Brief Report

Better theoretical models and protein design experiments can help to understand protein folding

Konda Mani Saravanan, Samuel Selvaraj[1]

Centre of Excellence in Bioinformatics, School of Biotechnology, Madurai Kamaraj University, Madurai, [1]Department of Bioinformatics, School of Life Sciences, Bharathidasan University, Tiruchirappalli, Tamil Nadu, India

Address for correspondence:
Dr. Samuel Selvaraj, Department of Bioinformatics, School of Life Sciences, Bharathidasan University, Tiruchirappalli - 620 024, Tamil Nadu, India. E-mail: selvarajsamuel@gmail.com

Abstract

In our study, we have concluded that two proteins with 88% homology choose different energetically favorable pathways in the very early stage of the folding process to attain their native folds. Subsequent reports from other investigators by performing folding and unfolding kinetics experiments concur with our findings. We herewith discuss the key papers revealing computational and experimental analysis of two designed proteins with similar sequence distant folds. Further we suggest that the theoretical/computational analysis of protein sequences and structures along with the relevant experiments provide a better understanding of the relationship between protein sequence, folding, and structure.

Key words: Designed proteins, molecular dynamics and simulations, protein folding, sequence and structural analysis, theoretical models

CASE REPORT

In recent years, coupling of theoretical and experimental approaches in the study of protein folding has resulted in providing fruitful clues. Experimental and computational protein design provides vital clues to understand the protein folding process, and it is of considerable interest in the area of protein science to engineer proteins with novel folds and desired functions. The field of protein design has a unique history where researchers from diverse discipline come together to explore novel catalytic, pharmaceutical, structural, and sensing properties of amino acids in proteins.[1]

Interestingly, a designed eleven amino acid sequence folded as a helix in one position and as a sheet in another position in the protein sequence.[2] This work has enabled to explore the role of nonlocal interactions in the formation of secondary structure. Subsequently, helices were transmuted into sheets to understand the conformation change phenomenon and illustrate that not all the amino acids play an equal role in specifying a fold.[3]

Utilizing the knowledge offered by several protein design groups, Kuhlman *et al.*[4] in 2003 have computationally designed a 93-residue α/β protein called Top7 and found that the protein could be experimentally folded and extremely stable. This pioneering work has enabled further research to understand the contribution of each amino acid residue in a protein to adopt a certain fold. Hence, emphasizing that protein design could be a powerful experiment to understand the processes that underlie conformational plasticity in proteins. Considering these facts, by using protein design and mutation experiments Alexander *et al.*[5-8] explored how two proteins with almost similar amino acid sequences change their fold and function.

Following the contribution of various theoretical and experimental protein science research groups, several such engineered proteins with selective folding and distinctive function are designed.[7,8] Nevertheless, the design of such a pair of proteins with high sequence identity with completely different topologies can be viewed as a challenge to the well-accepted paradigm that similar sequences always tend to fold into similar three-dimensional structures.

An analysis of the literature reveals that the design of two highly identical proteins with different folds and functions is challenging and time bound[5,6] as shown in [Table 1].

Table 1: Literature review of design of two proteins with high sequence identity adopting different folds

Paper title	Journal name	Published year	Sequence identity %
Folding and binding reactions for a pair of homologous heteromorphs is presented	Biochemistry	2005	59
Solution NMR structures of pair of homologous heteromorphs (55% identity) is presented	Biochemistry	2005	59
Design and characterization of pair of homologous heteromorphs (88% identity) is presented	Proc. Natl. Acad. Sci. USA	2007	88
Solution NMR structures of pair of homologous heteromorphs (88% identity) is presented	Proc. Natl. Acad. Sci. USA	2008	88
Design and characterization of pair of homologous heteromorphs (98% identity) via single amino acid mutation is presented	Proc. Natl. Acad. Sci. USA	2009	98.21

Homologous heteromorphs means sequence-similar, distant-structure proteins. NMR: Nuclear magnetic resonance

Streptococcus protein G contains two types of domains (GA and GB) that bind to serum proteins in blood. The natural versions of GA and GB domains share no significant sequence homology and have different folds, 3 − α and 4β + α, respectively. From the above two parent proteins, high-identity versions of GA and GB were synthesized.[5,6,9] Interestingly, small and critical differences in the sequences of the two proteins determine the topology of the protein early on the folding pathway.In addition, two proteins named G_A88 (PDB ID: 2JWS) and G_B88 (PDB ID: 2JWU) by mutation experiments from the *Streptococcus protein G* with 88% sequence identity adopt different structures and functions and these proteins are valuable tools to understand the contribution of residues to adopt a particular fold.[7] These two proteins vary only at seven positions out of 56 amino acids, which are shown in [Figure 1]. This design has made a breakthrough in the field of protein science and contradicts the general statement that "similar sequences adopt similar structures." Following this, we have carried out computational sequence and structural analysis on these two designed proteins.[10] We have performed secondary structure prediction of these two proteins and observed that the methods such as multivariate linear regression combiner can predict some regions as extended structures for the helical protein sequence GA, which gave us a clue that there may be structural plasticity at the region of first 15 residues, which are identical in the both proteins.[10] We also discovered some patterns in the nonidentical positions of two proteins with a rare combination of residues that are not present in any publicly available sequence databases. By analyzing the structures of the two designed proteins, we predicted nucleation sites at various positions in the sequence, which may start or terminate secondary structural elements (helix, sheet, and coil). We also observed drastic difference in the surrounding environment of nonidentical residues (7 out of 56) and difference in interaction energy. By observing the structural plasticity at the amino and carboxyl terminal of the sequences of two designed proteins and the influence of surrounding environment

