22. Gianinazzi-Pearson, V., Dumas-Gaudot, E., Gollotte, A., Tahiri-Alaoui and Gianinazzi, S., *New Phytol.*, 1996, **133**, 45–57.
23. Schellenbaum, L., Berta, G., Ravolanirina, F., Tisserant, B., Gianinazzi, S. and Fitter, A. H., *Ann. Bot.*, 1991, **68**, 135–141.
24. Vestberg, M. and Estaun, V., in *Impact of Arbuscular Mycorrhiza on Sustainable Agriculture and Natural Ecosystem* (eds Gianinazzi, S. and Schuepp, H.), Birkhäuser Verlag, Basel, 1994, pp. 217–226.

ACKNOWLEDGEMENT. We thank UGC and DBT, New Delhi, Govt. of India, for partial funding.

NIRMAL SHANKAR SAHAY
AJIT VARMA*

Microbiology Unit,
School of Life Sciences,
Jawaharlal Nehru University,
New Delhi 110 067, India
*For correspondence.
(e-mail: avarma@jnuniv.ernet.in)

Inter-residue interactions in protein structures

The attainment of secondary and tertiary structures of globular proteins is the result of inter-residue interactions between amino acid residues along the polypeptide chain. Inter-residue interactions in proteins have been viewed from different perspectives, such as development of empirical potentials[1,2], partitioning of energetic components to the folding and stability of globular proteins[3], spatial distribution of residues between the interior and exterior of protein molecules[4], etc. Miller *et al.*[4] analysed the preference of residues to occur in the interior and surface of proteins based on the concept of solvent accessibility. Ponnuswamy and Gromiha[3] estimated the relative contributions of non-covalent interactions to the folding and stability of globular proteins. Recently, Bahar and Jernigan[5] have derived residue-specific potentials and utilized them to successfully discriminate the correct sequences in inverse protein folding experiments. Further, the importance of inter-residue interactions has been stressed by several researchers in the field[6–8].

Tanaka and Scheraga[9] categorized the inter-residue interactions into short, medium and long-range and proposed a hypothesis for protein folding by a three-step mechanism based on these interactions. For the past two decades, the concept of inter-residue interactions has been the main focus to understand the mechanism of protein folding. During the process of protein folding, residues which are sequential neighbours as well as those far away in the sequence interact cooperatively to form the native stable structure. Recently, the short, medium and long-range interactions have been classified according to the distance of separation between the residues along the polypeptide chain. This classification has been used successfully to address the problem of protein folding and sequence recognition[10,11]. In this communication, we reveal the contribution of inter-residue interactions in globular proteins belonging to different structural classes.

The crystallographic data of 150 globular proteins form the source of our study. The selected proteins were non-homologous and the structures were determined to a high resolution (resolution < 2.5 Å) and belong to four different structural classes, namely all-α, all-β, $\alpha+\beta$ and α/β based on the criteria of Kneller *et al.*[12] and Chou[13]. Atomic coordinates of all the proteins have been taken from the recent release of the Protein Data Bank (PDB) of Brookhaven National Laboratory[14,15]. Details about the PDB code, structural class and the fold[16,17] of all the 150 proteins and the secondary structural assignments[18] have been described in our previous article[19].

The inter-residue contacts have been analysed from the composition of surrounding residues for each amino acid residue in a protein[10,19,20]. The residues in a protein molecule are represented by their α-carbon atoms. Using the C_α coordinates, a sphere of radius 8 Å is fixed around each residue and the composition of surrounding residues associated with each residue is calculated[21]. It has been shown that the influence of each residue over the surrounding medium extends effectively only up to 8 Å (refs 22–24).

From the composition of surrounding residues within the sphere of 8 Å radius contributions from $< \pm 3$ residues in sequence level are treated as short-range contacts, ± 3 or ± 4 residues as medium-range contacts and $> \pm 4$ residues are treated as long-range contacts[10,25]. Further, the long-range contacts ($> \pm 4$ residues) are classified into several intervals with a step of 10 (4–10; 11–20; 21–30; 31–40; 41–50 and > 50). The number and percentage of long-range contacts in each interval for all the residues in the 150 globular proteins belonging to four different structural classes were computed. Further, contributions of all the 20 amino acid residues towards the long-range interactions in different intervals were estimated for the entire database.

