ELSEVIER

Biophysical Chemistry 99 (2002) 189–198

Biophysical Chemistry

www.elsevier.com/locate/bpc

# Influence of medium- and long-range interactions in different folding types of globular proteins

T.S. Kumarevel[a,*], M. Michael Gromiha[b], S. Selvaraj[c], K. Gayatri[a], P.K.R. Kumar[a]

[a]*National Institute of Advanced Industrial Science and Technology (AIST), Institute of Molecular and Cell Biology, Functional Nucleic Acids Group, Tsukuba Central 6, 1-1 Higashi, Tsukuba Science City, Ibaraki 305-0045, Japan*
[b]*Computational Biology Research Center (CBRC), AIST, 2-41-6 Aomi, Koto-Ku, Tokyo 135-0064, Japan*
[c]*Department of Physics, Bharathidasan University, Tiruchirapalli 620 024, Tamil Nadu, India*

Received 21 February 2002; received in revised form 22 May 2002; accepted 7 June 2002

**Abstract**

Recognition of protein fold from amino acid sequence is a challenging task. The structure and stability of proteins from different fold are mainly dictated by inter-residue interactions. In our earlier work, we have successfully used the medium- and long-range contacts for predicting the protein folding rates, discriminating globular and membrane proteins and for distinguishing protein structural classes. In this work, we analyze the role of inter-residue interactions in commonly occurring folds of globular proteins in order to understand their folding mechanisms. In the medium-range contacts, the globin fold and four-helical bundle proteins have more contacts than that of DNA–RNA fold although they all belong to all-α class. In long-range contacts, only the ribonuclease fold prefers 4–10 range and the other folding types prefer the range 21–30 in α/β class proteins. Further, the preferred residues and residue pairs influenced by these different folds are discussed. The information about the preference of medium- and long-range contacts exhibited by the 20 amino acid residues can be effectively used to predict the folding type of each protein.


*Keywords:* Structural class; Folds; Medium-range contacts; Long-range contacts

## 1. Introduction

A unique stable three-dimensional structure of a protein can be attained by inter and intra molecular interactions between the constituent of amino acid residues along the linear polypeptide chain. The fold of the protein is defined by the arrangement of its major elements of secondary structures and by the topology of the connections between them. In most proteins the α-helices and β-sheets pack together in one of the different possible ways [1]. The connections between secondary structures obey a set of empirical topological rules in almost all cases. This means that it is common for sets of unrelated proteins to have similar, if not identical, folds. Based on the crystal structures available in the Protein Data Bank (PDB), a number of investigations have been carried out to understand the

*Corresponding author. Tel.: +81-298-61-6085; fax: +81-298-61-6095.
*E-mail address:* kumarevel-thirumananseri@aist.go.jp (T.S. Kumarevel).

0301-4622/02/$ - see front matter 
PII: S0301-4622(02)00183-7

role of different kind of interactions in the folding and stability of globular proteins [2–8]. The importance of long-range interactions has been stressed on various aspects such as the folding and stability of proteins [9,10], deriving potentials for fold recognitions [11–14] of globular proteins establishing a relationship between protein sequence and structure [15] and predicting protein folding rates [16].

Recently, we have analyzed the influence of medium- and long-range interactions in different structural classes of proteins and showed that while medium-range interactions predominate in all α-class proteins, long-range interactions predominate in the all β-class [17]. Based on this fact, an evaluation of the performance of several secondary structure prediction methods revealed that all the methods predict the secondary structure of all proteins more accurately than other classes [18]. Furthermore, the concept of inter-residue contacts has been successfully applied for discriminating globular and membrane proteins [19] and to predict the folding rates of two-state protein [16].

In this work, we have analyzed the influence of medium- and long-range interactions in several commonly occurring folds of globular proteins. The results show the presence of distinct inter-residue contacts in different folds among the same structural class, which can be utilized for fold recognition.

## 2. Materials and methods

### *2.1. Database*

It is now well established that protein domains having more than 30% of their sequence in common adopt the same fold structures [20–23]. In the present study, the classification method of Murzin et al. [24] was used for the distinction between different structural classes (SCOP, version 1.38). The structural classification in SCOP is entirely manual and does not incorporate any of the 'hard and fast' rules, such as those described by different authors [25–27]. Rather it focuses on what structural elements are within the 'core' of the protein. A protein with seven strands and one helix may be 'α+β' if the helix is integral to the core, while it would be 'all β ' if the helix was a non-conserved elaboration. The distinction of α/β and α+β is made on the interactions between the α and β sections of the structure. Therefore, in comparison with other classifications that only based on the percentages of secondary structures, the classification in SCOP is more natural, better reflects the objective reality, and provides a more reliable database for the study of proteins in different structural classes. At a sequence identity level of 30%, 675 (155 all-α, 156 all-β, 184 α+β and 180 α/β) unique sequences corresponding to the four known structural classes have been found in the PDB40D_1.37 database of SCOP [28]. These 675 protein coordinates taken for the PDB of Brookhaven National Laboratory [29] form the source for our present study. Further, we obtained the information about the fold of all proteins from SCOP [30] and CATH [31] databases.