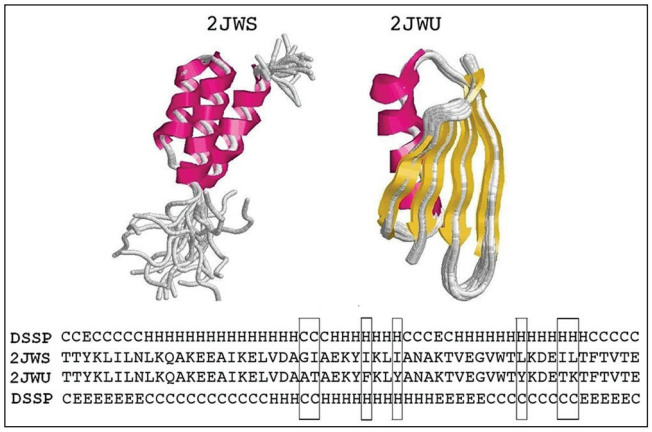

Figure 1: Sequence, dictionary of secondary structure of proteins assigned secondary structures and tertiary structures of pair of homologous heteromorphs (the seven residues that vary in both sequences are indicated in rectangular boxes)

of each residue, we concluded that early on during the process of folding, both proteins may choose different energetically favorable pathways to attain the different folds.[10]

Other researchers have characterized the folding of these two proteins using biophysical and computational experiments.[11] They also indicated that the final native structures of these proteins were dictated very early along the folding pathway by performing equilibrium unfolding of G_A88 and G_B88, folding and unfolding kinetics and molecular dynamics simulations experiments. Concurrently, energy calculations were performed on the two designed proteins in a vacuum,[12] which indicated that current computer modeling/simulations experiments cannot explain why two highly similar sequences fold into different structures. However, it was suggested that improved modeling/simulations tools should be developed to predict the pair of sequences with different structures, which differ, by only few residues.

In a recent study,[13] folding and unfolding kinetics experiments performed on these two designed proteins indicated a detectable residual structure in the denatured

state of G_B88 whereas the denatured state of G_A88 is unstructured. Interestingly, they explored these two proteins by Φ value analysis based on 132 site directed mutants and concluded that the protein's topology is committed very early along the folding pathway.

Based on the above studies, we suggest that, along with the suitable protein design experiments, better theoretical models including folding simulations coupled with structure prediction and sequence search in databases can shed light on the phenomenon of protein folding and conformation switching which may ultimately lead us to understand the contribution of each amino acid in these proteins to adopt a specific fold.

ACKNOWLEDGMENTS

K.M. Saravanan thanks University Grants Commission for the award of Dr. D.S. Kothari Post-Doctoral Fellowship (Grant No. F.13-932/2013 (BSR)).

REFERENCES

1. Samish I, MacDermaid CM, Perez-Aguilar JM, Saven JG. Theoretical and computational protein design. Annu Rev Phys Chem 2011;62:129-49.
2. Minor DL Jr, Kim PS. Context-dependent secondary structure formation of a designed protein sequence. Nature 1996;380:730-4.
3. Dalal S, Balasubramanian S, Regan L. Transmuting alpha helices and beta sheets. Fold Des 1997;2:R71-9.
4. Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, Baker D. Design of a novel globular protein fold with atomic-level accuracy. Science 2003;302:1364-8.
5. Alexander PA, Rozak DA, Orban J, Bryan PN. Directed evolution of highly homologous proteins with different folds by phage display: Implications for the protein folding code. Biochemistry 2005;44:14045-54.
6. He Y, Yeh DC, Alexander P, Bryan PN, Orban J. Solution NMR structures of IgG binding domains with artificially evolved high levels of sequence identity but different folds. Biochemistry 2005;44:14055-61.
7. He Y, Chen Y, Alexander P, Bryan PN, Orban J. NMR structures of two designed proteins with high sequence identity but different fold and function. Proc Natl Acad Sci U S A 2008;105:14412-7.
8. Alexander PA, He Y, Chen Y, Orban J, Bryan PN. A minimal sequence code for switching protein structure and function. Proc Natl Acad Sci U S A 2009;106:21149-54.
9. Scott KA, Daggett V. Folding mechanisms of proteins with high sequence identity but different folds. Biochemistry 2007;46:1545-56.
10. Saravanan KM, Balasubramanian H, Nallusamy S, Samuel S. Sequence and structural analysis of two designed proteins with 88% identity adopting different folds. Protein Eng Des Sel 2010;23:911-8.
11. Morrone A, McCully ME, Bryan PN, Brunori M, Daggett V, Gianni S, *et al*. The denatured state dictates the topology of two proteins with almost identical sequence but different native structure and function. J Biol Chem 2011;286:3863-72.
12. Allison JR, Bergeler M, Hansen N, van Gunsteren WF. Current computer modeling cannot explain why two highly similar sequences fold into different structures. Biochemistry 2011;50:10965-73.
13. Giri R, Morrone A, Travaglini-Allocatelli C, Jemth P, Brunori M, Gianni S. Folding pathways of proteins with increasing degree of sequence identities but different structure and function. Proc Natl Acad Sci U S A 2012;109:17772-6.

How to cite this article: Saravanan KM, Selvaraj S. Better theoretical models and protein design experiments can help to understand protein folding. J Nat Sc Biol Med 2015;6:202-4.

Source of Support: Nil. **Conflict of Interest:** None declared.

Access this article online	
Quick Response Code:	Website: www.jnsbm.org
	DOI: 10.4103/0976-9668.149122