Influence of Medium and Long Range Interactions in $(\alpha/\beta)_8$ Barrel Proteins

M. MICHAEL GROMIHA¹ and S. SELVARAJ²

¹ *Tsukuba Life Science Center, The Institute of Physical and Chemical Research (RIKEN), Tsukuba, Japan;* ² *Department of Physics, Bharathidasan University, Tiruchirapalli, Tamil Nadu, India*

Accepted in final form 3 November 1997

Abstract. The residue-residue contacts and the role of medium and long range interactions in 36 $(\alpha/\beta)_8$ barrel proteins have been analysed. The influence of long range contacts in the formation of physico-chemically similar clusters, and the preference of amino acid residues towards long range contacts have also been studied. The results reveal a nearly uniform level of medium and long range contacts in most of the proteins. The residues Gln and Ala have highest medium range contacts and the residue Pro has the lowest medium range contacts. The residue Cys has the highest long range contact followed by other hydrophobic residues namely Val, Ile and Leu. In the physico-chemically similar clusters identified in these proteins, 25–40 percent residues are influenced by long range contacts, and the residues Cys, Ile, Val and Met are the most preferred ones.

Key words: $(\alpha/\beta)_8$ barrel proteins, Interactions, Medium and long range, Residue contacts, Tertiary structure

1. Introduction

The attainment of the secondary and tertiary structures of globular proteins is the result of intra molecular interactions between the constituent residues along the polypeptide chain. A number of investigations based on the structural co-ordinates obtained from x-ray crystallography of these proteins have been carried out in the past to analyse the residue contacts and the role of various interactions in the folding and stability of globular proteins [1–5]. Further, the free energy of contacts between pairs of amino acid residues [6], the preferred environment and co-operative behaviour of amino acid residues [7], interactions of each kind of side chain with specified atom type and other side chains [8], residue-residue preference/association potentials [1, 2, 9, 10], effective inter residue contact energies [11], aromatic-aromatic and polar interactions [12], side chain clusters [13, 14] and side chain structures on residue-residue associations [15] have been reported. Karlin's group proposed a number of distance measures between residues in protein structures [16] and developed methods for identifying and analysing statistically significant residue clusters that occur in three dimensional protein structures [17, 18]. Moreover, the distribution of short, medium and long range interactions in extra and intra cellular localizations of proteins were reported [19]. Recently, we have analysed the influence of medium and long range interactions in different

structural classes of globular proteins based on a simple distance criterion [20]. Among these, the $(\alpha/\beta)_8$ barrel proteins belonging to the (α/β) structural class are distinct in that they all possess a common fold and yet do not show any similarity at the sequence level. In a recent work, based on the distance criterion, we have analysed the amino acid clustering pattern in these proteins and found a number of physico-chemically similar cluster pairs [21]. In the present work, the influence of medium and long range interactions in these proteins and especially among the similar cluster pairs have been analysed with a view to understand the underlying mechanism(s) responsible for the common fold.

2. Materials and methods

Protein structural data base. The crystallographic data of 36 $(\alpha/\beta)_8$ barrel proteins, derived from the information available in the literature [22–24] forms the source for the present study. The proteins were selected on the criteria that their structures were solved to high resolution (resolution < 2.5 Å). The three dimensional structures of all the 36 $(\alpha/\beta)_8$ barrel proteins were taken from the recent release of the Protein Data Bank of Brookhaven National Laboratory [25–27].

Computation of surrounding residues. The residues in a protein molecule are represented by their α -carbon atoms. The centre is fixed at the α -carbon atom of the first (N-terminal) residue and the distance between this atom and the rest of the α -carbon atoms in the protein molecule is computed. The composition of the surrounding residues associated with this residue is calculated for a sphere of radius 8 Å, which has been shown to be the required volume of the medium within which a residue in a protein molecule is noted to exert a detectable influence [7, 28]. The procedure is repeated each time by moving the centre to the successive α -carbon atom along the polypeptide chain to compute the composition of surrounding residues, for all the residues in a given protein.

