complexes including the spliceosome (*35*), the yeast u1 snrp (*36*), yeast spindle pole body (*37*) and the yeast anaphase promoting complex (*38*), and trans Golgi network–derived transport vesicles (*39*).

Other than its capital cost and the high level of technical sophistication required to use it well, the most significant limitation of mass spectrometric analysis of protein interactions is that the protein complexes first need to be isolated by physical methods such as electrophoresis(*33*). Many proteins will not be detected because, for example, they are too scarce, large, small, acidic, or alkaline to be studied by two-dimensional gel analysis, or the interaction is too weak and transient to survive affinity purification.

Clearly, mass spectrometric methods have tremendous potential. As their sensitivity and ease of use improves, mass spectrometry will come to complement biological methods for detecting and analyzing protein interactions, and may eventually supplant them.

Evanescent Wave Methods

Evanescent waves can be created at interfaces of transparent media of two different refractive indexes (here, glass and aqueous medium). When the angle of incidence of a light beam in the glass exceeds a critical angle, the light is reflected from the interface back into the glass. The electromagnetic field associated with the light creates an "evanescent" wave in the aqueous medium, which decreases in strength exponentially away from the interface. The energy from the evanescent wave is available to probe the volume near the interface (*40*) and can be used to detect protein interactions. In "surface plasmon resonance" (*41*) devices, there is a thin layer of metal at the interface between the glass and the aqueous medium. There exists an angle of incident light (a "resonance angle") at which some of the energy in the evanescent wave is dissipated into the electron cloud ("plasmon") in the metal. This angle depends on the local refractive index—which in turn depends on the mass of proteins bound to the interface. If a layer of bait protein is bound near the metallic surface and a solution containing an interacting protein flows past this surface, the resonance angle changes as the protein in solution binds the bait. The rate of increase in mass is the association rate. When a solution without the interacting protein flows past, the mass at the surface gradually decreases as the interacting protein dissociates and is washed away. The rate of that decrease is the dissociation rate. These rates give the interaction affinity.

Evanescent wave techniques are fast becoming the methods of choice (*41*) for quantifying interactions between known proteins. First-generation instruments required relatively large (milligram) amounts of putative interacting proteins, and the resolution with which the changes in mass was detected was far too low to allow identification of specific proteins from complex mixtures without a subsequent analytical step such as mass spectrometry (*42*). However, these methods are likely to become even more widely used, because thin optical fibers are being used to make cheap devices with which an evanescent wave interrogates a glass/aqueous medium interface (*43*). Bundles of such fibers, each conjugated with a different bait, may be able to simultaneously detect in vivo different proteins in extracellular compartments (such as blood) in organisms and even—if the fibers can be made small enough—in living cells.

Toward the End Game

New technologies to identify and characterize protein interactions will become available. For example, Roberts *et al.* described a method in which translated proteins are covalently coupled to the mRNA that encodes them in vitro (*44*). Individual mRNAs associated with individual proteins can be identified by reverse transcribing and amplified by polymerase chain reaction. The mRNA tags can thus, in principle, identify those proteins encoded by a large pool of mRNAs, whose interaction is not blocked by association with the large, negative cloud of mRNA, and that can interact under the dilute conditions of this assay (*44*).

Many protein interaction technologies are naturally combined. For example, we can imagine using a split enzyme approach to select and then characterize in vivo the proteins that interact with a particular bait, and in parallel, using that bait as an affinity tag to isolate protein complexes whose constituents are characterized by mass spectrometry. Development of these combined approaches will hasten the day when scientists can inventory all the protein interactions in selected organisms and cell types.

