The Porous Borders of the Protein World

Matthew H.J. Cordes[1,*] and Katie L. Stewart[1]
[1]Department of Chemistry and Biochemistry, University of Arizona, Tucson, AZ 85721, USA
*Correspondence: cordes@email.arizona.edu
DOI 10.1016/j.str.2012.01.015

Fold switching may play a role in the evolution of new protein folds and functions. He et al., in this issue of *Structure*, use protein design to illustrate that the same drastic change in a protein fold can occur via multiple different mutational pathways.

The vast structural repertoire of proteins underlies, at least in part, their role as the primary functional expression of genetic information, as well as their susceptibility to harmful misfolding. The different types of protein structure, or folds, number in the thousands, but the mechanisms by which that diversity originated and evolved are shrouded in mystery. Some folds may have no definable common ancestry, and instead arose independently. Recent studies of homologous sequences with different folds (Bolgurov et al., 2007; Roessler et al., 2008), however, and of individual proteins that can switch fold in the course of normal function (Tuinstra et al., 2008), support an emerging view that some structures could have evolved from others through fold switching mediated by sequence mutation (Murzin, 2008). In this issue of *Structure*, He et al. (2012) illuminate multiple pathways by which evolutionary fold switching might occur, using protein design to find not one or two, but three different sequence mutations that connect two very different protein folds. The unique construction of their system also allows these researchers to track changes in function associated with fold switching.

Protein evolution is a combination of freedom and restriction. Amino-acid sequence encodes protein structure, and a stable protein structure can act as a scaffold for function, though it is not an absolute prerequisite. Each protein fold can be specified by a huge number of sequences, many of them interconnected by simple pathways of mutation, allowing a sequence some freedom to drift while retaining the same stable structure. The vast majority of possible sequences, however, probably do not encode folded or functional proteins at all, and if one mapped stable protein structures onto the space of all possible sequences, the great majority of that sequence space would be empty. One could easily hypothesize that, within this vast void, galaxies of sequences encoding different folds would be confined and well separated from each other, hindering fold switching. Experimental and theoretical studies, however, suggest that differently folded regions of sequence space may approach, intersect, or overlap each other, allowing fold switching to occur without large jumps in sequence (Bryan and Orban, 2010; Meyerguz et al., 2007).

In a series of studies dating from 2007 (Alexander et al., 2007), Orban, Bryan, and coworkers have used a design approach to seek out these borderlands. Beginning with two natural sequences from differently folded domains G_A (3α) and G_B ($4\beta+\alpha$) within *Streptococcus* cell surface protein G, they engineered the two sequences to be more similar by incorporating elements from each sequence into the other, while retaining the original folds and preserving functional epitopes for ligand binding. Now, in a pièce de résistance, He et al. (2012) report structural data for four G_A/G_B variants related by a series of three single amino-acid substitutions, and use NMR spectroscopy to demonstrate that each successive substitution switches the fold. The resulting mutational pathway crisscrosses the border between the two folds three times, showing that multiple simple routes exist for changing the structure (see Figure 1).

Each substitution achieves complete shifting of the structure within detection limits, except that in G_B98-T25I, the alternate fold is populated at 5% ($\sim$2 kcal/mol difference). The single substitutions appear to both stabilize one fold and destabilize the other, leading to large relative free energy changes (>4 kcal/mol). Thus, the mutational pathways traversed here cross fairly well-delineated borders between folds, rather than regions encoding both. Surprisingly, however, all four sequences bind the ligand (IgG) associated with G_B better than they bind the ligand (HSA) for G_A, though one sequence binds both ligands. These behaviors probably owe to ligand-induced fold switching combined with degradation of the HSA-binding function, although this is not explored in depth here. Because of such factors, the borders between folds, and the borders between functions linked to those folds, may not always coincide.

The term "marches" or "marchlands" can refer to tracts of borderland claimed by two countries. In the protein world, one equivalent of a march might be a set of similar sequences, or a region of sequence space, encoding two different detectable folds and/or two detectable associated functions. Some of the sequences characterized by He et al. (2012) stand at the edge of, or arguably within, such a marchland. It seems only a matter of time before this research team reports sequences that have very small free energy differences between the two folds (<1 kcal/mol). Most of the switch substitutions in the current work are on the periphery of the hydrophobic cores; achieving a fine balance between the two folds might be aided by further substitution of residues that are highly solvent-exposed in one structure or the other. In Arc repressor, for example, population of both the native fold and an alternate fold resulted from a surface polar-to-hydrophobic substitution that stabilized the alternate fold but did not appear to destabilize the native conformation (Cordes et al., 2000).

Examples of fold change in design, mutagenesis, function, and evolution involve different degrees of topological change, ranging from substitutions of one or two secondary structure elements to replacement of entire subdomains, to wholesale domain remodeling (Bryan and Orban, 2010). Designs, including those of He et al. (2012) on the protein G domains, often involve more dramatic, comprehensive rearrangements than natural fold switches. Although artificial in nature, these designed systems have a great deal to tell us about the plausibility of large-scale structural changes as well as potential pathways of sequence change connecting structures. He et al. (2012) note that a distant evolutionary relationship between G_A and G_B remains possible. The NusG/RfaH system is an apparent case of a complete structural rearrangement in natural domain fold evolution (Belogurov et al., 2007).