The four letter PDB codes (fifth and sixth letters indicate the chain identification and domain number, respectively) for the proteins of different folds used in the present study are given below.

*Four-helical bundle:* 1BUC, 1FAP, 1LE4, 1LPE, 1VTM, 256B, 2CCY, 2HMQ, 2LIG, 2MHR.

*Globin:* 1BAB, 1CPC, 1ECO, 1FCS, 1HBG, 1LH1, 1MBS, 2HCOA, 2HCOB, 2LHB, 3SDH, 4MBN.

*DNA–RNA:* 1AOY, 1BIA, 1CGP, 1DPRA1, 1ETD, 1FOW, 1HCR, 1HST, 1IGN, 1LEA, 1MSE, 1OPC, 1PDN, 1RNL, 1SFE, 1TNS, 1XGS, 2HTS.

*Immunoglobulin:* 1A45, 1ACX, 1BGLA1, 1CD1, 1CD8, 1CID, 1ENH, 1GGT, 1GOF_1, 1HIL, 1MAM, 1MSP, 1NCI, 1NEU, 1NOA, 1REI, 1TEN, 1TLK, 1VCAA1, 1VCAA2, 1XSO, 2HFT, 2SOD, 3DPA_1, 3HHR, 4FAB, 4KBP.

*Oligosaccharide binding (OB):* 1ASY, 1CKM, 1CUK, 1ESF, 1LTS, 1MJC, 1PYP, 1RIP, 1STY, 1YHB, 3ULL.

*Ferredoxin:* 1AB8, 1AFI, 1APS, 1DAR_4, 1DUR, 1FCA, 1FD2, 1FNB, 1FWP, 1MLA, 1MLI, 1NPK, 1NRC, 1PBA, 1PIL, 1PSD, 1RAA, 1REG, 1RIS, 1UP1, 1VAOA1, 1VJW, 1XER, 2BOP, 2PIA, 3RUB, 3RUBL2.

*Flavodoxin:* 1BMT, 1CUS, 1ESC, 1GPMA2, 1ORDA1, 1QRD, 1REQA2, 1SCUA1, 2FX2, 2NACA1, 3CHY.

*Ribonuclease:* 1ASU, 1BCO, 1CHM, 1GLA, 1HJR, 1HPM, 1KFD, 1NOY, 1SFE, 2RN2.

*FAD–NAD:* 1COY_1, 1COY_2, 1DNP, 1GAL, 1GD1, 1GES, 1GND_1, 1GND_2, 1HRDA1, 1KIF, 1MBB_1, 1PBE, 1PVDA1, 1SCUA2, 1VAOA2, 1XEL, 1YVE, 2CMD, 2NACA2, 2OHX, 2TMDA2.

*Rossmann:* 1CSE, 1DHR, 1ETU, 2DRI, 2HAD, 2LIV, 2PGD, 2SBT, 3ADK, 3COX, 3LDH, 3PGK, 4GPD, 4ICD, 5ABP, 5ADH, 5NLL.

$(\alpha/\beta)_8$ *barrel:* 1ADD, 1AK5, 1BKS, 1BPL, 1BYB, 1CBG, 1CDG, 1DIK_1, 1DJX, 1DOR, 1DOS, 1EBH, 1FCB, 1GOX, 1GYM, 1HVQ, 1LUC, 1MNS, 1NAL, 1NFP, 1PII, 1PKYA2, 1PTA, 1PUD, 1QAP, 1QBA, 1REQA1, 1SFT, 1TIM, 1TPF, 1TRE, 1UCW, 1XYZ, 2AAA, 2ACR, 2KAU, 2KAUC1, 2MNR, 2TAA, 2XIS, 3RUB.

### *2.2. Computation of surrounding residues and long-range contacts*

The computation of surrounding residues in a protein molecule has been described in our earlier article [8]. The residues in a protein molecule are represented by their Cα atoms. Using the Cα coordinates, a sphere of radius 8 Å is fixed around each residue and the composition of surrounding residues associated with all the residues is calculated. It has been shown that the influence of each residue over the surrounding medium extends effectively only up to 8 Å [32–34].

For a given residue, the composition of surrounding residues is analyzed in terms of the location at the sequence level and the contributions from $<\pm 3$ residues are treated as short range contacts, $\pm 3$ or $\pm 4$ residues as medium-range contacts and $>\pm 4$ residues are treated as long-range contacts [17].

### *2.3. Long-range contacts in different ranges of residue distances*

The long-range contacts ($>\pm 4$ residues) are further classified into several intervals with a step of 10 (4–10; 11–20; 21–30; 31–40; 41–50 and $>50$). The number of long-range contacts in each interval for all the residues in different folding types was computed. Furthermore, the percentage of long-range contacts for all the folds in each interval were calculated.

