Protein-Protein Interactions

From Genome to Interactome

Genomic sequences specify protein sequences, but they don't necessarily clarify how proteins interact. Protein-protein interactions form the basis of many important activities, not only within cells, but in organisms as a whole, including antigen-antibody reactions, enzyme activation and inhibition, cellular signaling, and protein localization. Deciphering these interactions is challenging, in part due to the size and flexibility of many proteins of interest, the wide range of concentrations possible for partners in an interaction, the multiple pathways a given protein may affect, and the difficulties in predicting binding sites from structures. Approaches that are being employed include inhibiting specific protein-protein interactions by designing molecules that are "larger than the small molecules chemists love," according to Samuel H. Gellman, Professor, Department of Chemistry, University of Wisconsin, Madison, WI, and engineering new proteins containing interactive domains from other proteins and mining databases for potential interactions that can be confirmed experimentally.

Nature's Building Blocks

The approach taken by Gellman and his group is to design inhibitors of therapeutically relevant protein-protein interactions that he calls "foldamers," polymers with a strong tendency to adopt a specific, well-defined, compact conformation. This is useful where small molecules won't work, such as when the contact surface between the interacting proteins is large. The use of foldamers "embodies molecular design the way nature builds its most powerful molecules, nucleic acids and proteins," which are polymeric, Gellman notes. His group synthesizes building blocks of oligomers that fold into set shapes and display particular side chains and uses them to probe cellular functions. By using β -amino acids, which are not recognized by natural cellular proteases, and incorporating small ring structures, they have been able to synthesize short oligomers (β -peptides) that form stable helical structures in an aqueous environment. These helical β -peptides can be used as scaffolds to display functional groups mimicking protein surfaces

and structures and to replace one partner in an interactive pair. Two areas of research for Gellman's group concern members of the Bcl-2 family, which are involved in apoptosis (programmed cell death), and proteins participating in the membrane fusion steps that allow cytomegalovirus, a human pathogen in the herpesvirus family, to enter and infect cells.

Enough to Go Around

Although, like Gellman, Tony Pawson, Senior Investigator, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, ON, Canada, recognizes the therapeutic potential of understanding protein-protein interactions in human cells, he deems examining how these cells are organized interesting in their own right. His lab focuses on the role of particular protein domains in protein-protein interactions. Now that the human genome is sequenced, he notes, "one of the really interesting, obvious issues is how do so few genes specify the complicated behaviors that cells display? It seems like there is very little to go around, so how do you get such huge biological complexity with limited information?" He suggests that "everything must be combinatorial." That is, if each gene and its product only did one thing, there wouldn't be enough genes and proteins to do everything cells are capable of in all tissues and organs. Part of the answer lies in posttranslational modifications (e.g., phosphorylation, ubiquitination, acetylation, and SUMOylation), allowing a given protein to display different modifications for different cellular activities, states, or tissues. This allows a relatively simple initial repertoire of proteins or their component domains to expand into more complex, cell-specific interactions.

With a long-term goal of understanding how proteins form large networks, Pawson acknowledges that they are just starting to scratch the surface. Some of the technologic advances he sees as furthering the understanding of protein-protein interactions are chip-based approaches and using biological mass spectrometry, which he calls a "fantastic tool to identify proteins and

Courtesy of Jack Sadowsky, University of Wisconsin, Madison, WI.

Image of β -peptide foldamer docked into the BH3-recognition cleft on the protein Bcl-xL, generated as part of an effort to identify foldamers that inhibit interactions between Bcl-xL and BH3-containing proteins involved in the apoptosis pathway.

posttranslational modifications that one couldn't imagine a few years ago." A high-throughput tool called LUMIER, developed by a colleague, Jeff L. Wrana, Senior Scientists of Molecular Biology and Cancer, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, and a professor at the University of Toronto, is a mass spectroscopic-based approach that allows one to take a given protein and look at all of its binding partners. Pawson likens using it to being able to perform yeast two hybrid-like experiments in mammalian cells.