Delineation of medium and long range contacts in $(\alpha/\beta)_8$ barrel proteins. The residues coming within a sphere of radius 8 Å for each residue in all the $(\alpha/\beta)_8$ barrel proteins (comprising of a total of 14221 residues) were computed as described above. For a given residue, the composition of surrounding residues is analysed in terms of their location at the sequence level and the contribution due to short (± 2 residues along the sequence [29]), medium (± 4 residues along the sequence [30]) and long range ($> \pm 4$ residues [20]) interactions are computed. The number of short, medium and long-range contacts for all the residues in all 36 $(\alpha/\beta)_8$ barrel proteins and the contribution of all the 20 amino acid residues towards short, medium and long range interactions were computed for the entire database.

Long range contacts in physico-chemically similar clusters of $(\alpha/\beta)_8$ barrel proteins. Physico-chemically similar clusters in $(\alpha/\beta)_8$ barrel proteins were selected

as detailed in our previous work [21]. We used five sets of physico-chemical properties of amino acid residues used by Argos [31]. Basically, the surrounding residues around each residue were replaced by numerical values of the physico-chemical properties of amino acid residues (for example the property, hydrophobicity) and the correlation coefficient (CC) was computed among all pairs of residues. Pairs of residues having a $CC > 0.95$ were selected and designated as physico-chemically similar clusters. The computations were repeated with other properties, as well as for the combination of all the five properties. For all similar clusters with each property, the surrounding residues within 8 Å are selected and the influence of long range contacts ($> \pm 4$ residues along the sequence) to form similar clusters are analysed. Further, the preference of each amino acid residue contributing far from the sequence to form similar clusters were analysed.

3. Results and discussions

3.1. OCCURRENCE OF RESIDUES IN THE SHORT, MEDIUM AND LONG RANGE INTERACTIONS

The total number of contacts and average contacts per residue for the short, medium and long range interactions computed for the set of 36 $(\alpha/\beta)_8$ barrel proteins are given in Table I.

For all the 14221 residues, 56616 short range contacts, 28252 medium range contacts and 60668 long range contacts were found. On an average, within the sphere of 8 Å radius, 4 residues contribute towards short range interactions, 2 residues towards medium range and 4 residues towards long range interactions. This result shows that a significant number of residues far away in the sequence can come together through long range interactions to maintain the tertiary structure of $(\alpha/\beta)_8$ barrel proteins. The average residue contacts for $(\alpha/\beta)_8$ barrel proteins are similar to other (α/β) proteins. A perusal of Table I further reveals that while average residue contacts due to short range interactions remain constant, those due to medium and long range contacts seem to balance each other in such a way that the net contacts remain around the value 6. This trend seems to be present in all β -proteins, $\alpha + \beta$ proteins and α/β proteins [20]. Further, even though most of the proteins used in this study did not form part of the database of proteins used in our previous study [20], the average residue contacts for the $(\alpha/\beta)_8$ barrel proteins are similar to other (α/β) proteins. This may reflect the common intra molecular interaction pattern in these proteins.

The relationship between the number of residues and the total number of interactions are shown in Figure 1. We note from the figure that the total number of contacts are proportional to the total number of residues ($r = 0.996$) in the considered 36 $(\alpha/\beta)_8$ barrel proteins, a similar trend seen for other globular proteins [19].

In the medium range, the highest value of 2.44 contacts/residue is found for the protein Tryptophan synthase (1 WSY) and the lowest value of 1.43 is found for

Table I. Average residue contacts in the short, medium and long range interactions in a set of 36 $(\alpha/\beta)_8$ barrel proteins