### *2.4. Preference of surrounding residues influenced by long-range contacts*

The residues coming within a sphere of 8 Å for each residue in all the four structural classes and different fold types were computed and the residues which contribute towards long range are selected as described above. For a given residue, the preference of all the 20 amino acid residues to form long-range contacts is computed and the total preference for all the 20 amino acid residues is estimated. The average preference of surrounding residues is computed using the expression:

$$\langle N\rangle_{i,j}=\frac{\sum N_{i,j}}{\sum N_i+\sum N_j}$$

where $N_{i,j}$ is the number of surrounding residues of type $j$ around residue $i$. $N_i$ and $N_j$ are, respectively the total number of residues of type $i$ and $j$. The top ten residue pairs were selected and used for further analysis. A similar analysis was carried out for residues influenced by medium-range contacts.

## 3. Results and discussion

### *3.1. Occurrence of residues in medium- and long-range interactions in different folding types*

The average numbers of medium- and long-range contacts for 11 commonly occurring folding types in different structural classes are given in Table 1.

Among the 11 folding types, the folds such as, four-helical bundle, globin and DNA–RNA belong to all-α class, immunoglobulin and OB belong to all-β, ferridoxin belongs to $\alpha+\beta$ and the remaining 5 folding types (flavodoxin, ribonuclease, FAD–NAD, Rossmann and $(\alpha/\beta)_8$ barrel) belong to $\alpha/\beta$ class. We have analyzed the folding types that have sufficient number ($\geqslant 10$) of proteins.

We found that the average medium- and long-range contact for DNA–RNA fold differs appreciably from that of all-α proteins. Further, the

Table 1
Average medium- and long-range contacts for different structural classes and different folding types of globular proteins

| Structural class/fold | $N_p$ | $N_{res}$ | Number of contacts/residue | |
|---|---|---|---|---|
| | | | Medium | Long |
| *All-α* | 155 | 22 335 | 2.72±0.662 | 2.05±0.853 |
| Four-helical bundle | 10 | 1309 | 3.15±0.358 | 1.88±0.377 |
| Globin | 12 | 1786 | 3.17±0.089 | 2.08±0.221 |
| DNA–RNA | 18 | 1401 | 2.26±0.347 | 2.22±0.570 |
| *All-β* | 156 | 26 004 | 0.91±1.459 | 4.81±1.969 |
| Immunoglobulin | 27 | 3481 | 0.75±0.401 | 4.72±0.841 |
| OB | 11 | 1290 | 1.04±0.389 | 4.19±0.553 |
| α+β | 184 | 27 637 | 1.64±1.173 | 3.78±1.224 |
| Ferredoxin | 27 | 2862 | 1.58±0.280 | 3.68±0.569 |
| α/β | 180 | 47 245 | 1.95±1.000 | 3.88±1.009 |
| Flavodoxin | 11 | 2036 | 1.94±0.276 | 3.85±0.514 |
| Ribonuclease | 10 | 1921 | 1.86±0.139 | 3.66±0.397 |
| FAD–NAD | 21 | 4000 | 1.74±0.414 | 3.89±0.492 |
| Rossmann | 17 | 5151 | 1.95±0.328 | 4.23±0.698 |
| $(\alpha/\beta)_8$ barrel | 41 | 13 620 | 2.03±0.312 | 4.06±0.422 |

$N_p$, number of proteins taken for the analysis; $N_{res}$, number of residues in each data set.

globin fold and four-helical bundle proteins have more medium-range contacts than that of RNA–DNA fold although all these proteins belong to the same structural class. The four-helical bundle proteins have less long-range contacts compared with other folding types of proteins in all-α class.

In all-β class proteins, OB fold proteins have remarkably low long-range contacts. Among the different folds belong to α/β class, FAD–NAD has the lowest medium-range contacts and Rossmann fold has the highest long-range contacts.

### *3.2. Long-range contacts for different residue intervals in different folding types*

The average percentage of long-range contacts for different residue intervals in different fold types of all structural classes are given in Table 2.

In all-α class, all the folding types prefer the range of 4–10 and the four-helical bundle fold has the lowest contact in the same range. Further, globin fold proteins show insignificant number of contacts in the range of 11–20. The immunoglobulin and OB fold proteins prefer the long-range contacts in the range of 11–20 as observed in all-β class proteins [10]. In α/β class folds, it is

Table 2
Percentage of long-range contacts for different residue intervals in four structural classes and different folding types

