A guided tour in protein interaction space: Coiled coils from the yeast proteome

James C. Hu*

Department of Biochemistry and Biophysics, and Center for Advanced Biomolecular Research, Biochemistry/Biophysics Building, Room 443A, Texas A&M University, College Station, TX 77843-2128

Exploring the interactions among the proteins identified by genome sequencing can be overwhelming. Like a visitor arriving in a new country, there are many different ways to take in the points of interest. In the pregenomic era, we tended to stay in particular neighborhoods and study them in great detail. By analogy, we've learned a lot about protein–protein interactions from studying specific proteins or biochemical activities from many model organisms. Complete genome sequences have whetted our appetite for a different kind of exploring by giving us the genes for every protein encoded by the genome. Now we can complement our knowledge of the famous landmarks with a directory of every building. Recent papers from Uetz *et al.* (1) and Ito *et al.* (2) have described the early progress at generating a comprehensive census of the protein–protein interactions from the budding yeast *Saccharomyces cerevisiae*. These two groups both used the yeast two-hybrid system (3) as a rapid and generalizable assay to detect interactions between any two proteins, called the bait and the prey. Pairwise combinations of proteins were examined where a full-length ORF of every known or predicted gene from *S. cerevisiae* was used as bait and/or prey.

The paper by Newman *et al.* (4) in this issue of PNAS provides a complementary approach that uses bioinformatics to guide the search for interactions to a particular class of interacting proteins: those with coiled-coil domains. Coiled coils are perhaps the most common motif used to hold proteins together. Using the program MULTICOIL (5) to predict coiled coils from the yeast genome, Newman *et al.* found that 1 of every 11 yeast proteins is predicted to contain a coiled coil; they estimate that more than 5% of putative ORFs in sequenced genomes contain coiled coils. Because coiled coils are known to interact with each other in homotypic and heterotypic complexes, focusing a two-hybrid screen on predicted coiled coils had an excellent chance of finding both partners for any interaction based on this motif.

Newman *et al.* examined interactions among 162 putative coiled-coil regions from 121 yeast proteins of the 550 or so that were predicted to contain coiled coils. The subset was selected to include proteins that had high MULTICOIL scores and/or annotations in the Yeast Proteome Database that suggested involvement in the spindle pole body. From the $162 \times 162 = 26{,}244$ pairwise tests suggested by this computationally directed screen (fewer were actually used due to activation by 11 of the bait fusions in the absence of any prey), 213 interactions were found involving 100 putative coiled coils from 77 different proteins. Of the identified interactions, 33 were homotypic and 175 were heterotypic. Many of the tested coiled-coil sequences interacted with more than one partner; the biological significance of these multipartner interactions is not yet clear.

As noted by the authors, an obvious extension of this work is to perform directed interaction screens based on other known interaction motifs. Given the ongoing efforts to generate comprehensive interaction maps, one might wonder whether the directed approach championed by Newman *et al.* is merely looking for a subset of the interactions that have been or eventually will be found by the comprehensive projects. In fact, there are several good reasons why this kind of directed approach will play a central role in the future exploration of protein interaction space.

First, any two-hybrid assay for protein–protein interactions scales as the square of the number of candidate interactions that are being examined. Screening every possible ORF $\times$ ORF interaction for the approximately 6,000 genes in the yeast genome requires examining on the order of 3.6×10^7 pairs. Despite the heroic efforts of the comprehensive searches, it is clear that they are far from saturation, judging by the fact that there are many nonoverlapping hits identified by the two studies. Indeed, none of the interactions identified by Newman *et al.* were found in either of the comprehensive screens. A 10-fold enrichment for proteins that are likely to be involved in an interaction gives a 100-fold improvement in the size of the search that has to be performed. The number of pairwise possibilities for the 550 or so two- and three-stranded coiled coils predicted by MULTICOIL is only on the order of 3×10^5. Extrapolating these calculations to larger genomes increases the advantage of a directed approach exponentially.

Second, any interaction identified by the directed approach described by Newman *et al.* localizes the site of interaction within the ORF. Localization can provide clues about function in cases where a protein interacts with different partners. If the interactions share the same target, the proteins may be involved in a subunit exchange process to generate a series of different complexes with shared components. The paradigm for this kind of subunit sharing is provided by the bZip transcription factors, where different combinations of subunits are held together by mixing and matching coiled coils (6). On the other hand, if different targets are used by two partners to interact with a particular protein, it suggests that all three could be part of a single complex.