It was observed that the average residue contacts in the medium-range are more in the all-α proteins (2.8 contacts/residue) compared to all-β proteins (0.92 contacts/residue) whereas average long-range contacts are more for the all-β proteins (5.2 contacts/residue) than all-α proteins (2.4 contacts/residue). The average residue contacts in the $\alpha+\beta$ and α/β proteins lie in the range between all-α and all-β proteins for both medium and long-range interactions[10].

Among the 20 amino acid residues, Met has the highest medium-range contact followed by Leu, Ala, Glu and Gln. Interestingly these residues are observed to be helix formers[26,27]. Pro has the lowest medium-range contact, indicating that it is not a favoured residue in α-helical conformation[26,28]. The residue preferences for each of the structural

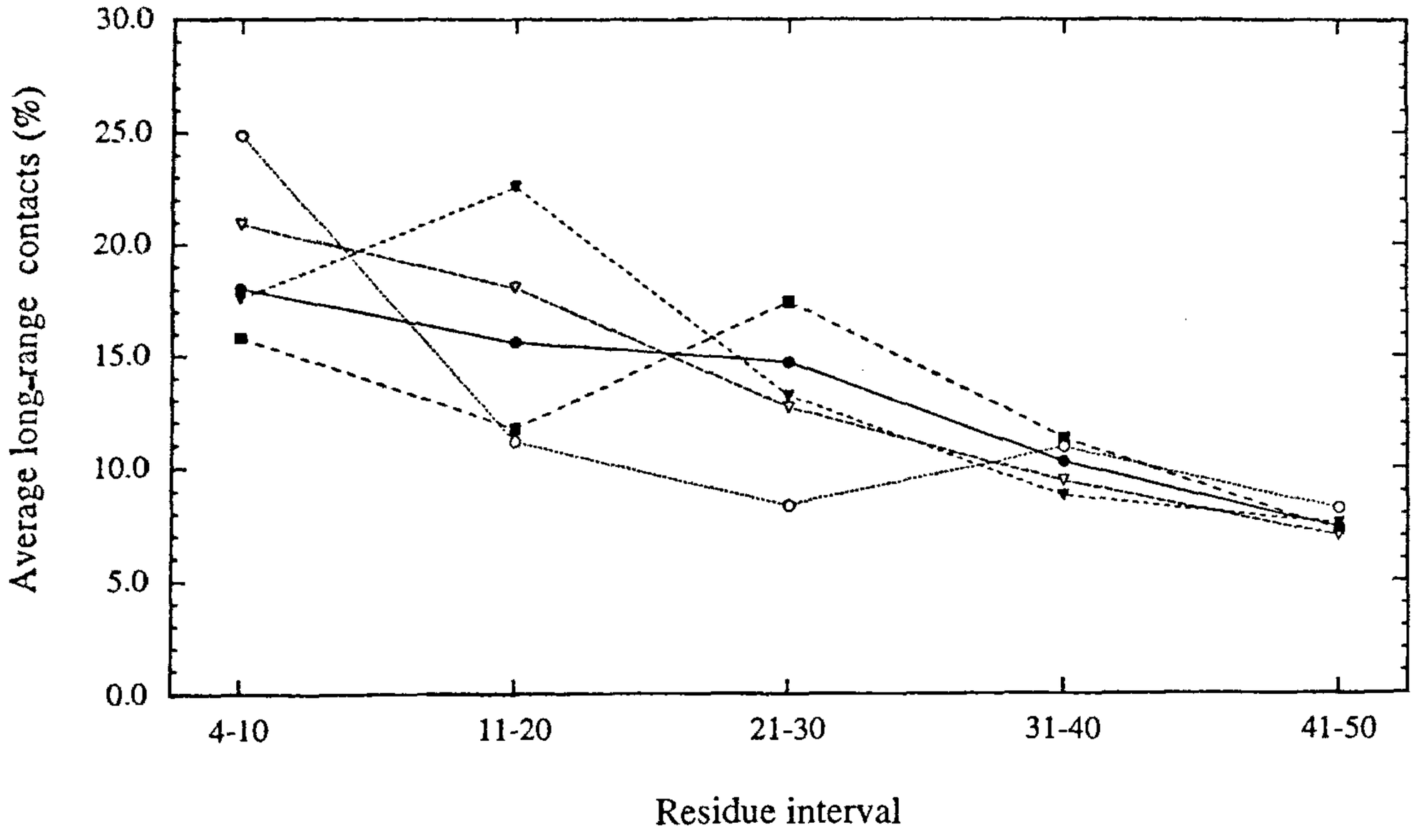

Figure 1. Average percentage of long-range contacts in different intervals for the four structural classes of globular proteins. O, all-α; ▼, all-β; ∇, $\alpha+\beta$; ■, α/β; ●, combined set.

classes have been analysed in our earlier article[10].