| Structural class/folding type | $N_{long}$ | Percentage long-range contacts | | | | | |
|---|---|---|---|---|---|---|---|
| | | 4–10 | 11–20 | 21–30 | 31–40 | 41–50 | >50 |
| *All-α* | 55 324 | 31.95±16.531 | 14.84±13.908 | 11.16±8.043 | 10.75±8.698 | 7.78±8.033 | 21.57±17.343 |
| Four-helical | 2482 | 20.17±4.991 | 12.58±8.683 | 12.10±5.128 | 12.23±5.356 | 11.85±4.647 | 31.06±11.760 |
| Globin | 3716 | 24.03±4.340 | 5.98±1.745 | 5.72±1.423 | 12.18±3.750 | 9.25±3.568 | 42.83±7.635 |
| DNA–RNA | 3202 | 36.77±11.218 | 19.51±7.996 | 8.13±6.859 | 15.75±10.595 | 11.65±7.792 | 8.18±11.373 |
| *All-β* | 130 028 | 18.14±7.060 | 22.59±10.317 | 13.49±10.324 | 10.21±7.847 | 8.57±5.578 | 26.97±15.908 |
| Immunoglobulin | 16 870 | 17.13±6.916 | 23.21±3.814 | 13.45±5.870 | 10.33±5.295 | 10.66±4.776 | 25.21±10.297 |
| OB | 5434 | 22.38±4.287 | 25.54±5.632 | 11.72±5.167 | 12.92±7.118 | 10.35±2.812 | 17.09±12.215 |
| α+β | 108 970 | 21.85±7.987 | 20.09±9.031 | 12.83±9.065 | 11.19±7.781 | 8.02±7.217 | 26.02±14.737 |
| Ferredoxin | 10 370 | 18.82±5.182 | 14.43±4.536 | 13.17±8.988 | 12.19±6.296 | 11.27±5.078 | 30.12±12.909 |
| α/β | 187 404 | 17.22±4.093 | 11.99±5.949 | 19.54±11.198 | 11.45±11.626 | 7.26±7.106 | 32.53±15.522 |
| Flavodoxin | 7946 | 16.22±4.265 | 11.25±8.000 | 26.26±12.526 | 11.84±9.986 | 12.34±5.967 | 22.08±10.716 |
| Ribonuclease | 7058 | 22.28±5.045 | 18.45±4.731 | 13.15±8.269 | 7.26±5.585 | 5.48±3.525 | 33.38±8.221 |
| FAD–NAD | 15 644 | 19.50±8.359 | 14.31±5.305 | 23.90±15.907 | 8.22±5.238 | 6.01±5.229 | 28.06±11.130 |
| Rossmann | 22 344 | 16.33±2.305 | 13.40±9.441 | 19.00±8.649 | 9.54±6.398 | 7.08±6.764 | 34.63±12.673 |
| $(\alpha/\beta)_8$ barrel | 55 286 | 16.17±2.447 | 12.06±5.017 | 19.53±9.767 | 15.58±7.103 | 10.17±7.388 | 26.48±9.980 |

$N_{long}$, number of long-range contacts.

interesting to note that only the ribonuclease fold prefers 4–10 range and the other fold types prefer the range 21–30 as observed in other $\alpha/\beta$ class proteins [10]. In consistent with the previous analysis [35] that the $(\alpha/\beta)_8$ barrel fold proteins have the highest long-range contacts in the range of 21–30. The analysis on the interval $>50$ showed that among the 11 folding types, the globin fold has the highest number of long-range contact whereas DNA–RNA fold has the lowest contact, both belong to all-$\alpha$ class. The subdivision of interval $>50$ into small bins yielded insignificant number of contacts and hence the results are not included in the analysis.

### *3.3. Preference of residues in the medium- and long-range contacts*

#### *3.3.1. Folds belong to all-$\alpha$ class proteins*

The average medium- and long-range contacts for all the 20 amino acid residues for four-helical bundle, globin and DNA–RNA fold types along with all-$\alpha$ proteins are presented in Fig. 1a and b. All the 20 amino acid residues in the four-helical bundle and Globin folds have higher medium-range contacts and the DNA–RNA fold has lower medium-range contacts compared to all-$\alpha$ class proteins. The residue His has the highest medium-range contact in four-helical bundle fold followed by Leu, Cys, Met, Ala and Arg. The residue, Pro has the lowest medium-range contacts due to its helix breaking tendency [36]. In globin fold, Cys has the highest medium-range contacts followed by Trp, Ile, Leu and Val. The residue Leu has the highest medium-range contacts in DNA–RNA fold followed by Gln and Ala. We noticed that Pro has higher medium-range contacts in globin fold compared to other folds. Although all the above three-folds belong to all-$\alpha$ class, the residue preferences in the medium-range changed significantly and this information may be helpful to identify the fold recognition.

In long-range contacts, the residue Cys has observed the highest long-range contacts in all-$\alpha$ class proteins followed by Val, Ile, Tyr and Phe. The negatively charged residues Glu and Asp have low long-range contacts. In four-helical bundle fold, all the 20 amino acid residues have less long-range contacts compared to all-$\alpha$ class proteins. In globin fold, the residues Cys, Lys, Met, Gln and Val have appreciable number of long-range contacts. Interestingly, the residue Tyr has the highest long-range contacts in DNA–RNA fold followed by Ile, Val, Cys and Phe. It is noteworthy that the residue Tyr, Ile, Val, Phe and Glu have more number of long-range contacts in DNA–RNA compared with all-$\alpha$ class proteins.