Third, the use of full-length ORFs in a two-hybrid assay can often obscure real interactions between proteins that can be detected by using protein fragments. The reasons for this are probably different on a case-by-case basis, but there are a number of plausible explanations. For example, interaction surfaces are not always accessible in an isolated protein subunit; an allosteric interaction may be required to reveal them. A well-known example is the interaction of the σ^{70} subunit of *Escherichia coli* RNA polymerase with promoter DNA (7). A domain in σ^{70} blocks its interaction with the promoter until it binds to core RNA polymerase. Fragments lacking the inhibitory domain bind promoter DNA. Interactions may also escape detection if either the bait or prey is

See companion article on page 13203.

*E-mail: jimhu@tamu.edu.

mislocalized by interactions with endogenous yeast proteins. Elsewhere, arguments have been presented for why this possibility of mislocalization suggests that searches for interactions among yeast proteins should use one of the newer bacterial two-hybrid systems (8). Whereas the directed approach does not completely obviate this problem, full-length ORFs present more ways for a fusion to be trapped by endogenous proteins through noncompetitive interactions with different domains.

Newman *et al.* point out that the yeast two-hybrid system may underrepresent homotypic interactions. A striking anomaly was their failure to observe a homotypic interaction for the best-studied coiled coil in the yeast proteome, the leucine zipper from Gcn4p. The rationale for this underrepresentation is based on a local concentration effect (Fig. 1). It's easy to forget that the DNA-binding proteins used to localize baits to a reporter gene are themselves dimeric. Barring steric constraints, a monomer bait that forms homodimers will prefer to bind to an identical bait on an adjacent subunit over a prey coming from solution. In addition, dimerization of the prey will decrease its effective concentration. My own work with leucine zippers began with the observation that λ repressor fusions provide a convenient genetic system to examine the formation of homodimers by the *GCN4* leucine zipper (9). At the time, we considered using the yeast two-hybrid system; given the result reported by Newman *et al.*, using the λ system was fortunate in hindsight. Recently, we've found that the coiled coil predicted by MULTICOIL in Bbp1p, but not observed as an interaction in Newman *et al.*, is among the homotypic oligomers detectable in repressor fusions (L. Mariño-Ramírez, unpublished data).

Despite any interactions that may have been missed, the data from Newman *et al.* provide a wealth of useful new information. Further analysis of the interactions is clearly needed, but the data already raise interesting questions. Obvious questions are: What do the complexes look like? Which ones form dimers, trimers, or higher order homotypic or heterotypic complexes? Are they parallel or antiparallel? The baits and preys tested range in size from 31 aa (one of two putative coiled coils from *SIN3*) to 712 aa (from *IMH1*). Many of the observed interactions involve predicted coiled coils of dramatically different sizes. Because by definition a coiled coil involves two or more helices, the longer subunit will contain significant amounts of unfolded polypeptide in the areas where it is not covered by the shorter subunit. Thus, the registration of the helices in the complexes is also an important unknown. Although biochemical and biophysical methods can be used to address these questions, the answers may come first from structural genomics projects.