No.	PDB file name	N	Average residue contacts		
			Short	Medium	Long
1	5XIA	393	3.985 (1566)	2.260 (888)	3.374 (1326)
2	2TAA	478	3.987 (1906)	1.611 (770)	4.423 (2114)
3	1MLI	96	3.938 (378)	1.646 (158)	2.792 (268)
4	1YPI	247	3.976 (982)	2.065 (510)	3.773 (932)
5	1ALD	363	3.983 (1446)	2.110 (766)	4.391 (1594)
6	1RUB	437	3.959 (1730)	2.069 (904)	4.078 (1782)
7	1WSY	248	3.927 (974)	2.444 (606)	3.629 (900)
8	1CDG	686	3.991 (2738)	1.426 (978)	5.076 (3482)
9	1FCB	494	3.976 (1964)	1.976 (976)	3.814 (1884)
10	1GOX	350	3.966 (1388)	2.063 (722)	4.011 (1404)
11	1MNS	357	3.983 (1422)	1.972 (704)	4.482 (1600)
12	1TIM	247	3.976 (982)	2.105 (520)	3.765 (930)
13	2TMD	729	3.992 (2910)	1.874 (1366)	4.658 (3396)
14	4ENL	436	3.986 (1738)	2.096 (914)	4.693 (2046)
15	1BTM	251	3.976 (998)	2.183 (548)	4.454 (1118)
16	1PII	457	3.982 (1820)	1.952 (892)	4.276 (1954)
17	1NAR	289	3.979 (1150)	1.896 (548)	3.875 (1120)
18	1CTN	538	3.989 (2146)	1.561 (840)	4.922 (2648)
19	1BYB	490	3.988 (1954)	1.959 (960)	4.127 (2022)
20	3CBH	365	3.984 (1454)	1.956 (714)	4.241 (1548)
21	2CMD	312	3.981 (1242)	2.128 (664)	4.071 (1270)
22	3ICD	414	3.986 (1650)	2.014 (834)	4.338 (1796)
23	1TSY	316	3.981 (1258)	1.930 (610)	3.367 (1064)
24	1NAL	291	3.979 (1158)	2.302 (670)	3.732 (1086)
25	8RUC	467	3.987 (1862)	1.953 (912)	4.206 (1964)
26	1PKY	464	3.974 (1844)	1.866 (866)	4.664 (2164)
27	1GHR	306	3.980 (1218)	2.026 (620)	4.268 (1306)
28	2ACQ	315	3.981 (1254)	1.924 (606)	3.746 (1180)
29	1MUC	360	3.978 (1432)	2.050 (738)	4.506 (1622)
30	1SCU	388	3.985 (1546)	1.912 (742)	4.201 (1630)
31	1MIN	522	3.989 (2082)	2.284 (1192)	3.598 (1878)
32	2DRI	271	3.978 (1078)	2.118 (574)	4.804 (1302)
33	1BNH	456	3.987 (1818)	2.026 (924)	5.013 (2286)
34	1DIK	869	3.986 (3464)	2.290 (1990)	4.518 (3926)
35	1DHR	236	3.975 (938)	1.915 (452)	4.186 (988)
36	1NIP	283	3.979 (1126)	2.028 (574)	4.021 (1138)
Complete set		14221	3.981 (56616)	1.987 (28252)	4.267 (60668)

N is the total number of residues in the protein. The number of contacts are given in brackets.

Notes to Table I: 5XIA – Xylose isomerase; 2TAA – Taka amylase; 1MLI – Muconolactone isomerase; 1YPI – Triose phosphate isomerase (yeast); 1ALD – Aldolase A; 1RUB – Rubisco (ribulose-1,5-bisphosphate) carboxylase; 1WSY – Tryptophan synthase; 1CDG – Cyclodextrin glycosyl transferase; 1FCB – Flavocytochrome b2; 1GOX – Glycolate oxidase; 1MNS – Mandelate racemase; 1TIM – Triose phosphate isomerase (chicken); 2TMD – Trimethylamine dehydrogenase; 4ENL – Enolase; 1BTM – Triose phosphate isomerase (Escherichia coli); 1PII – Anthranilate isomerase; 1NAR – Narbonin; 1CTN – Chitinase A; 1BYB – Beta amylase; 3CBH – Cellobio hydrolase; 2CMD – Malate dehydrogenase; 3ICD – Isocitrate dehydrogenase; 1TSY – Thymidylate synthase; 1NAL – Acetyl neuraminatase lyase; 8RUC – Ribulose-1,5-bisphosphate carboxylase oxygenase; 1PKY – Pyruvate kinase; 1GHR – Beta-glucanase; 2ACQ – Aldose reductase; 1MUC – Muconate lactonizing enzyme; 1SCU – Succinyl-co A synthetase; 1MIN – Nitrogenase molybdenum iron protein; 2DRI – Ribose-binding protein; 1BNH – Ribonuclease inhibitor; 1DIK – Pyruvate phosphate dikinase; 1DHR – Dihydropteridine reductase; 1NIP – Nitrogenase iron protein.

