Structure
Previews

Searchlight on Domains

Monica Riley[1,*]
[1] Bay Paul Center, Marine Biological Laboratory, Woods Hole, MA 02543, USA
*Correspondence: mriley@mbl.edu
DOI 10.1016/j.str.2006.12.003

In this issue of *Structure*, Bashton and Chothia (2007) examine in detail the functions of selected domains within proteins both when they are alone and when in combination with others. Domain function is relevant to molecular evolution and to annotation of proteins known only by sequence.

In the changing world of biological science, disciplines are interconnected such that information on protein structure applies beyond the field itself to questions of molecular evolution and to prediction of properties of proteins known only from genome sequences. The report by Matthew Bashton and Cyrus Chothia in this issue asks how domains affect each other in multidomain proteins, finding that often they interact in ways that create related but new functional attributes.

Molecular evolution of proteins is thought of as having used two main paths to create variety and new function over time: (1) duplication of an entire protein with subsequent functional divergence of the copies, and (2) recombination and combination within proteins to reassort functional domains. Bashton and Chothia (2007) use the SCOP superfamily database (in which the domains are defined as evolutionary elements organized by their structure, function and sequence) to look closely at the effects of combining domains, one domain modifying the function of another through broadly categorized interactions.

In this meticulous and detailed study, the authors assembled 45 sets of proteins, each containing a multidomain protein and corresponding homologous 1 domain proteins. The sets were chosen so that functional information was available both for the domains in the multidomain protein and for at least one of the homologous 1 domain proteins. In some cases, the combination and interaction of domains changes the function of a single domain to a related but different one in the context of the more complicated protein. A common change is substrate specificity where a domain recognizes one substrate when in a 1 domain protein, but changes to recognize a similar but different substrate when spatially altered by interaction with another domain in a multidomain protein.

There are several ways that domain interaction affects the action of a single domain. One case describes two amidating enzymes, a 1 domain and a 2 domain. The domain in common is specific for aspartic acid in one enzyme and aspartyl-tRNA in the other. Both enzymes use an aminoacyl-adenylate intermediate as amide donor, but the catalytic action differs in that the amidation takes place at the β-carboxylate for one and the α carboxylate for the other. The function of the catalytic sites is related but different in the two proteins. In the 2 domain protein, the geometry of the structure has been affected by the second domain. Besides changing the specificity of a binding site, examples are given of changing the number or kind of subunit contacts, setting the spacing and orientation of two other domains and participating in an activation process that enables substrate binding or active site function through conformation changes.

A complex case of alternative binding of polypeptides effects the transfer of copper from a chaperone to the active site of superoxide dismutase (SOD). Both the chaperone and the enzyme are homodimers. The central domain of SOD contains both the active site and the binding site for dimerization. The homologous domain in the chaperone has no catalytic activity, but shares the binding site. A chaperone monomer binds to an SOD monomer, forming a heterodimer through the dimerization interface, which leads to major conformational changes and sets the stage for transfer of a copper atom from the chaperone to the catalytic site of the SOD monomer. After dissociation of the chaperone, upon reforming the SOD homodimer, the essential copper has activated the enzyme (see Figure 5 in Bashton and Chothia, 2007).

Clearly the study of interacting domains has relevance to the second kind of evolution mentioned above: the generation of new function by reassorting domain partners. Simple reassortment will generate new molecules, but beyond that we see that the interactions between domains can bring about change in the properties of domains so that new domain functions also emerge.

The work also applies to the practice known as annotation, by which functions of unknown proteins are predicted from their sequences. More detailed information about domains and their interactions can fruitfully transfer to this arena. Today in the genomic era, we are flooded with protein sequences derived from genomic DNA sequences. Prediction of unknown protein function done either with automated systems or manually is an art that is constantly improving. However, the methods of gene annotation have long been understood by its practitioners to be crude (e.g., Galperin and Koonin, 1998). A valuable type of information in this context is identification of domains in a protein by sequence and knowledge of domain function to help predict the protein function.