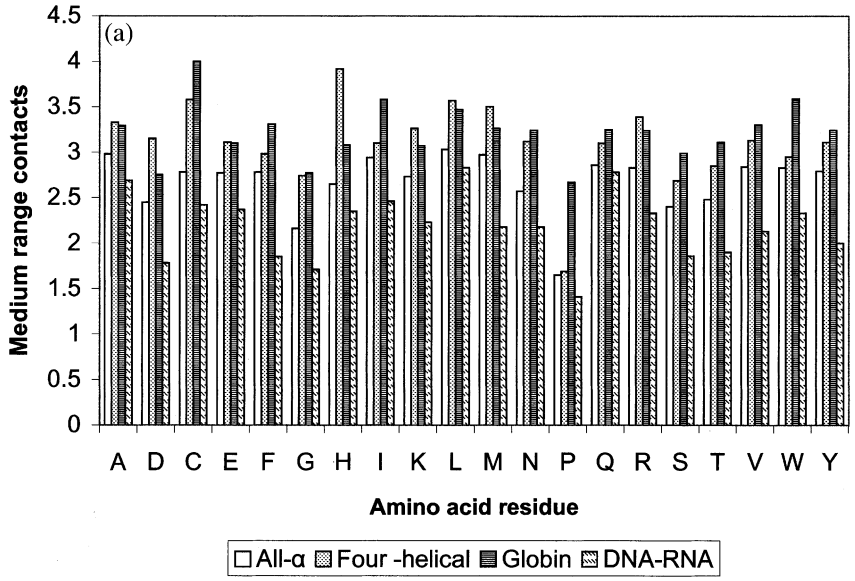
#### *3.3.2. Folds belong to $\alpha/\beta$ class proteins*

The average medium- and long-range contacts for all the 20 amino acid residues in different folds of $\alpha/\beta$ class of proteins are presented in Fig. 2a and b. The residues Glu, Lys, Met, Asn, Arg and Thr in flavodoxin, residues Lys, Asn, Ser and His in ribonuclease, residue His in FAD–NAD, residues Phe, Gly, His, Lys, Met, Pro, Trp and Tyr in Rossmann and the residues Ala, Asp, Glu, Ile, Lys, Asn, Gln, Arg, Val and Tyr in $(\alpha/\beta)_8$ barrel have more number of medium-range contacts compared to $\alpha/\beta$ class proteins.

In long-range contacts, the residues Cys, Leu, Met, Gln, Val, Trp and Tyr in flavodoxin, Glu, Phe, and Thr in ribonuclease, Ala and Ile in FAD–NAD have higher preference compared to $\alpha/\beta$ class proteins. The residues Ala, Gly, His, Ile, Asn, Arg, Ser, Val and Trp have more long-range contacts in Rossmann fold than other folds belong to $\alpha/\beta$ class proteins. The residues Cys, Gln and Tyr in flavodoxin and the residues Glu, Phe and Thr have significant number of long-range contacts in ribonuclease fold. The residue Pro has appreciable number of long-range contacts in $(\alpha/\beta)_8$ barrel fold proteins. It is interesting to note that the residue preferences vary with each folding type in the average medium- and long-range contacts although they all belong to one structural class. Hence, the information about the preference of medium- and long-range contacts exhibited by the 20 amino acid residues can be effectively used to predict the folding type of each proteins.

### *3.4. Residue pairs influenced by medium- and long-range contacts*

The preference of each amino acid residue to be surrounded by all the 20 amino acid residues



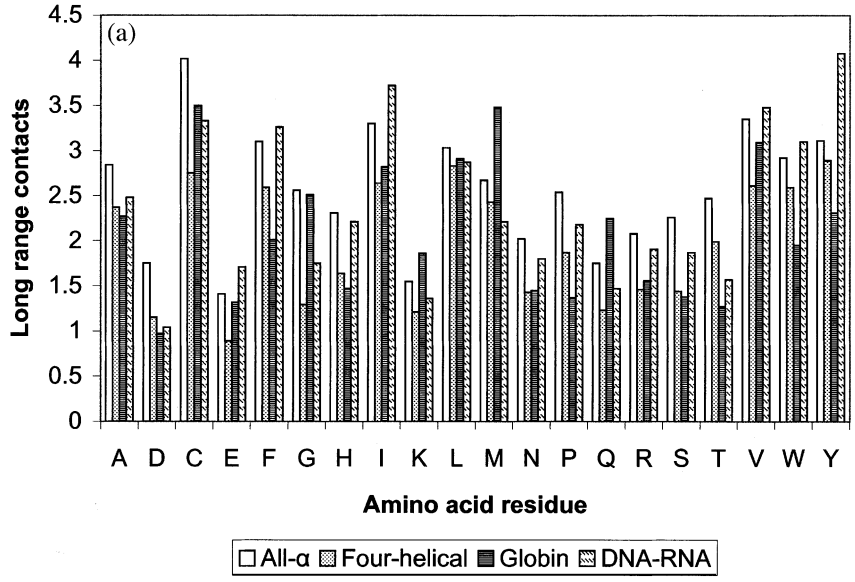




Fig. 1. (a) Average medium-range contacts in different folds belong to all-α class. (b) Average long-range contacts in different folds belong to all-α class.