Figure 1 Plot showing the relationship between the number of residues and the total number of interactions in $(\alpha/\beta)_8$ barrel proteins.

1 CDG. These values are in the range seen in α/β class of globular proteins [20]. In the long range the highest value of 5.08 is observed for 1 CDG, and the lowest value of 2.79 is found for Muconolactone isomerase (1 MLI). The highest value is comparable to 4 ADH (5.107) and the lowest contact is similar to that of 3 ADK (2.784), both belonging to α/β class [20].

Comparison of medium and long range contacts in $(\alpha/\beta)_8$ barrel proteins

Residue	Medium	Long	Ratio
Ala	2.41	4.36	1.81
Asp	1.94	3.07	1.58
Cys	1.90	6.62	3.48
Glu	2.20	2.77	1.26
Phe	2.00	4.59	2.29
Gly	1.67	4.57	2.74
His	2.00	3.80	1.91
Ile	1.89	5.65	2.99
Lys	2.14	3.01	1.41
Leu	2.21	4.83	2.19
Met	2.29	4.31	1.88
Asn	1.91	3.85	2.01
Pro	1.39	4.23	3.03
Gln	2.43	3.30	1.36
Arg	2.23	3.62	1.62
Ser	1.77	3.96	2.24
Thr	1.66	4.49	2.71
Val	1.82	5.69	3.12
Trp	1.79	4.68	2.62
Tyr	1.81	4.74	2.62

3.2. PREFERENCE OF RESIDUES IN THE MEDIUM AND LONG RANGE CONTACTS

Medium range contacts. The average medium range contacts computed for all the 20 types of amino acid residues in $(\alpha/\beta)_8$ barrel proteins are given in Table II. From this table we observe that the residue Gln has the highest medium range contacts than other residues. Also the residues Ala, Met, Arg, Leu and Glu have values greater than 2.2. These 6 residues with topmost priority are found in the case of α/β globular proteins with highest contacts/residue. The residue Pro has the lowest medium range contacts. This may be due to the influence of the presence of 8 alpha helical segments in $(\alpha/\beta)_8$ barrel proteins, which are mainly occupied by residues Met, Leu, Ala, Glu (strong helix formers) and Gln, a helix former [32, 33] and Pro is not a favoured residue in α helical conformation.

Long range contacts. We also find from Table II that the residue Cys has the highest long range contact (6.62) followed by Val and Ile. This may be due to the formation of disulfide bridges and hydrophobic clusters. It is noteworthy that low values are obtained for the charged residues and the lowest is for Glu (2.77).

Long range contacts in physico-chemically similar clusters of $(\alpha/\beta)_8$ barrel proteins

Property	N _{sim}	N _{tot}	N _{long}	% _{long}
H_p	504	9800	3418	34.87
t	3943	77483	28706	37.05
B	2468	50865	19773	38.87
μ	3581	63823	19007	29.78
β	2822	57279	22042	38.48
Combined	126	2051	566	27.60

N_{sim} – number of similar cluster pairs.

N_{tot} – total number of surrounding residues in similar clusters.

N_{long} – number of residues from long range contacts.

%_{long} – percentage of residues contributed from long range contacts.

Preference of amino acid residues to form similar clusters of $(\alpha/\beta)_8$ proteins due to long range contacts

