Sequence and structural analysis of two designed proteins with 88% identity adopting different folds

Saravanan K. Mani, Harihar Balasubramanian, Saranya Nallusamy and Selvaraj Samuel

Department of Bioinformatics, School of Life Sciences, Bharathidasan University, Tiruchirappalli, TN 620024, India

1To whom correspondence should be addressed. E-mail: sel_emi@yahoo.co.uk

Received April 16, 2010; revised August 2, 2010; accepted September 12, 2010

Edited by P. Balaram

Protein folding is a natural phenomenon by which a sequence of amino acids folds into a unique functional three-dimensional structure. Although the sequence code that governs folding remains a mystery, one can identify key inter-residue contacts responsible for a given topology. In nature, there are many pairs of proteins of a given length that share little or no sequence identity. Similarly, there are many proteins that share a common topology but lack significant evidence of homology. In order to tackle this problem, protein engineering studies have been used to determine the minimal number of amino acid residues that codes for a particular fold. In recent years, the coupling of theoretical models and experiments in the study of protein folding has resulted in providing some fruitful clues. He et al. have designed two proteins with 88% sequence identity, which adopt different folds and functions. In this work, we have systematically analysed these two proteins by performing pentapeptide search, secondary structure predictions, variation in inter-residue interactions and residue–residue pair preferences, surrounding hydrophobicity computations, conformational switching and energy computations. We conclude that the local secondary structural preference of the two designed proteins at the N and C-terminal ends to adopt either coil or strand conformation may be a crucial factor in adopting the different folds. Early on during the process of folding, both proteins may choose different energetically favourable pathways to attain the different folds.

Keywords: conformational switch/nucleation site/pentapeptide search/protein folding/surrounding hydrophobicity

Introduction

It is of great importance to discover how proteins fold to their native three-dimensional (3D) structures to perform their biological function. The amino acid residues in a protein contribute to different extents in coding for a particular fold. Since the conformational freedom of a protein is vast, it is difficult to understand the whole process of protein folding. Some natural proteins alter their fold in response to changes in solution conditions or as a consequence of mutation (Dalal et al., 1997). The combination of these factors results in proteins sharing high sequence identity adopting different folds.

The ability of polylysine to adopt different secondary structures depends on the solution environment and method of sample preparation (Rosenheck and Doty, 1961). Conformational switching peptides were synthesized, which undergo a spontaneous, reversible transition from an α-helix to a β-sheet structure (Mutter and Hersperger, 1990). Recently, attention has focused on the possibility that conformationally sensitive sequences may play a role in diseases such as Alzheimer’s dementia and prion diseases (Nguyen et al., 1995).

A database search showed that the balance of local and long-range interactions could alter a protein’s conformation (Kabsch and Sander, 1984). Also they observed that identical pentapeptide sequences could be found in either an α-helical or in a β-sheet conformation. Although the turn and loop connections between elements of secondary structure often do not play a principal role in protein’s stability, the packing of residues in the hydrophobic core appears to be extremely important (Munson et al., 1996).

Each amino acid has a different propensity to adopt an α-helical or a β-sheet conformation. Such propensities were originally identified through statistical surveys of proteins of known structure and more recently have been quantified via experimentally determined energy scales. An 11-amino acid sequence could adopt either an α-helical or a β-sheet conformation, depending on its placement within a small protein (Minor and Kim, 1996).

The structure adopted by designed peptides and proteins depends not only on the primary sequence, but also on the conditions such as the solvent polarity or the method of sample preparation (Cerpa et al., 1996). Regan and coworkers (Dalal et al., 1997) altered the natural fold in the pair of small protein targets around 50 amino acids of length with response to the changes in solution conditions or as a consequence of mutation. A dramatic change in the secondary structure of proteins was detected by circular dichroism and nuclear magnetic resonance techniques (Dalal et al., 1997).

He et al. designed and synthesized two proteins, G$_{\Lambda}$88 and G$_{\beta}$88, with 88% sequence identity and yet adopting different folds and functions. They pointed out that the non-identities between these two amino acid sequences would be responsible for coding one fold topology over the other and also they illustrated dramatically that not all amino acids play an equal role in specifying a fold (He et al., 2008). In a very striking recent paper, they proved that a single mutation at the right position can completely change the fold of a protein (Alexander et al., 2009).