due to long-range contacts in different folding types of proteins belonging to all-α class and α/β class are computed and the topmost 10 residue pairs are given in Table 3a and b. We found that the preferences of residue pairs to form medium- and long-range contacts shows a significant difference among different folds of all-α and α/β class proteins. In general, the hydrophobic and polar residues have an equal preference in forming medium-range contacts whereas in long-range con-

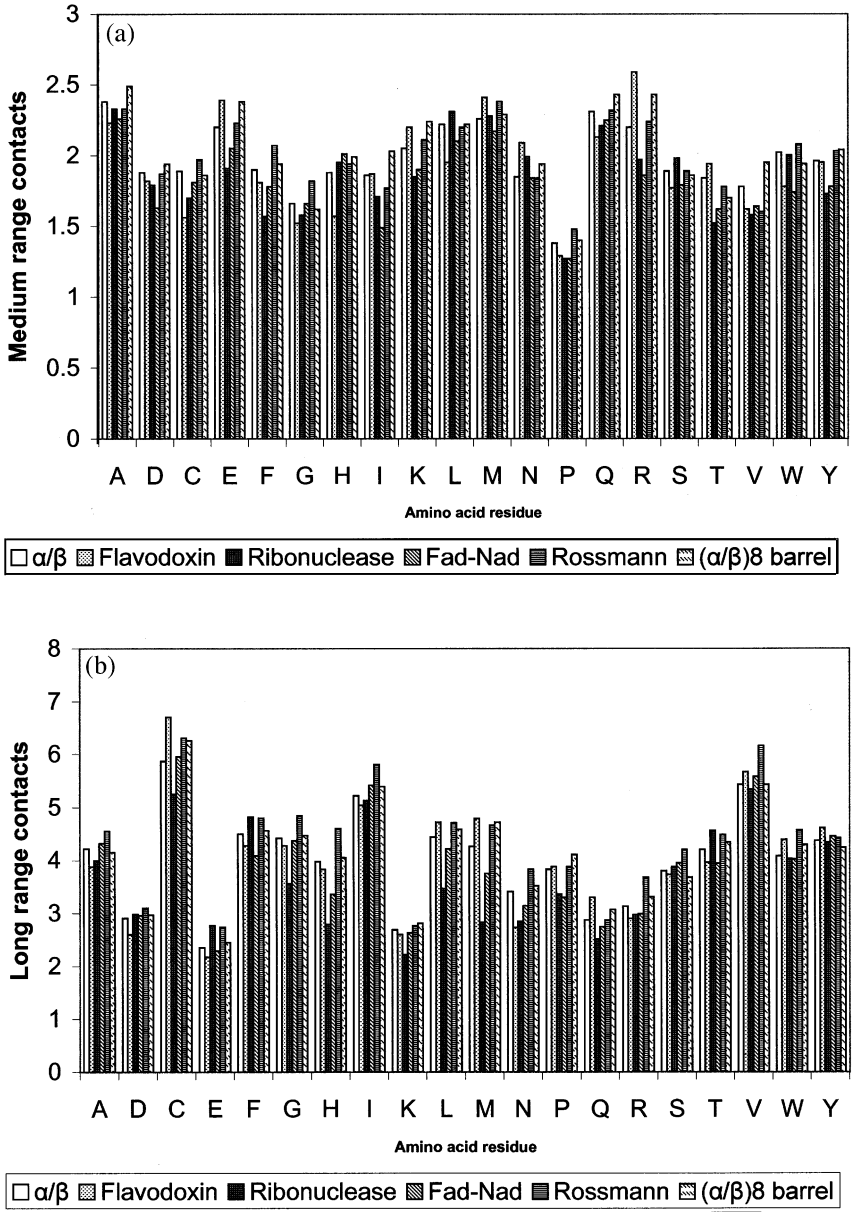



Fig. 2. (a) Average medium-range contacts in different folds belong to α/β class proteins. (b) Average long-range contacts in different folds belong to α/β class proteins.

tacts, most of the contribution is influenced by hydrophobic residues.

The analysis on the preference of amino acid residues in the medium-range contacts showed that the residues Asp and Gln in four-helical bundle, Gly, Phe, and His in globin and Ser in DNA–RNA fold make appreciable contribution. In long-range contacts, Pro, Met, Ser, His and Tyr in four-helical bundle, Gly, Lys, Glu in Globin and Asn, Trp and Ser in DNA–RNA fold make appre-

Table 3
Topmost 10 residue pairs influenced by different folding types

(a) Folds belong to all-α class

| Medium-range contacts | | | | Long-range contacts | | | |
|---|---|---|---|---|---|---|---|
| All-α | Four-helical | Globin | DNA–RNA | All-α | Four-helical | Globin | DNA–RNA |
| L–L | A–A | A–A | L–L | L–L | A–A | A–L | L–L |
| A–A | D–K | A–K | K–K | C–C | L–L | A–V | A–A |
| E–K | L–L | L–L | A–L | A–L | A–L | G–K | I–I |
| A–L | D–R | V–V | E–R | A–A | A–V | L–L | I–V |
| E–R | L–Q | G–K | I–L | I–L | V–V | L–V | L–V |
| A–E | K–K | G–L | A–K | A–V | P–P | A–A | N–W |
| L–V | A–L | F–L | L–V | L–V | M–S | V–V | A–L |
| I–L | E–E | H–L | R–V | A–I | L–R | A–G | I–L |
| E–E | A–K | L–V | S–S | I–I | H–H | E–K | S–S |
| A–V | L–R | A–G | I–I | V–V | Y–Y | A–I | I–R |

