The Scientist
www.the-scientist.com

Volume 17 | Issue 2 | 34 | Jan. 27, 2003

Previous | Next

Unraveling Protein-Protein Interactions

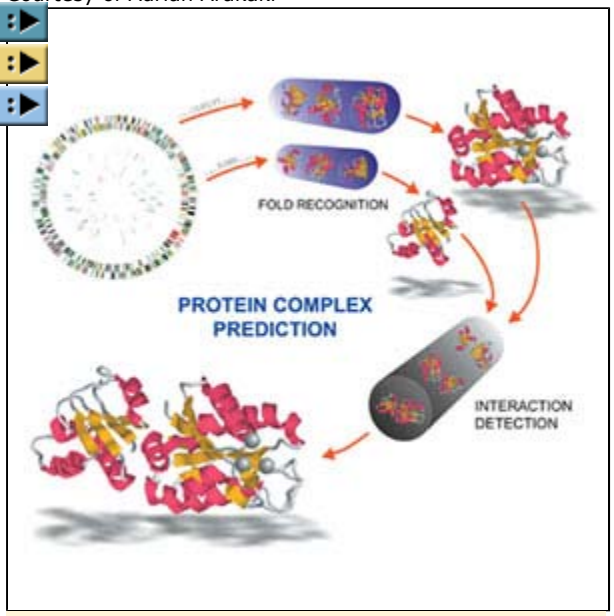
New genome-scale approaches expose the wiring of multiprotein complexes on a cellular scale | By Leslie Gray

Daily News
Upfront
Feature
Research
HotPapers
Lab Consumer
Profession
Archives

Search
::Go
Advanced Search

Off the Cuff
Snapshot

Courtesy of Adrian Arakaki



THERE'S GOLD IN THEM THERE COMPLEXES: Digging up protein-protein interactions with MULTIPROSPECTOR.

Using a computer instead of a pipette, Jeffrey Skolnick contemplates the subtle forces that bring proteins together. His first computational forays helped decipher the quaternary structure of proteins--the interactions between subunits in molecules such as tropomyosin. Now Skolnick, executive director of the Buffalo Center of Excellence in Bioinformatics, Buffalo, NY, works on a far bigger problem: understanding, modeling, and predicting the currently unfathomable rules of etiquette that govern protein alliances inside the cell on a genomic scale. "When

Tools
Comment
E-mail article

Article Extras
PDF:
The Six Degrees of CDC34?, (166K)

Need to download Adobe Acrobat?
Click here.

you look inside a cell, it's not like you are looking at an isolated individual trapped on a desert island," he says. "It's more like a crowded party on New Year's Eve."

A party of proteins, that is. And understanding what those proteins are doing involves much more than simply reading the guest list. That's because most of the heavy lifting in the cell is performed not by individual proteins, but by multimeric protein complexes. "There is much more structural and functional organization in a cell than traditionally believed," says computationalist Sandor Vajda of Boston University, "and these are not simply pair-wise interactions." In some instances, as many as 50 proteins interact to form one large, well-organized machine.

Such complexes make attractive drug targets for the biopharmaceutical industry. Yet the vast majority of proteins are known by their amino acid sequences alone; little, if any, information is known about their appearances or interactions. Traditional methods for detecting interactions in vivo, such as the venerable yeast two-hybrid screen, are effective but are low-throughput.[1] Researchers need new, faster ways of detecting and mapping protein interactions on a genomic scale, and scientists--both in academia and the commercial sector--have heeded the call.

New genome-wide, high-throughput, experimental and computational approaches have given scientists the ability to analyze interactions along the entire proteome in a single assay. The yeast two-hybrid protocol, still a

Detecting Interactions In Vitro

Sometimes researchers want to determine whether two proteins can interact, without placing them in a cellular milieu. A variety of in vitro methodologies exist, and one popular approach is the protein microarray.[1] The arrays consist of tens to thousands of proteins, individually spotted at unique addresses in a micro- or nano-scale matrix, so that interactions between the chip and the test samples can easily be identified.