The formation of stable secondary structures and a unique tertiary structure of proteins is dictated by intra-molecular interactions between the constituent amino acid residues
along the polypeptide chain as well as their interaction with the surrounding medium (Gromiha and Selvaraj, 1997). For maintaining the native structure of a protein, an indispensable role is played by inter-fragment interactions, most of which are long range (Epand and Scheraga, 1968). Long-range interactions are interactions formed by the residues that are far in sequence and closer in space in the 3D structure of the proteins (Gromiha and Selvaraj, 1999).

Knowledge of how frequently different types of residues are found near each other in protein structures has been widely used in simulating protein folding (Vijayakumar and Zhou, 2000). Recent experimental and theoretical studies on protein folding have shown that a specific nucleus of residues is formed during the transition state of folding and the subsequent formation of the structure is very fast. Further it has been shown that the location of the specific nucleus is highly dependent on the fold and the residues that form the folding nucleus usually have hydrophobic environment (Poupan and Mormon, 1999). The hydrophobic clusters in the proteins are mostly found in the middle of the strands and they are predicted as folding nucleus (Kannan et al., 2001). The residues in the hydrophobic clusters are mainly influenced by long-range interactions and these residues have high thermal stability (Selvaraj and Gromiha, 2003).

Our studies focus on the two proteins G\textalpha88 and G\textbeta88 of 56 residue length (He et al., 2008), which differ only at seven positions that have 88% sequence identity and adopt two different folds. The 3D structures of two proteins with such high sequence identity but different fold topology are exceptional. In the context of such considerations, the aim of the present work is to investigate the sequence and structure of these two proteins by a variety of computational studies to bring out significant features that could have influenced the structures adopted by the two sequences. We found some rare combinations of pentapeptide residue patterns, which are absent, while searching sequence databases for the occurrence of overlapping stretches of pentapeptides existing in the two proteins. Since the presence of nucleation sites at various positions is so crucial in determining the fold, the difference between hydrophobic residue packing between two proteins is investigated. We have tested various secondary structure prediction methods against these small proteins that are delicately balanced between \(\alpha\) and \(\alpha/\beta\) structures encoded with only 12% of their amino acids.

**Materials and methods**

**Data set**

Atomic coordinates of the two designed proteins G\textalpha88 and G\textbeta88 (PDB ID: 2JWS and 2JWU) with 88% sequence identity but possessing different folds and functions (He et al., 2008) and atomic coordinates of the parent proteins PSD-1 and GB1 (PDB ID: 2FS1 and 1PGB) from which the above two proteins are derived were obtained from the Protein Data Bank (Henrick, K., Feng, Z., Bluhm, W.F., et al., 2008) forms the source for our present study. G\textalpha88 has a three-\(\alpha\)-helical structure, whereas G\textbeta88 has an \(\alpha/\beta\) fold and each is shown to have a high similarity to the structure of the parent proteins namely PSD-1 (mainly \(\alpha\)) and GB1 (\(\alpha/\beta\) fold) from which it was derived. The lengths of the two designed proteins are 56. We have extracted C-\(\alpha\) coordinates from the PDB files 2JWS and 2JWU to perform various computations.

**Secondary structure prediction**

In order to assess whether the difference in secondary structures between two proteins could be predicted by the different secondary structure prediction methods, we have used the NPS web server (Combet et al., 2000).

**Consensus pentapeptide secondary structure**

The sequence and secondary structure assignments for the proteins were downloaded from RCSB web server (Berman et al., 2007). The secondary structure of protein residues corresponds to the dictionary of protein secondary structure (DSSP) method (Kabsch and Sander, 1983). We grouped the eight possible secondary structural states into three types: \(\alpha\)-helix (H), \(\beta\)-strands (E) and coil (C). We included B in coils because, as an isolated bridge, it does not represent a regular geometric structure (Selbig, 2000). Overlapping pentapeptide segments from each of the two designed proteins were searched against the PDB sequences. Secondary structures of identically matched segments were obtained and a consensus secondary structure was predicted for each residue in the two proteins.