There are many systems for characterizing shorter and longer sequences within proteins as carrying out specific functions. The Uni-Prot knowledgebase (http://www.expasy.ch/)

provides information on motifs and domains within each protein sequence, drawn from many sources. Functional units within proteins have been identified and grouped into families and superfamilies using various criteria, appearing in a variety of classifications such as Pfam (Finn et al., 2006), Worldwide PDB (Berman et al., 2006), CCD (Marchler-Bauer et al., 2006), SCOP (Wilson et al., 2006), and many others (see UniProt). The addition of information on the consequences of interactions between SCOP superfamily domains adds yet deeper and more complex understanding of domain function, which no doubt will be incorporated into the practice of the prediction of protein function from sequence.

In order to improve application of knowledge about domains to genome annotation work, a useful step would be to expand databases to provide explicit information on domain function. In bridging biological fields, there is a problem in communicating information in a form that can be used by the nonexpert genome annotator. Some domains are well known and well described; for example, the various NAD(P)H binding site domains. However, in the SCOP superfamily listings, domains are systematized by alphanumeric coding and are also given brief names: sometimes mnemonics, sometimes understandable abbreviations, and sometimes seemingly opaque labels. PFAM also assigns names, but these are often specific to the function in the protein(s) studied first and might not be literally transferable. Consequently, the nonspecialist may not understand the biological activity of most domains by their labels. Over time and for the sake of the scientific community, descriptions will need to be expanded in an effort to inform the nonspecialist about attributes expected of a domain in an unknown protein, making better use of one biological field's knowledge for another.

Thus, the elegant paper by Bashton and Chothia provides highly specific information about domains and their interactions, particularly for multidomain proteins. In some cases, domains did not change their actions when combined in multidomain proteins, but in other cases, a variety of effects on function resulted from interaction. The detailed information provided in these many examples contributes not only to the field of structural chemistry of proteins, but it also presages the kind of careful and detailed information that will accelerate our understanding of evolutionary mechanisms and will aid the practice of predicting functions of unknown proteins from their sequences.

REFERENCES

Bashton, M., and Chothia, C. (2007). Structure *15*, this issue, 85–99.

Berman, H., Henrick, K., Nakamura, H., and Markley, J.L. (2006). Nucleic Acids Res., in press. Published online November 16, 2006. 10.1093/nar/gkl971.

Finn, R.D., Mistry, J., Schuster-Bockler, B., Griffiths-Jones, S., Hollich, V., Lassmann, T., Moxon, S., Marshall, M., Khanna, A., Durbin, R., et al. (2006). Nucleic Acids Res. *34*, D247–D251.

Galperin, M.Y., and Koonin, E.V. (1998). In Silico Biol. *1*, 55–67.

Marchler-Bauer, A., Anderson, J.B., Derbyshire, M.K., Deweese-Scott, C., Gonzales, N.R., Gwadz, M., Hao, L., He, S., Hurwitz, D.I., Jackson, J.D., et al. (2006). Nucleic Acids Res., in press. Published online November 29, 2006. 10.1093/nar/gkl951.

Wilson, D., Madera, M., Vogel, C., Chothia, C., and Gough, J. (2006). Nucleic Acids Res., in press. Published online November 10, 2006. 10.1093/nar/gkl910.

Tying the Knot That Binds

Lisa M. Gloss[1,*]
[1] School of Molecular Biosciences, Washington State University, Pullman, WA 99164, USA
*Correspondence: lmgloss@wsu.edu
DOI 10.1016/j.str.2006.12.001

The recent determination of protein structures with knots in their backbone topology has defied previous conventional wisdom. How proteins can fold with a knot is an intriguing question that has been explored for YibK from *Haemophilus influenzae* in this issue of *Structure* (Mallam and Jackson, 2007a).

It has been over 40 years since Anfinsen and colleagues demonstrated that a protein's sequence contains all the necessary information to determine its structure, stability, and folding mechanism. Deciphering how this information is encoded by the sequence is a holy grail of structural biology. Since the mid-90s, many efforts have focused on studying small, single domain, monomeric proteins. These simple structures often fold by two-state kinetic mechanisms, with no transiently populated intermediates (Jackson, 1998). These experimental systems are also amenable to detailed computational studies, and this synergy has provided new insights and