Uppsala, Sweden-based Biacore's surface plasmon resonance technology, offers researchers another chip-based method to detect protein-protein interactions. Biacore offers a variety of surface chemistries, and though each chip can harbor only one target, the systems are capable of high-throughput processing of up to 192 chips in unattended runs.

contender thanks to robotic automation, has been joined by such new approaches as high-throughput mass spectrometry and "in silico" computational genome analysis.[2,3] Each of these methods has its strengths and weaknesses. In a recent review, Peer Bork of the European Molecular Biology Laboratory, Germany, and colleagues conclude: "To increase coverage and to improve confidence in detected or predicted protein interactions, as many complementary methods as possible should be used."[4]

Other approaches include phage display, far western blotting--in which a Western blot is probed with labeled proteins, glutathione-S-transferase "pull down" assays, and coimmunoprecipitation.

--Leslie Pray

1. A. Constans, "The chipping news," *The Scientist*, 16 [9]:28-31, April 29, 2002.

EXPERIMENTAL METHODS Two-hybrid systems were originally used to test for interactions between two known proteins--the so-called bait and prey proteins. More recently, it has been possible to screen for interactions between the bait and an entire library of cDNA prey,[1] using such commercial tools as: BD Matchmaker™ library construction and screening kits from BD Biosciences Clontech™ of Palo Alto, Calif.; Rockville, Md.-based OriGene's DupLEX-A™ Yeast Two-Hybrid System; Carlsbad, Calif.-based Invitrogen's ProQuest Two-Hybrid System; and La Jolla, Calif.-based Stratagene's HybiZAP®, CytoTrap®, and BacterioMatch™ systems.

All such kits have drawbacks, says Peter Uetz, group leader, functional genomics and proteomics, Institute for Toxicology and Genetics, University of Karlsruhe, Germany. Identifying the interacting proteins in traditional two-hybrid assays is a laborious, time-consuming task, explains Uetz, because each positive clone must be picked out individually and its prey DNA separately isolated and sequenced. In 2000, Uetz and collaborators described a faster and more efficient approach. They employed a colony array--a defined set of cloned prey open-reading frames or ORF fragments, in which interactions are detected by systematically mating the arrays to the bait strain.[5,6] With this approach, the user needs only to sequence the clones when the array is initially constructed. Thereafter, since the same array is used repeatedly, positive transformants can be readily identified without additional sequencing.

Sequencing aside, what really gives the array format its power, says Uetz, is its ability to be automated with robotics. Uetz and

his coauthors used the Beckman Coulter® Biomek 2000 Laboratory Automation Workstation.[5] The MegaMate system from UK-based Genetix is the only commercially available robotic two-hybrid array kit. A fully automated two-hybrid system, MegaMate screens up to 100,000 clones per day unattended. Up to 27,000 prey and bait colonies are robotically arrayed onto plates, 22 x 22 cm, and then replica-gridded directly on top of each other for mating. Positive colonies--those containing interacting proteins--can be picked at a rate of 4,000 colonies per hour.

Some companies offer robotic tools that can be used with extant two-hybrid systems. The VersArray Colony Picker and Arrayer from Hercules, Calif.-based Bio-Rad Laboratories, which generates and replicates ordered matrices of high-density colonies in an array format, can be applied to any array-based protein analysis, including two-hybrid assays. According to company literature, the arrayer can pick 9,000 colonies during one unattended run, at a rate of 576 colonies per hour.