**Computation of inter-residue interactions**

Each residue in a protein molecule is represented by its \(\alpha\)-carbon atom. The centre is fixed at the \(\alpha\)-carbon atom of the first (N-terminal) residue, and the distances between this atom and the rest of the \(\alpha\)-carbon atoms in the protein molecule are calculated. The composition of the residues associated with this residue is calculated for a sphere of 8 Å radius, which has been shown to be the required volume of the medium within which a residue in a protein molecule is known to exert a detectable influence (Manavalan and Ponnuswamy, 1978). The procedure was repeated each time by moving the centre to the successive carbon atom along the polypeptide chain to compute the composition of surrounding residues, for all residues in a given protein. From the computation of surrounding residues within the sphere of 8 Å radius, the contribution due to short-range (\(\alpha\) atom \(\pm 2\) residues along the sequence), medium-range (\(\alpha\) atom \(\pm 3\) or \(\pm 4\) residues along the sequence) and long-range (\(\alpha\) atom \(> \pm 4\) residues) interactions were computed (Gromiha and Selvaraj, 2004).

**Computation of pairwise residue preferences**

For computing pairwise residue preferences, the long-range contacts computed as above were considered. The residue-residue contact preferences of seven residues that are different in the two proteins were computed in the form of a two-dimensional matrix.

**Computation of surrounding hydrophobicity**

Representing each residue in the protein molecule by its \(\alpha\) atom and fixing the first residue as the centre, the distances between this atom and the rest of the \(\alpha\) atoms in the protein molecules are computed. The residues appearing around the reference residue within a sphere of 8 Å radius are found out. These residues are assigned with their respective hydrophobic indices (Tanford, 1962) and the sum of
the hydrophobic indices of those residues that are present within the above sphere is taken to be the ‘surrounding hydrophobicity’ of that particular residue. Thus, the surrounding hydrophobicity of the $j$th residue of a given protein molecule in its folded form is given by,

$$H_f^j = \sum_{n_{ij}} h_i$$

where $n_{ij}$ is the sum of the surrounding residues of type $i$ around the $j$th residue of the protein, $h_i$ is the hydrophobic index of the $i$th residue and $f$ refers to the folded state of the protein molecule (Manavalan and Ponnuswamy, 1978). The residues having surrounding hydrophobicity values equal to or greater than twice the average value for all the residues in the protein were assumed to form ‘hydrophobic domains’ and the residue of the highest surrounding hydrophobicity within a domain is taken to represent a nucleation site of the protein (Ponnuswamy and Prabhakaran, 1980).

The surrounding hydrophobicity values for each residue in the two designed proteins and their parent proteins were computed and cross-correlated with each other. In order to study the hydrophobic core residue packing between designed and parent proteins, the residues having high surrounding hydrophobicity values were considered and their contacts between them were plotted as a two-dimensional graph that is similar in principle to the work of Soundararajan et al. (2010).

**Computation of phi and psi values and analysis of conformational switching**

Analysis was made on the phi and psi values in two designed protein structures based on the computation performed using DSSP continuous secondary structure assignment server (Carter et al., 2003). The difference between phi and psi values for each residue in both proteins were obtained and plotted as a function of residue position.

**Energy computation before and after mutation**

Both PDB structures (2JWS and 2JWU) were subjected to energy minimization by using Swiss PDB Viewer (Guex and Peitsch, 1996). The helical protein 2JWS was mutated at each non-identical position with the corresponding residues from the $\alpha/\beta$ protein and vice versa. The total energy and the energy of two designed proteins at each non-identical position were computed before and after performing mutation by using Gromacs force field in Swiss PDB viewer.

**Results**

**Secondary structure prediction**

The methods such as double prediction method (DPM), hierarchical neural network (HNNC), multivariate linear regression combination (MLRC), predicting 1D protein structure using profile based neural networks and self-optimized prediction method with alignment (SOPM) predict residues 3–6 as an extended structure in the $\alpha$-fold protein. The methods DPM, DSC, HNNC and SOPM predict residues at 1–8 in the $\alpha/\beta$ fold protein as an extended structure, which are shown in Fig. 1.