The residue Ile has the highest long-range contact followed by the hydrophobic residue Cys, Val, Tyr, Trp, Leu and Phe. It is noteworthy that all the aromatic residues have higher long-range contacts and they are important for the formation of a hydrophobic core during the process of protein folding. The statistical preference of the 20 amino acid residues to form medium and long-range contacts has been used to understand the stability of proteins caused by buried and partially buried mutations[29,30] and the thermal stability of globular proteins[31].

Figure 1 shows the average percentage of long-range contacts in different residue intervals for the four structural classes and the whole set of proteins. A perusal of Figure 1 clearly reveals the opposite trends between the folding of all-α and all-β proteins. The all-α class proteins have more long-range contacts in the 4–10 range and the all-β class proteins have more long-range contacts in the 11–20 range. This may be due to the specific hydrogen bonding pattern of α-helices and β-strands in these classes of proteins. The behaviour of proteins in $\alpha+\beta$ and α/β classes is surprising. The $\alpha+\beta$ class of proteins prefer the range 4–10 while the α/β class of proteins prefer the 21–30 range. The helical and strand segments are segregated into separate domains in $\alpha+\beta$ proteins and the proteins in this class behave like either all-α or all-β type. In the present analysis we found that the features of $\alpha+\beta$ proteins are similar to those of all-α proteins. In α/β class, the α-helices and β-strands occur alternatively and some residue distances are necessary to form β-strand and barrel, which leads to having higher contacts in the 21–30 range. A similar trend was also observed in our previous study of $(\alpha/\beta)_8$ barrel proteins[32,33]. These results indicate that the long-range contacts from different intervals play a considerable role in the folding of proteins belonging to different structural classes.

Further, we observed that the residues Cys, Ile and Val prefer the 11–20 range and all the other residues prefer the 4–10 range. Interestingly, Cys, Ile and Val are the topmost three hydrophobic residues[34]. These residues have a higher tendency of forming hydrophobic clusters and disulfide bridges due to long-range contacts and hence prefer the range 11–20.

The above results reveal the extent of the influence of inter-residue interactions in different structural classes as well as in different secondary structures. This provides a basis to understand the process of secondary structure formation[35], particularly the characteristic residue separation required for the formation of α-helical and β-strand structures.

Summarizing, the environment around each residue in a globular protein as defined in a sphere of 8 Å radius can be conveniently partitioned as comprising residues that contribute to short, medium and long-range interactions. The dominance of medium-range interactions in the formation of α-helices and that of long-range interactions in the formation of β-strands has been brought out. Further, the stabilization of proteins belonging to different structural classes through non-covalent interactions, and the distinguishing characteristics of each class have been understood. It is envisaged that these results may be incorporated in protein

structure prediction algorithms as well as in protein design experiments.