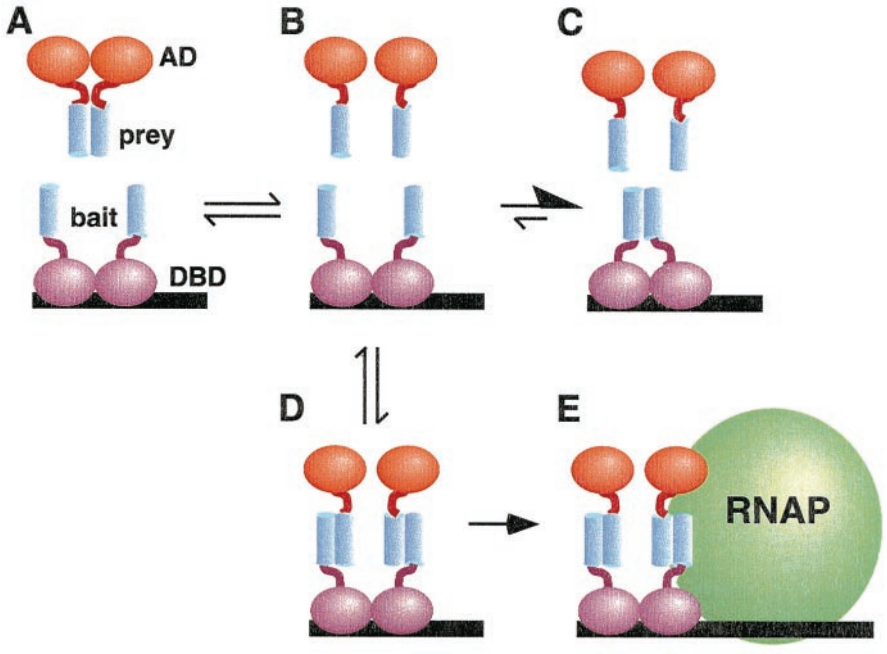

Fig. 1. Self-squelching of homodimers in a two-hybrid system. Hybrid proteins containing baits and preys from the same homodimeric protein are in equilibrium between species shown in *A*, *B*, and *C*. Only *B* is competent to form complex *D*, where the activation domain (AD) is recruited to the promoter. Only complex *D* can recruit the general transcription machinery (RNAP) to form complex *E*, which is transcriptionally activated. Because the DNA-binding domains (DBD) used in two-hybrid systems are homodimers, the local concentration of the bait favors the formation of complex *C*, which cannot recruit the activator-prey fusion. In addition, formation of complex *A* competes with formation of complex *B*.

The wealth of new data may be particularly useful for extending prediction algorithms beyond evaluating which sequences belong to a class of structures toward predicting the detailed partnering interactions themselves. Several groups have reengineered model coiled coils to generate new assembly specificities (10–15). It will be interesting to see whether the principles learned from protein design allow us to generate predictive algorithms that correctly match the different partnerships seen by Newman *et al.*

I thank Leonardo Mariño-Ramírez and Debby Siegele for useful discussions and suggestions. Work in my lab was supported by National Science Foundation Grant MCB-9808474 and an Advanced Research Program award (000517-0116-1997) from the Texas Higher Education Coordinating Board.

1. Uetz, P., Giot, L., Cagney, G., Mansfield, T. A., Judson, R. S., Knight, J. R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., *et al.* (2000) *Nature (London)* **403,** 623–627.
2. Ito, T., Tashiro, K., Muta, S., Ozawa, R., Chiba, T., Nishizawa, M., Yamamoto, K., Kuhara, S. & Sakaki, Y. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 1143–1147.
3. Fields, S. & Song, O. (1989) *Nature (London)* **340,** 245–246.
4. Newman, J. R. S., Wolf, E. & Kim, P. S. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 13203–13208.
5. Wolf, E., Kim, P. S. & Berger, B. (1997) *Protein Sci.* **6,** 1179–1189.
6. Lamb, P. & McKnight, S. L. (1991) *Trends Biochem. Sci.* **16,** 417–422.
7. Dombroski, A. J., Walter, W. A. & Gross, C. A. (1993) *Genes Dev.* **7,** 2446–2455.
8. Hu, J. C., Kornacker, M. G. & Hochschild, A. (2000) *Methods* **20,** 80–94.
9. Hu, J. C., O'Shea, E. K., Kim, P. S. & Sauer, R. T. (1990) *Science* **250,** 1400–1403.
10. O'Shea, E. K., Lumb, K. J. & Kim, P. S. (1993) *Curr. Biol.* **3,** 658–667.
11. Vinson, C. R., Hai, T. & Boyd, S. M. (1993) *Genes Dev.* **7,** 1047–1058.
12. Harbury, P. B., Zhang, T., Kim, P. S. & Alber, T. (1993) *Science* **262,** 1401–1407.
13. Monera, O. D., Kay, C. M. & Hodges, R. S. (1994) *Biochemistry* **33,** 3862–3871.
14. Oakley, M. & Kim, P. (1998) *Biochemistry* **37,** 12603–12610.
15. Zeng, X., Herndon, A. M. & Hu, J. C. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 3673–3678.