Interestingly while the residues at positions 1–8 are identical in both proteins, the method MLRC predicts them as an extended structure. This suggests that residues 1–8 possess extended structure-forming propensity. Also residues from 39 to 43 are predicted as extended structure except by GOR 1 and GOR 3 methods, which may be due to the occurrence of rare amino acid patterns at these positions. Most of the methods predict residues 42–46 and 51–55 as extended
structure in both proteins. Hence, the residues at this position are so crucial in the transition of one secondary structure over another. Alexander et al. (2009) have shown experimentally that a single mutation [L45Y] can switch the conformation from one fold to the other.

Consensus pentapeptide secondary structure prediction

The result of this analysis (Fig. 2) indicates that the C-terminal 15 residues of the proteins may play a vital role in the transition of helical regions into extended structures. The consensus pentapeptide secondary structure prediction results reveal that residues 1–5 possess sheet-forming propensity and residues 42–45 show transition from helix to extended structure. Also we found that the residues 25–35 in 2JWU are more helix-dominant region compared with that of 2JWS. Hence, there is a long central helix in that region flanked by extended structures on both sides. The residues 41–50 in 2JWS are more dominant for helix than 2JWU. Similar kind of results was observed by the DSSP continuous assignment server.

Also we found that there are some residue patterns that are not present in the database. So we further performed several other similarity search methods to identify whether these combinations of residues are occurring in any other protein sequence databases. Our search revealed that there are no such combinations of residues in other sequence databases available publicly on internet.

In the helical protein, the pattern AEKYI occurs in helix 2 and VEGVW occurs at helix 3, whereas in the α/β fold protein, the patterns WTYK and WTYKD occur at strand 2, which are shown in Fig. 2. Hence the presence of such rare combination of amino acid residues at the secondary structural element positions may lead some methods not to predict the secondary structural states accurately. The pattern WTYKD (43–47th position) contains an unusual combination of residues that may have the potential to adopt different conformations depending on the sequence context and interactions with other segments in the protein.

Inter-residue contact analysis of two designed proteins

The number of contacts and number of long-range contacts with their secondary structural states for non-identical residues between the two derived proteins are shown in Table I. The number of long-range contacts of non-identical residues in two proteins shows high variation at residues 33 and 50. In the helical protein, residue 33Ile forms nine long-range contacts, in which two of the residues form long-range contact with helix 1 and next two residues form long-range contact with helix 2 and rest of the five residues contact with helix 3. The amino acid isoleucine forms long-range contact with strong helix-forming residues, which are shown in Fig. 3. In the α/β fold protein, the 33Ile is replaced with tyrosine, which forms only one long-range contact with the N-terminal region of the protein.

The residues forming long-range contacts with 33Ile in the helical protein are strong helix formers. Also this residue strongly interacts with the third helix with five long-range contacts. The replacement of this residue with tyrosine forms one long-range contact, which does not interact with other secondary structural elements. Also the 50th residue of both proteins shows variation in the number of long-range contacts as well as residue separation between contacting residues.

Pairwise residue—residue preferences

The non-identical residues in 2JWS and 2JWU possess a completely different residue—residue pair preference as shown in Table II. In the helical protein, most of the contacts were found with the residues Ala, Ile, Leu and Lys, whereas in the case of α/β fold protein, major contacts are with the residues Ala, Glu, Ile, Leu, Lys, Thr, Tyr and Val.

The pair preference of Gly with other residues in the two proteins does not show much variation. This is because of the secondary structural propensity of the residue in the two proteins. In the case of Ile, the maximum numbers of contacts were found in the α/β fold protein than in the helical protein. Thus, a vast difference in the pair-preference interaction between residues was observed in two designed proteins.

Surrounding hydrophobicity models and nucleation sites

Analysing the surrounding hydrophobicity of the two designed proteins with 88% identity adopting different folds provides an interesting clue that there is considerable difference in hydrophobic residue packing between two proteins at the first five residues (1–5) and the last five residues (52–56), which are shown in Fig. 4. For the smaller proteins, the
folded units possess relatively small hydrophobic cores, and their folding is likely to be influenced by the environment more than that of the normal-sized proteins with a solid hydrophobic core.