References

1. M. Vidal and P. Legrain, *Curr. Opin. Chem. Biol.* **3**, 64 (1999); B. L. Drees, *Nucleic Acids Res.* **27**, 919 (1999); P. Colas and R. Brent, *Trends Biotechnol.* **16**, 355 (1998); R. Brent and R. L. Finley Jr., *Annu. Rev. Genet.* **31**, 663 (1997); S. Fields and O. Song, *Nature* **340**, 245 (1989).
2. J. W. Harper *et al.*, *Cell* **75**, 805 (1993); R. L. Finley Jr. and R. Brent, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 12980 (1994). See also review by Vidal and Legrain in (*1*).
3. P. L. Bartel *et al.*, *Nature Genet.* **12**, 72 (1996).
4. M. Fromont-Racine, J. C. Rain, P. Legrain, *ibid.* **16**, 277 (1997).
5. J. R. Hudson Jr. *et al.*, *Genome Res.* **7**, 1169 (1997).
6. M. Vidal *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 10315 (1996); M. Vidal *et al.*, *ibid.*, p. 10321; J. Huang and S. L. Schreiber, *ibid.* **94**, 13396 (1997); C. A. Leanna and M. Hannink, *Nucleic Acids Res.* **24**, 3341 (1996).
7. H. M. Shih *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 13896 (1996).
8. R. Geyer, A. Colman-Lerner, R. Brent, *Proc. Natl. Acad. Sci. U.S.A.*, in press.
9. R. Jiang and M. Carlson, *Genes Dev.* **10**, 3105 (1996).
10. C. W. Xu, A. R. Mendelsohn, R. Brent, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 12473 (1997).
11. C. Inouye *et al.*, *Genetics* **147**, 479 (1997).
12. J. Zhang and S. Lautar, *Anal. Biochem.* **242**, 68 (1996).
13. A. Aronheim *et al.*, *Mol. Cell. Biol.* **17**, 3094 (1997); A. Aronheim, *Nucleic Acids Res.* **25**, 3373 (1997).
14. Y. C. Broder, S. Katz, A. Aronheim, *Curr. Biol.* **8**, 1121 (1998).
15. N. Johnsson and A. Varshavsky, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 10340 (1994); M. Dunnwald, A. Varshavsky, N. Johnsson, *Mol. Biol. Cell* **10**, 329 (1999).
16. I. Stagljar *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 5187 (1998).
17. F. Rossi, C. A. Charlton, H. M. Blau, *ibid.* **94**, 8405 (1997).
18. I. Remy and S. W. Michnick, *ibid.* **96**, 5394 (1999); J. N. Pelletier, F. X. Campbell-Valois, S. W. Michnick, *ibid.* **95**, 141 (1998).
19. G. Karimova *et al.*, *ibid.* **95**, 5752 (1998).
20. P. Wu and L. Brand, *Anal. Biochem.* **218**, 1 (1994).
21. R. M. Clegg, *Curr. Opin. Biotechnol.* **6**, 103 (1995).
22. O. Shimomura, F. H. Johnson, Y. J. Saiga, *J. Cell. Comp. Physiol.* **59**, 223 (1962).
23. D. C. Prasher *et al.*, *Gene* **111**, 229 (1992).
24. R. Y. Tsien, *Annu. Rev. Biochem.* **67**, 509 (1998); B. A. Pollok and R. Heim, *Trends Cell Biol.* **9**, 57 (1999).
25. N. P. Mahajan *et al.*, *Nature Biotechnol.* **16**, 547 (1998).
26. X. Xu *et al.*, *Nucleic Acids Res.* **26**, 2034 (1998).
27. P. I. Bastiaens and A. Squire, *Trends Cell Biol.* **9**, 48 (1999).
28. G. Turcatti *et al.*, *J. Biol. Chem.* **271**, 19991 (1996).
29. J. D. Bain *et al.*, *Nature* **356**, 537 (1992); S. A. Benner, *Trends Biotechnol.* **12**, 158 (1994).
30. B. A. Griffin, S. R. Adams, R. Y. Tsien, *Science* **281**, 269 (1998).
31. Y. Xu, D. W. Piston, C. H. Johnson, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 151 (1999).
32. J. S. Andersen, B. Svensson, P. Roepstorff, *Nature Biotechnol.* **14**, 449 (1996); J. R. Yates III, *J. Mass Spectrom.* **33**, 1 (1998); J. A. Loo, *Mass Spectrom. Rev.* **16**, 1 (1997).
33. O. N. Jensen *et al.*, *Methods Mol. Biol.* **112**, 571 (1999); A. Shevchenko *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 14440 (1996); M. Wilm *et al.*, *Nature* **379**, 466 (1996).
34. M. Muzio *et al.*, *Cell* **85**, 817 (1996).
35. G. Neubauer *et al.*, *Nature Genet.* **20**, 46 (1998).
36. G. Neubauer *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 385 (1997).
37. P. A. Wigge *et al.*, *J. Cell Biol.* **141**, 967 (1998).
38. W. Zachariae, *Science* **279**, 1216 (1998).
39. A. Shevchenko *et al.*, *Electrophoresis* **18**, 2591 (1997).
40. N. L. Thompson and B. C. Lagerholm, *Curr. Opin. Biotechnol.* **8**, 58 (1997); P. B. Garland, *Q. Rev. Biophys.* **29**, 91 (1996); A. M. Hutchinson, *Mol. Biotechnol.* **3**, 47 (1995).
41. P. Schuck, *Curr. Opin. Biotechnol.* **8**, 498 (1997); *Annu. Rev. Biophys. Biomol. Struct.* **26**, 541 (1997).
42. R. W. Nelson, J. R. Krone, O. Jansson, *Anal. Chem.* **69**, 4369 (1997); *ibid.*, p. 4363.
43. E. Squillante III, *Drug Dev. Ind. Pharm.* **24**, 1163 (1998). See also review in (*37*).
44. R. W. Roberts and J. W. Szostak, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 12297 (1997).