Mass spectrometry is a relatively recent addition to the protein interaction detection arsenal.[7] Last January two independent groups described a genome-wide mass spectrometry method for characterizing distinct multiprotein interactions in *Saccharomyces cerevisisae*.[8-10] Between the two groups, thousands of interactions were identified using a baiting technique that involved attaching tags to hundreds of different bait proteins, introducing the library of bait clones into yeast, allowing the resulting protein complexes to form, fishing out the complexes using the tag, and finally identifying the isolated proteins with standard mass spectrometry methods. As with the genome-wide two-hybrid system, robotics made the assays high-throughput.[11]

Courtesy of Genetix

COMPUTATIONAL METHODS Some scientists are working out ways to screen entire genomes and predict protein- protein interactions from amino acid sequence information alone. David Eisenberg and Todd Yeates of the University of California, Los Angeles, and their collaborators, for instance, have developed several computational

The Genetix Megamate System

methods, including the phylogenetic profile and Rosetta Stone techniques.[3,12] Both approaches infer functional--not necessarily direct physical--linkage between protein pairs. Phylogenetic profiling, for example, infers functional linkage based on protein co-occurrences in multiple organisms. The idea, says Eisenberg, is that if two proteins always appear in the same organism together, the two are probably part of the same pathway.

One of the newest in silico methods for identifying protein partners is a threading approach developed by Skolnick and his collaborators, then at the Danforth Plant Science Center, St. Louis.[13] Traditional threading involves finding the closest match between a protein sequence of interest and a library of known structures. Skolnick and his colleagues have taken the technique one step further with their new algorithm, called MULTIPROSPECTOR (bioinformatics.buffalo.edu/proint), which identifies potentially interacting proteins using traditional threading and then, after finding matches, computes the interfacial energy between the protein pairs and rethreads both partners. The interfacial potential and threading score are used to determine whether the partners form a stable complex.

LINGERING DOUBT Some scientists are skeptical of the new computational approaches. They yield too many false positives, says Uetz, and their predicted interactions have yet to be verified experimentally--though he concedes the same criticism could be made of two-hybrid screens. Nonetheless, he says, "I appreciate these methods if they are applied cautiously."

Eisenberg says in silico predictions can be quite powerful, achieving perhaps 10 times more coverage than experimental approaches. "A direct experimental look is always a wonderful thing," he says. "But coverage is usually limited. Each protein requires a separate method. It is hard to get very broad coverage that way even though high-throughput methods are improving."

The computationalists themselves agree that, no matter how sophisticated their methodologies become, as Skolnick says, "at the end of the day you still need a carefully validated benchmark set where you really know that the interactions are there. Experiments are absolutely essential for that." That aside, he adds, "Computational methods are certainly no worse than experimental methods."

In silico or not, most protein biologists agree there is no single

best tool for detecting protein interactions. Even with the whole arsenal at one's disposal, the new genome-wide approaches reveal only whether an interaction exists and then, between which proteins. They say nothing about what the interaction looks like nor how its thousands of atoms fit together, let alone whether it is biologically important. Yet it is precisely these questions that drug designers must answer if genomic-level proteomics are to lead to new drug development.

Leslie Pray (lpray@nasw.org) is a freelance writer in Leverett, Mass.

References

1. B.A. Cunningham, "Finding a mate," *The Scientist*, 15[7]:26, April 2, 2001.

2. Y. Ho et al., "Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry," *Nature*, 415:180-3, 2002.

3. M. Pellegrini et al., "Assigning protein functions by comparative genome analysis: protein phylogenetic profiles," *Proc Natl Acad Sci*, 96:4285-8, 1999.

4. C. von Mering et al., "Comparative assessment of large-scale data sets of protein-protein interactions," *Nature*, 417:399-403, 2002.

5. P. Uetz et al., "A comprehensive analysis of protein- protein interactions in *Saccharomyces cerevisiae*," *Nature*, 403:623-7, 2000.

6. B.A. Maher, "A tale of two-hybrid," *The Scientist*, 16[10]:38-9, May 13, 2002.

7. J.M. Perkel, "Mass spectrometry applications for proteomics," *The Scientist* 15[16]:31, Aug. 20, 2001.

8. T. Ito et al., "A comprehensive two-hybrid analysis to explore the yeast protein interactome," *Proc Natl Acad Sci*, 98:4569-74, 2001.

9. A.-C. Gavin et al., "Functional organization of the yeast proteome by systematic analysis of protein complexes," *Nature*, 415:141-7, 2002.