The residues at positions 16, 20, 33 and 46 are having peaks, which indicate that there may be a nucleation site that starts or terminates α-helix or β-sheet, whereas in the case of α/β fold protein only two peaks are found at residues 5 and 52. Because these proteins are small, the fraction of residues that change their structure may be large. As pointed out by Alexander et al., the C-terminal five residues are unstructured in the smaller GA conformation, but form important hydrophobic interactions in the core of the GB conformation (Alexander et al., 2009).

**Table I.** Inter-residue contact analysis of non-identical residues in two designed proteins

<table>
<thead>
<tr>
<th>Residue</th>
<th>Total number of contacts</th>
<th>Number of long-range contacts</th>
<th>Number of secondary structure elements contacts</th>
</tr>
</thead>
<tbody>
<tr>
<td>24 GLY</td>
<td>24 ALA</td>
<td>7 6</td>
<td>0 0</td>
</tr>
<tr>
<td>25 ILE</td>
<td>25 THR</td>
<td>7 9</td>
<td>2 1</td>
</tr>
<tr>
<td>30 ILE</td>
<td>30 PHE</td>
<td>12 11</td>
<td>4 3</td>
</tr>
<tr>
<td>33 ILE</td>
<td>33 TYR</td>
<td>16 9</td>
<td>9 1</td>
</tr>
<tr>
<td>45 LEU</td>
<td>45 TYR</td>
<td>11 8</td>
<td>3 4</td>
</tr>
<tr>
<td>49 LEU</td>
<td>49 THR</td>
<td>10 7</td>
<td>2 2</td>
</tr>
<tr>
<td>50 LEU</td>
<td>50 LYS</td>
<td>11 12</td>
<td>3 6</td>
</tr>
</tbody>
</table>

Residue: residue along with its position in sequence is given for two designed proteins 2JWS and 2JWU;

<table>
<thead>
<tr>
<th>Residue</th>
<th>Total number of contacts</th>
<th>Number of long-range contacts</th>
<th>Number of secondary structure elements contacts</th>
</tr>
</thead>
<tbody>
<tr>
<td>50 ILE</td>
<td>50 THR</td>
<td>11 12</td>
<td>3 6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Residue</th>
<th>Total number of contacts</th>
<th>Number of long-range contacts</th>
<th>Number of secondary structure elements contacts</th>
</tr>
</thead>
<tbody>
<tr>
<td>2JWS</td>
<td>2JWU</td>
<td>2JWS 2JWU 2JWS 2JWU 2JWS 2JWU 2JWS 2JWU</td>
<td>2JWS 2JWU</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PDB ID</th>
<th>Res</th>
<th>ALA</th>
<th>ARG</th>
<th>ASN</th>
<th>ASP</th>
<th>CYS</th>
<th>GLN</th>
<th>GLU</th>
<th>GLY</th>
<th>HIS</th>
<th>ILE</th>
<th>LEU</th>
<th>LYS</th>
<th>MET</th>
<th>PHE</th>
<th>PRO</th>
<th>SER</th>
<th>THR</th>
<th>TRP</th>
<th>TYR</th>
<th>VAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>2JWS</td>
<td>GLY</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2JWS</td>
<td>ILE</td>
<td>11</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>6</td>
<td>2</td>
<td>0</td>
<td>6</td>
<td>11</td>
<td>10</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>2JWS</td>
<td>LEU</td>
<td>8</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>7</td>
<td>2</td>
<td>0</td>
<td>11</td>
<td>8</td>
<td>9</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>2</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>2JWU</td>
<td>ALA</td>
<td>14</td>
<td>0</td>
<td>6</td>
<td>5</td>
<td>0</td>
<td>3</td>
<td>15</td>
<td>3</td>
<td>0</td>
<td>13</td>
<td>15</td>
<td>24</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>7</td>
<td>2</td>
<td>5</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>2JWU</td>
<td>LYS</td>
<td>24</td>
<td>0</td>
<td>7</td>
<td>4</td>
<td>0</td>
<td>4</td>
<td>20</td>
<td>1</td>
<td>0</td>
<td>13</td>
<td>17</td>
<td>14</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>16</td>
<td>2</td>
<td>13</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>2JWU</td>
<td>PHE</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>5</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>2JWU</td>
<td>THR</td>
<td>7</td>
<td>0</td>
<td>2</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>14</td>
<td>3</td>
<td>0</td>
<td>5</td>
<td>16</td>
<td>16</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>18</td>
<td>4</td>
<td>11</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>2JWU</td>
<td>TYR</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>8</td>
<td>13</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>11</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Residue</th>
<th>Total number of contacts</th>
<th>Number of long-range contacts</th>
<th>Number of secondary structure elements contacts</th>
</tr>
</thead>
<tbody>
<tr>
<td>49 LEU</td>
<td>49 THR</td>
<td>10 7</td>
<td>2 2</td>
</tr>
</tbody>
</table>