He et al. (2012) and many other switch and chameleon design studies (Anderson et al., 2011; Ambroggio and Kuhlman, 2006) utilize mostly or entirely the binary sequence space defined by a pair of aligned protein sequences. Because actual sequence space is vastly larger, these proteins are designed with one proverbial hand tied behind the back. Considering the expanded possibilities inherent in higher dimensional explorations of sequence space (Ambroggio and Kuhlman, 2006), and given the increasing number of examples of natural fold switching, the future seems wide open. The fog is lifting to reveal the beautiful and rugged, yet passable, borderlands of the protein universe. Some proteins may stay clear of the marchlands during their evolutionary travels; others may wander through; still others may remain in these in-between places.

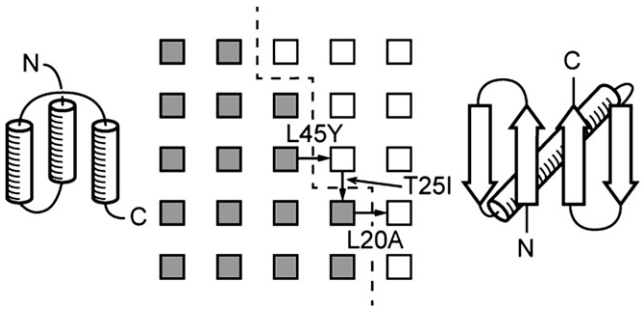

Figure 1. Multiple Mutational Pathways for Fold Switching
Crossing back and forth between 3α (left) and 4β+α (right) folds via a series of single amino-acid substitutions at different positions in the protein sequence. Shaded boxes represent sequences encoding the 3α fold; open boxes represent sequences encoding the 4β+α fold.

REFERENCES

Alexander, P.A., He, Y., Chen, Y., Orban, J., and Bryan, P.N. (2007). Proc. Natl. Acad. Sci. USA *104*, 11963–11968.

Ambroggio, X.I., and Kuhlman, B. (2006). Curr. Opin. Struct. Biol. *16*, 525–530.

Anderson, W.J., Van Dorn, L.O., Ingram, W.M., and Cordes, M.H. (2011). Protein Eng. Des. Sel. *24*, 765–771.

Belogurov, G.A., Vassylyeva, M.N., Svetlov, V., Klyuyev, S., Grishin, N.V., Vassylyev, D.G., and Artsimovitch, I. (2007). Mol. Cell *26*, 117–129.

Bryan, P.N., and Orban, J. (2010). Curr. Opin. Struct. Biol. *20*, 482–488.

Cordes, M.H., Burton, R.E., Walsh, N.P., McKnight, C.J., and Sauer, R.T. (2000). Nat. Struct. Biol. *7*, 1129–1132.

He, Y., Chen, Y., Alexander, P.A., Bryan, P.N., and Orban, J. (2012). Structure *20*, this issue, 283–291.

Meyerguz, L., Kleinberg, J., and Elber, R. (2007). Proc. Natl. Acad. Sci. USA *104*, 11627–11632.

Murzin, A.G. (2008). Science *320*, 1725–1726.

Roessler, C.G., Hall, B.M., Anderson, W.J., Ingram, W.M., Roberts, S.A., Montfort, W.R., and Cordes, M.H. (2008). Proc. Natl. Acad. Sci. USA *105*, 2343–2348.

Tuinstra, R.L., Peterson, F.C., Kutlesa, S., Elgin, E.S., Kron, M.A., and Volkman, B.F. (2008). Proc. Natl. Acad. Sci. USA *105*, 5057–5062.

Capturing a Virus while It Catches Its Breath

Theodore C. Pierson[1,*] and Richard J. Kuhn[2,*]
[1]Viral Pathogenesis Section, Laboratory of Viral Diseases, National Institutes of Health, Bethesda, MD 20892, USA
[2]Markey Center for Structural Biology, Department of Biological Sciences, Purdue University, West Lafayette, IN 47907, USA
*Correspondence: piersontc@mail.nih.gov (T.C.P.), kuhnr@purdue.edu (R.J.K.)
DOI 10.1016/j.str.2012.01.014

The four serotypes of dengue virus present a formidable challenge for the development of efficacious human vaccines. Cockburn and colleagues, in this issue of *Structure*, describe the structural basis of a cross-reactive neutralizing antibody, providing greater insight into immune protection and pathogenesis.

Dengue virus (DENV) is a mosquito-borne member of the flavivirus genus responsible for roughly 50 million human infections each year. Four antigenically-related serotypes of DENV circulate in tropical and subtropical regions of the globe. Infection by any of the four DENV serotypes may cause a self-limiting febrile illness that is rarely fatal, and is thought to confer immunity to reinfection by the same serotype of DENV. However, the secondary infection of a DENV-seropositive individual with a heterologous DENV serotype may lead to more severe clinical manifestations of disease, including