(b) Folds belong to α/β class

| Medium-range contacts | | | | | | Long-range contacts | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| α/β | Flavodoxin | Ribonuclease | FAD–NAD | Rossmann | $(\alpha/\beta)_8$ barrel | α/β | Flavodoxin | Ribonuclease | FAD–NAD | Rossmann | $(\alpha/\beta)_8$ barrel |
| A–A | A–A | A–A | A–A | A–A | A–A | V–V | V–V | V–V | V–V | V–V | V–V |
| L–L | G–L | D–L | L–L | E–K | A–L | L–L | A–V | L–V | G–G | G–G | A–V |
| A–L | I–E | E–R | G–S | A–G | L–L | G–G | I–V | A–L | I–I | A–V | A–A |
| E–K | A–K | I–V | A–G | G–G | E–K | L–V | L–L | A–V | I–V | I–V | A–L |
| A–G | E–K | A–L | V–V | A–L | A–E | A–V | I–I | I–V | L–L | I–I | L–L |
| A–R | G–L | E–K | G–G | E–L | A–G | I–V | A–I | G–G | A–V | I–L | I–V |
| A–E | E–E | L–R | A–V | V–V | D–K | A–G | A–A | L–L | L–V | L–V | L–V |
| A–V | A–L | L–L | I–I | G–L | A–K | A–L | A–L | I–L | A–I | L–L | I–L |
| G–L | D–K | D–R | A–L | D–K | A–V | A–A | I–L | A–I | A–A | A–L | A–I |
| L–V | T–T | A–E | G–V | A–K | A–R | I–I | A–G | A–A | A–G | A–I | I–I |

ciable contribution. Although these threefolds belong to all-α class, the preferred residue pairs significantly differ in each fold.

A–A and V–V are the most preferred residue pairs in all the folds belonging to α/β class proteins in the medium- and long-range contacts, respectively. The role of Gly in forming medium-range contacts is distinct in different folding types of α/β proteins; it has appreciable contribution in FAD–NAD while it is minimal in ribonuclease fold. The polar and hydrophobic residues influence the medium-range contacts equally whereas the long-range contacts are influenced mainly by hydrophobic residues in the different folds belong to α/β class proteins.

## 4. Conclusions

We have analyzed the role of medium- and long-range contacts in several commonly occurring folds of globular proteins. We found that the globin and four-helical bundle folds have appreciably higher medium-range contacts than that of DNA–RNA fold. Immunoglobulin fold in all-β proteins has the highest number of long-range contact among all the folding types. Ribonuclease fold prefers the range 4–10 while the residues in all other folds of α/β class proteins interact with distant residues in the range of 21–30 to form long-range contacts. The role of Gly is more important to form medium- and long-range contacts in globin fold than other folding types of all-α proteins. The observations found in the present study could be useful for predicting the fold recognition.