Fig. 3 Variation of long-range contacts at 33rd position of two designed proteins.
The cross-correlation coefficient of surrounding hydrophobicity values between parent proteins and designed proteins is shown in Table III. The surrounding hydrophobicity of residues in both parent and derived protein shows sensible correlation. For helical parent and designed proteins, the correlation coefficient is 0.72 and for α/β fold parent and designed proteins, the correlation coefficient is 0.52. This may be due to the conservation of contacts between hydrophobic residues of parent and designed proteins, which are shown in Fig. 5. Hence, the designed protein adopts the parent fold due to the effect of hydrophobicity, which is believed to play a major role in organizing the self-assembly of residues during the process of protein folding even when the sequence identity is high.

### Conformational switching

A protein with 56 residues, with only 7 non-identical residues adopts structure in totally different conformation. In these two proteins, psi values changes largely compared with phi values in most of the residues. Differences in the phi and psi values on these residues were found to be high (Fig. 6). Phi and psi value difference was very low for other non-identical residues. Differences in the phi and psi values in most of the residues. Differences in the phi and psi values on these residues were found to be high (Fig. 6). Phi and psi value difference was very low for other non-identical residues.

The second residue Tyr has large difference in their phi value of about 128° shift to adopt a β-sheet (2JWS) from the coil conformation seen in 2JWU. More than 100° shift in the phi value was observed for the residues at the positions 12, 36, 37 and 56. For both 12th and 36th residue transition between coil and α-helix, secondary structure occurs in both proteins, whereas for residues at 37 and 56 positions higher phi value difference was observed though they have same coil conformation in these two proteins. Phi value difference was found to be higher for the residues switching between α-helix and β-sheet conformation in 2JWS and 2JWU.

### Energy before and after mutation

The non-identical residues in 2JWS and 2JWU possess completely different force field energy before and after mutation, which are shown in Table IV. It is not easy to understand the energetics of the switch between two alternative folds that are thermodynamically linked (Alexander et al., 2009). In the α/β fold protein, the total energy of the wild-type protein is greater than the helical protein, which is −673.047 and −1333.709, respectively. Even though the energetic situation of the non-identical residues in both proteins shows drastic variation, the binding of N- and C-termini of the α/β fold protein is the primary energetic driving force for the conformational switch (Alexander et al., 2009).

### Discussion

Folding of proteins inside a test tube is completely different from the folding of protein inside a cell. Protein folding inside the cell is thought to be facilitated by molecular chaperones, whereas in the test tube, protein folding depends mainly on various properties of amino acid residues. Since there has been considerable progress made over the past years in linking experimental and theoretical approaches to protein folding, we made systematic computational studies to analyse biophysical properties of amino acid residues such as secondary structural propensities, inter-residue interactions, surrounding hydrophobicity, geometry and force field energy against two designed proteins, which adopt two different folds with different functions encoded with only 12% of their amino acids.