Work in Progress

"The notion of interaction is superficial, it's just the beginning step to understand what's going on in the cell," according to Lukasz Salwinski, UCLA-DOE Institute for Genomics and Proteomics, University of California, Los Angeles, who maintains DIP, the Database of Interacting Proteins. It's one thing to know that actin and myosin interact, he observes; the next level is to understand how they form a working muscle. DIP started around 1999 as an in-house research project for calibrating protein interaction prediction methods, soon went public, and is available online, containing nearly 56,000 interactions involving 19,000 proteins from over 100 organisms, with an aim of integrating experimental evidence on protein-protein interactions into a single, easily accessible database. "It's pretty painful to run a database," he admits, "and it would be pretty annoying to be doing this if no one cares." With recent funding from NIH and the hope of getting more people involved, Salwinski expects data entry, the most labor-intensive part of the operation, will become more efficient. Several steps are involved in increasing efficiency, the first of which is using a standard format for data exchange between databases; the second is gathering all the active public databases together in a common curation pipeline to avoid duplication. Next, it will be necessary to persuade the community to provide data in a preprocessed form. This is being done through efforts to get the editors of journals to buy into a common system, requiring authors to provide a specified

amount of information about their protein interaction system and in a format that can be plugged into databases, not unlike what is now in place for genomic and other sequence data. "The first thing to realize," notes Salwinski, "is that information about protein-protein interactions is very incomplete. We're not getting all the information, even in yeast." Still, he says, the quality of data from database analysis is better than from direct experimentation, but not yet complete enough to draw direct conclusions.

The Human Proteome Organisation Proteomics Standards Initiative (HUPO-PSI) was formed in 2001 to consolidate proteome organizations into a single worldwide body. As Sandra Orchard, Senior Scientific Database Curator, EMBL-EBI, Wellcome Trust Genome Campus, Cambridge, UK, observes, HUPO tissue groups, like proteomics groups, required standards to allow both exchange of data and public repositories where reference data sets could be stored, downloaded, and compared with those experimentally generated locally.

In the field of molecular interactions, data providers worked together to produce a common XML interchange format populated by controlled vocabularies to remove ambiguities when describing methodologies. The first implementation was published in 2004, allowing exchange of basic information on protein-protein interactions. The current implementation, in preparation, will add further complexity and flexibility, for example including the ability to deal with other interactors, such as nucleic acids or small molecules.

The data exchange procedures and infrastructure of the International Molecular Exchange Consortium (IMEx), which is an agreement among major public domain molecular interaction databases to share curation effort and exchange completed records in a manner similar to the successful global collaboration achieved by the nucleotide sequence databases, are currently still under development. The intention is to implement the exchange of curated records fully this year.

Future Directions

Gellman suggests that using foldamers may lead to identification of therapeutic targets and approaches, particularly in situations where pathology may be the result of aberrant cellular signaling or other cell surface interactions (e.g., in viral infections, cancer, and autoimmune diseases). "I'm not predicting that foldamers will be useful as therapeutics, but rather as tools," he says, adding, "I will be delighted either way."

Pawson sees future challenges in moving from static maps of protein interactions that can be drawn today, to creating dynamic pictures of how signals move in and out of cells, and understanding how different cell types interact. Ultimately, to understand how a human cell is organized, the whole network, including not just proteins but phospholipids, nucleic acids, and small molecules, will need to be integrated. "Finally, the issue is whether this will tell you anything about disease and how to treat it," he notes.

There are a few drugs available now that exploit protein-protein interaction, e.g., imatinib mesylate, which inhibits the abnormal bcr-abl tyrosine kinase in patients with Philadelphia chromosome-positive chronic myeloid leukemia. Rapamycin and its relatives, including mammalian targets of rapamycin (mTOR), which are serine-threonine kinases, have antibiotic, immunosuppressive, and anti-neoplastic activities that may result from their ability to form complexes containing proteins that would normally not interact with each other in the cell. One may expect that additional promising compounds affecting protein-protein interactions are in development or will be soon.

Lynne Lederman is a freelance medical writer in Mamaroneck, NY.