1. Zhang, L. and Skolnick, J., *Protein Sci.*, 1998, **7**, 111–122.
2. Seno, F., Maritan, A. and Banavar, J. R., *Proteins*, 1998, **30**, 244–248.
3. Ponnuswamy, P. K. and Gromiha, M. M., *J. Theor. Biol.*, 1994, **166**, 63–74.
4. Miller, S., Janin, J., Lesk, A. M. and Chothia, C., *J. Mol. Biol.*, 1987, **196**, 641–656.
5. Bahar, I. and Jernigan, R. L., *J. Mol. Biol.*, 1997, **266**, 195–214.
6. Gugolya, Z., Dosztanyi, Z. and Simon, I., *Proteins*, 1997, **27**, 360–366.
7. Selbig, J. and Argos, P., *Proteins*, 1998, **31**, 172–185.
8. Drablos, F., *Bioinformatics*, 1999, **15**, 501–509.
9. Tanaka, S. and Scheraga, H. A., *Proc. Natl. Acad. Sci. USA*, 1975, **72**, 3802–3806.
10. Gromiha, M. M. and Selvaraj, S., *J. Biol. Phys.*, 1997, **23**, 151–162.
11. Miyazawa, S. and Jernigan, R. L., *Proteins*, 1999, **36**, 347–356.
12. Kneller, D. G., Cohen, F. E. and Langridge, R., *J. Mol. Biol.*, 1990, **214**, 171–182.
13. Chou, K. C., *Proteins*, 1995, **21**, 319–344.
14. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, B. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. and Tasumi, M., *J. Mol. Biol.*, 1977, **112**, 535–542.
15. Abola, E., Bernstein, F. C., Bryant, S. H., Koetzle, T. F. and Weng, J., in *Crystallographic Databases – Information Content, Software Systems, Scientific Applications* (eds Allen, F. H., Bergerhoft, G. and Sievers, R.), 1987, pp. 107–132.
16. Hubbard, T. J. P., Ailey, B., Brenner, S. E., Murzin, A. G. and Chothia, C., *Nucleic Acids Res.*, 1999, **27**, 254–256.
17. Orengo, C. A., Pearl, F. M. G., Bray, J. E., Todd, A. E., Martin, A. C., Lo-Conte, L. and Thornton, J. M., *Nucleic Acids Res.*, 1999, **27**, 275–279.
18. Kabsch, W. and Sander, C., *Biopolymers*, 1983, **22**, 2577–2637.
19. Gromiha, M. M. and Selvaraj, S., *Biophys. Chem.*, 1999, **77**, 49–68.
20. Gromiha, M. M. and Selvaraj, S., *J. Biol. Phys.*, 1997, **23**, 209–217.
21. Selvaraj, S. and Gromiha, M. M., *J. Protein Chem.*, 1998, **17**, 407–415.
22. Manavalan, P. and Ponnuswamy, P. K., *Arch. Biochem. Biophys.*, 1977, **184**, 476–487.
23. Manavalan, P. and Ponnuswamy, P. K., *Nature*, 1978, **275**, 673–674.
24. Ponnuswamy, P. K., *Prog. Biophys. Mol. Biol.*, 1993, **59**, 57–103.
25. Gromiha, M. M. and Selvaraj, S., *Prep. Biochem. Biotech.*, 1999, **29**, 339–351.
26. Fasman, G. D., in *Prediction of Protein Structure and Principles of Protein Conformation* (ed. Fasman, G. D.), Plenum Press, NY, 1989, pp. 193–316.
27. Gromiha, M. M. and Ponnuswamy, P. K., *Int. J. Pept. Protein Res.*, 1995, **45**, 225–240.
28. MacArthur, M. W. and Thornton, J. M., *J. Mol. Biol.*, 1991, **218**, 397–412.
29. Gromiha, M. M., Oobatake, M., Kono, H., Uedaira, H. and Sarai, A., *Protein Eng.*, 1999, **12**, 549–555.
30. Gromiha, M. M., Oobatake, M., Kono, H., Uedaira, H. and Sarai, A., *J. Protein Chem.*, 1999, **18**, 565–578.
31. Gromiha, M. M., Oobatake, M. and Sarai, A., *Biophys. Chem.*, 1999, **82**, 51–67.
32. Selvaraj, S. and Gromiha, M. M., *J. Protein Chem.*, 1998, **17**, 691–697.
33. Gromiha, M. M. and Selvaraj, S., *Period. Biol.*, 1999 (in press)
34. Ponnuswamy, P. K. and Gromiha, M. M., *Int. J. Pept. Protein Res.*, 1993, **42**, 326–341.
35. Gromiha, M. M. and Selvaraj, S., *Protein Eng.*, 1998, **11**, 249–251.

S. SELVARAJ
M. MICHAEL GROMIHA*,†

Department of Physics,
Bharathidasan University,
Tiruchirappalli 620 024, India
**Tsukuba Life Science Center,*
The Institute of Physical and Chemical Research (RIKEN),
3-1-1 Koyadai, Tsukuba,
Ibaraki 305-0074, Japan
†For correspondence.
(e-mail: gromiha@rtc.riken.go.jp)