10. A. Kumar, M. Snyder, "Proteomics: Protein complexes take the bait," *Nature*, 415:123-4, 2002.

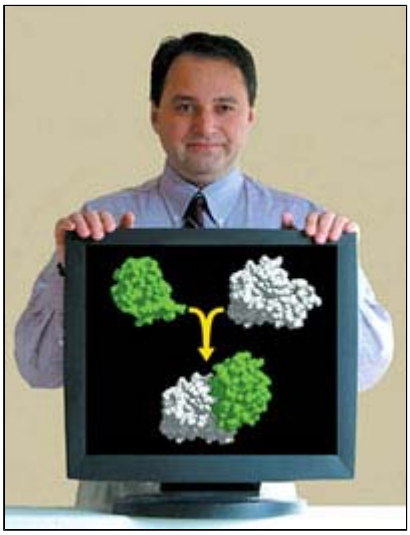
11. D. Wilkinson, "Pursuing proteomes," *The Scientist*, 14[12]:28, June 12, 2000.

12. E.M. Marcotte et al., "A combined algorithm for genome-wide prediction of protein function," *Nature*, 402:83-6, 1999.

13. L. Lu et al., "MULTIPROSPECTOR: An algorithm for the prediction of protein-protein interactions by multimeric threading," *Proteins*, 49:350-64, November 2002.

DOCKING PARTY

Courtesy of Ilya A. Vakser



Ilya A. Vakser

The three-dimensional atomic structure of a protein-protein complex provides crucial information for designing drugs that can fit into the nooks and crannies of the interaction interface. Typically researchers gather such information using X-ray crystallography,[1] but that technique has serious limitations, say Boston University's Sandor Vajda and Ilya Vakser of State University of New York at Stony Brook.

First, says Vakser, it is slow. "Crystallography can only produce a minute fraction of all the universe of protein complexes." The 20,000 protein structures currently stored in the Protein Data Base (www.rcsb.org/pdb) represent only a tiny fraction of the total number of known proteins. Second, most protein interactions are weak, transient complexes that cannot be crystallized. Computationalists are racing to develop reliable methods for theoretically predicting three-dimensional structures of interacting proteins using automated docking algorithms. (Docking is the process whereby two molecules fit together in three-dimensional space.)

Last September, 19 research teams put their approaches to the test. The groups submitted blind predictions of seven protein-protein complexes to independent assessors, who compared them to newly determined X-ray structures. The results were discussed at the First CAPRI (Critical Assessment of Predicted Interactions) Evaluation Meeting (capri.ebi.ac.uk).

Though the teams collectively achieved only about a 70% success rate, Vajda says that both the participation and results indicate rapid progress in the field. There is still a long way to go, however, before nonspecialists can use these algorithms reliably. And there is even further to go before the programs can be used genome-wide. That, says Jeffrey Skolnick of the Buffalo Center for Excellence in Bioinformatics, "is simply not practical today. It would take millions of days of computer time."

Currently, docking algorithms are used mainly to predict whether and how small molecules, such as drug candidates, interact with known protein targets. Available software titles include AutoDock

(www.scripps.edu/pub/olson-web/doc/autodock), DOCK (www.cmpharm.ucsf.edu/kuntz/dock.html), GOLD (www.ccdc.cam.ac.uk/prods/gold), and FLexX (cartan.gmd.de/flexx/). Programs written specifically for protein-protein interactions include DOT (www.sdsc.edu/CCMS/DOT), FTDOCK (www.bmm.icnet.uk/docking), GRAMM (reco3.ams.sunysb.edu/gramm/), and ZDOCK (zlab.bu.edu/~rong/dock/software.shtml).

Though docking programs are crucial for structure-based drug design, two-hybrid assays and their kin have a place in pharmaceutical research, too. Knowing which proteins are interacting, even if it is unclear what those interactions look like, is valuable information that narrows the playing field of drug targets to a much more manageable set of potential interaction partners.

--Leslie Pray

1. K. Miller, "Accelerating X-ray crystallography," *The Scientist,* 17[1]:39-41, Jan. 13, 2003.

Previous | Next