## References

[1] C. Chothia, Proteins. One thousand families for the molecular biologist, Nature 357 (1992) 543–544.
[2] P.K. Ponnuswamy, M.M. Gromiha, On the conformational stability of folded proteins, J. Theor. Biol. 166 (1994) 63–74.
[3] Bahar, M. Kaplan, R.L. Jernigan, Short-range conformational energies, secondary structure propensities, and recognition of correct sequence-structure matches, Proteins 29 (1997) 309–317.
[4] C. Zhang, J.L. Cornette, C. Delisi, Consistency in structural energetics of protein folding and peptide recognition, Protein Sci. 6 (1997) 1057–1064.
[5] T.S. Kumarevel, M.M. Gromiha, M.N. Ponnuswamy, Analysis of hydrophobic and charged patches and influence of medium and long range interactions in molecular chaperones, Biophys. Chem. 75 (1998) 105–113.
[6] R.B. Russell, M.A. Saqi, P.A. Sayle, M.J. Sternberg, Recognition of analogous and homologous protein folds—assessment of prediction success and associated alignment accuracy using empirical substitution matrices, Protein Engg. 11 (1998) 1–9.
[7] M.M. Gromiha, M. Oobatake, H. Kono, H. Uedaira, A. Sarai, Role of structural and sequence information in the prediction of protein stability changes: comparison between buried and partially buried mutations, Protein Engg. 12 (1999) 557–561.
[8] M.M. Gromiha, S. Selvaraj, Inter-residue interactions in the structure, folding and stability of proteins, Recent Res. Devel. Biophys. Chem. 1 (2000) 1–14.
[9] D.S. Gottfried, E. Hass, Nonlocal interactions stabilize compact folding intermediates in reduced unfolded bovine pancreatic trypsin inhibitor, Biochemistry 31 (1992) 12353–12362.
[10] M.M. Gromiha, S. Selvaraj, Importance of long-range interactions in protein folding, Biophys. Chem. 77 (1999) 49–68.
[11] A. Nayeem, H.A. Scheraga, A statistical analysis of side-chain conformations in proteins: comparison with ECEPP predictions, J. Protein Chem. 13 (1994) 283–296.
[12] B. Reva, A.V. Finkelstein, M. Sanner, A.J. Olson, Residue–residue mean-force potentials for protein structure recognition, Protein Engg. 10 (1997) 865–876.
[13] J. Skolinick, A. Kolinski, A.R. Ortiz, MONSSTER: a method for folding globular proteins with a small number of distance constraints, J. Mol. Biol. 265 (1997) 217–241.
[14] V.I. Abkevich, A.M. Gutin, E.I. Shakhnovich, Impact of local and non-local interactions on thermodynamics and kinetics of protein folding, J. Mol. Biol. 252 (1995) 460–471.
[15] J. Selbig, P. Argos, Relatioships between protein sequence and structure patterns based on residue contacts, Proteins 31 (1998) 172–185.
[16] M.M. Gromiha, S. Selvaraj, Comparison between long-range interactions and contact order in determining the folding rate of two-state proteins: application of long-range order to folding rate prediction, J. Mol. Biol. 310 (2001) 27–32.
[17] M.M. Gromiha, S. Selvaraj, Influence of medium and long-range interactions in different structural classes of globular proteins, J. Biol. Phys. 23 (1997) 151–162.
[18] M.M. Gromiha, S. Selvaraj, Protein secondary structure prediction in different structural classes, Protein Engg. 11 (1998) 249–251.
[19] M.M. Gromiha, S. Selvaraj, Role of medium and long-range interactions in discriminating globular and membrane proteins, Int. J. Biol. Macromol. 29 (2001) 25–34.

[20] C. Sander, R. Schneider, Database of homology-derived protein structures and structural meaning of sequence alignment, Proteins 9 (1991) 56–58.
[21] T.L. Blundell, M.S. Johnson, Catching a common fold, Protein Sci. 2 (1993) 877–883.
[22] T.P. Flores, C.A. Orengo, D. Moss, J.M. Thornton, Comparison of conformational characteristics in structurally similar protein pairs, Protein Sci. 2 (1993) 1811–1826.
[23] M. Hilbert, G. Bohm, R. Jaenicke, Structural relationships of homologous proteins as a fundamental principle in homology modeling proteins, Proteins 17 (1993) 138–151.
[24] A.G. Murzin, S.E. Brenner, T. Hubbard, C. Chothia, SCOP: a structural classification of protein database for the investigation of sequence and structures, J. Mol. Biol. 247 (1995) 536–540.
[25] H. Nakashima, K. Nishikawa, T. Ooi, The folding type of a protein is relevant to the amino acid composition, J. Biochem. 99 (1986) 153–162.
[26] P. Klein, C. Delisi, Prediction of protein structural class from amino acid sequence, Biopolymers 25 (1986) 1659–1672.
[27] K.C. Chou, A novel approach to predicting protein structural classes in a (20-1)-D amino acid composition space, Proteins 21 (1995) 319–344.
[28] Z.X. Wang, Z. Yuan, How good is prediction of protein structural class by the component-coupled method?, Proteins 38 (2000) 165–175.
[29] H.M. Berman, J. Westbrook, Z. Feng, et al., Protein data bank, Nucleic Acids Res. 28 (2000) 235–242.
[30] T.J.B. Hubbard, B. Ailey, S.E. Brenner, A.G. Murzin, C. Chothia, SCOP: a structural classification of proteins database, Nucleic Acids Res. 27 (1999) 254–256.
[31] C.A. Orengo, F.M.G. Pearl, J.E. Bray, et al., The CATH database provides insights into protein structure/function relationships, Nucleic Acids Res. 27 (1999) 275–279.
[32] P. Manavalan, P.K. Ponnuswamy, A study of the preferred environment of amino acid residues in globular proteins, Arch. Biochem. Biophys. 184 (1977) 476–487.
[33] P. Manavalan, P.K. Ponnuswamy, Hydrophobic character of amino acid residues in globular proteins, Nature 275 (1978) 673–674.
[34] P.K. Ponnuswamy, Hydrophobic characteristics of folded proteins, Prog. Biophys. Mol. Biol. 59 (1993) 57–103.
[35] S. Selvaraj, M.M. Gromiha, Importance of long-range interactions in $(\alpha/\beta)_8$ barrel fold, J. Prot. Chem. 17 (1998) 691–697.
[36] G.D. Fasman, The development of the prediction of protein structure, in: G.D. Fasman (Ed.), Prediction of Protein Structure and Principles of Protein Conformation, Plenum Press, New York, 1989, pp. 193–316.