The aspect of protein structure that makes the comparison problem inherently tractable is that protein structure is organized in a hierarchy of structural levels beginning with the basic unit of an amino acid, short stretches of these can adopt one of the two semi-regular local structures such as helices and strands. The results of pentapeptide sequence analysis reveal that there is an innate preference to adopt extended structure between residues 1–5 in both sequences. Although most of the secondary structure prediction methods predict the residues 1–8 to be in the coil conformation, the MLRC method predicts that these residues will adopt extended conformation. This highlights the plasticity of these residues so that it could adopt either coil or extended structure. Due to the occurrence of rare combination of amino acid residues at the regions 39–47, the secondary structural states of the residues are not well defined by any of the methods except GOR 1 and GOR 3. The results agree with the observations of Shortle (2009) who concluded that the N- and C-terminal stretches have the potential to be unstructured or to adopt different secondary structures. Alexander et al. (2009) have shown that a single mutation at the above pentapeptide stretch (L45Y) can switch the conformation from one fold to the other. It is not known, how far secondary structure formation is influenced by forces.
other than the sequences own intrinsic propensity. Understanding the degree to which the residue environment influences the structural unanimity of certain rarely occurring pentapeptides has implications for both protein design and the development of structure prediction methods (Zhou et al., 2000).

Evolutionarily closely related proteins appear to retain similar contact propensities in spite of sequence divergence and hence possess similar 3D structures; however, the sequences of Gp88 and Gp88 were derived from two different proteins, which are not evolutionary related. Even though the derived proteins share 88% sequence identity, they do not possess similar inter-residue contacts. Analysis of inter-residue interactions reveals the distinct preference in terms of sequence separation and the interacting residues. The replacement of ILE with TYR at 33rd position in α/β fold protein caused a drastic change in forming long-range contacts with other structural elements in the protein. Also the hydrophobic residue contacts may be formed early in folding and once these particular contacts have been formed, folding proceeds rapidly. Hence, the absence of these set of hydrophobic residues and their essential contacts, which form a compact core in the native state (specific nucleus) may be the reason for the protein to adopt extended structure rather than the helical structure.

In the helical protein, one could observe six hydrophobic rich regions, whereas in α/β fold protein there are only two hydrophobic rich regions at 5th and 52nd positions. This point to the innate plasticity of peptide sequences to adapt to their structural environment (hydrophobic environment in all α protein and the absence of hydrophobic environment in the α/β protein). Further, 50% of the proteins are occupied by the residues A, E, K, L and T. We propose from our analysis that conformational switching in these specified residues may convert one secondary structural state to other secondary structural state.

From this work, we conclude that the local secondary structural preference of two designed proteins at the N and C-terminal ends to adopt either coil or strand conformation may be a crucial factor in adopting the different folds. Early on during the process of folding, both proteins may choose different energetically favourable pathways to attain the different folds.

Fig. 5 Comparison of key hydrophobic core contacts between designed and parent proteins.

Fig. 6 Difference between phi and psi angles in two designed proteins.
Table IV. Force field energy before and after mutation

<table>
<thead>
<tr>
<th>Wild</th>
<th>New</th>
<th>Minimized energy</th>
<th>Total energy</th>
<th>Energy after mutation</th>
<th>Total energy</th>
</tr>
</thead>
<tbody>
<tr>
<td>2JWS (before mutation)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24 GLY</td>
<td>ALA</td>
<td>37.979</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25 ILE</td>
<td>THR</td>
<td>17.217</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30 ILE</td>
<td>PHE</td>
<td>7.535</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33 ILE</td>
<td>TYR</td>
<td></td>
<td>-673.047</td>
<td>20.187</td>
<td>421.447</td>
</tr>
<tr>
<td>5 LEU</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>49 ILE</td>
<td>THR</td>
<td>15.223</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50 LYS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2JWU (after mutation)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24 GLY</td>
<td></td>
<td>-6.417</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25 THR</td>
<td>ILE</td>
<td>-10.83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30 PHE</td>
<td>ILE</td>
<td>-35.899</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33 TYR</td>
<td>ILE</td>
<td>-93.216</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45 TYR</td>
<td>LEU</td>
<td>-93.364</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>49 THR</td>
<td>ILE</td>
<td>16.705</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50 LYS</td>
<td>LEU</td>
<td>-6.157</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Wild, wild-type residue along with its position in sequence; new, residues replaced during mutation; minimized energy, energy of each residue after energy minimization before mutation; total energy, energy of the protein before mutation.

**Funding**

This work was supported by the Department of Science and Technology, Government of India SR/SO/BB-09/2006 (K.M.S. and B.H.) and University Grants Commission, Government of India through Rajiv Gandhi National Fellowship (N.S.).

**References**