Residue	t	B	H _p	μ	β	Combined
Ala	34.68	35.74	30.49	28.06	35.72	33.23
Asp	28.47	23.52	19.41	20.33	25.50	24.22
Cys	49.35	49.05	43.53	49.58	51.50	49.53
Glu	28.46	33.40	29.77	21.85	30.44	28.25
Phe	41.40	40.03	35.57	32.55	41.75	38.41
Gly	37.26	31.02	33.77	28.42	31.24	34.27
His	35.08	50.09	33.98	29.14	42.48	38.34
Ile	41.73	48.86	45.78	39.32	50.37	45.11
Lys	31.51	27.52	23.41	19.60	26.96	25.42
Leu	35.50	42.59	39.47	32.96	39.62	38.19
Met	41.41	48.07	40.48	32.31	48.66	41.83
Asn	43.51	28.25	30.58	26.81	30.19	33.44
Pro	36.01	36.95	34.80	34.81	39.78	36.66
Gln	37.70	25.05	24.03	23.71	25.33	27.37
Arg	35.00	35.50	28.46	26.85	36.37	32.17
Ser	37.15	38.56	33.94	29.20	37.47	35.05
Thr	39.15	40.16	41.77	33.16	38.30	37.70
Val	39.71	48.26	44.47	38.17	47.60	44.35
Trp	33.63	54.85	43.64	29.52	52.61	38.39
Tyr	48.65	44.03	38.68	30.52	40.83	39.03

3.3. PHYSICO-CHEMICALLY SIMILAR CLUSTERS OF $(\alpha/\beta)_8$ BARREL PROTEINS

Long range contacts. The total number of contacts and the number of long range contacts observed in physico-chemically similar clusters of $(\alpha/\beta)_8$ barrel proteins are given in Table III. It is interesting to note that for all the properties 25–40%

of residues are influenced by long range contacts to form similar clusters. 39% residues are noted to contribute far from the sequence to form similar clusters for the property bulk (B) and antiparallel beta strand preference (β).

Residue preferences. The influence of all the 20 amino acid residues to contribute far from the sequence in forming similar clusters for various properties are given in Table IV. For the property bulk (B), the residue Trp is the most preferred residue followed by His, Cys, Ile and Val. For the property anti parallel beta strand preference (β), again Trp is the most preferred residue followed by Cys, Ile, Met and Val. The residues Cys, Ile, Val and Met are the most preferred ones for the combination of all the five properties.

4. Conclusions

The present study reveals the importance of medium range and long range interactions in the formation and stabilization of $(\alpha/\beta)_8$ barrel fold. The total number of residue-residue contacts are proportional to the number of residues. The total number of average residue contacts due to medium and long range interactions is maintained at an uniform level in these proteins. Among the physico-chemically similar clusters identified in these proteins, the residues Cys, Ile, Val and Met are the most preferred ones for contributing to long range interactions.

Acknowledgments

MMG is a recipient of an ICGEB fellowship. SS wishes to thank Professor M. Lakshmanan for his encouragement and support.

References

1. MacArthur, M.W. and Thornton, J.M.: *J. Mol. Biol.* **218** (1991), 397–412.
2. Karpeisky, M.Ya. and Ilyin, V.A.: *J. Mol. Biol.* **224** (1992), 629–638.
3. Vtyurin, N.: *Proteins* **15** (1993), 62–70.
4. Ponnuswamy, P.K. and Gromiha, M.M.: *J. Theor. Biol.* **166** (1994), 63–74.
5. Gobel, U., Sander, C., Schneider, R. and Valencia, A.: *Proteins* **18** (1994), 309–317.
6. Tanaka, S. and Scheraga, H.A.: *Macromolecules* **9** (1976), 945–950.
7. Manavalan, P. and Ponnuswamy, P.K.: *Arch. Biochem. Biophys.* **184** (1977), 476–487.
8. Warme, P.K. and Morgan, R.S.: *J. Mol. Biol.* **118** (1978), 289–304.
9. Narayana, S.V.L. and Argos, P.: *Int. J. Peptide Protein Res.* **24** (1984), 25–39.
10. Viswanathan, V.N.: *Int. J. Biol. Macromol.* **9** (1987), 39–48.
11. Miyazawa, S. and Jernigan, R.L.: *Macromolecules* **18** (1985), 534–552.
12. Burlley, S.K. and Petsko, G.A.: *Science* **229** (1985), 23–29.
13. Heringa, J. and Argos, P.: *J. Mol. Biol.* **220** (1991), 151–171.
14. Magalhaes, A., Margret, B., Hoflack, J., Gomes, J.N. and Scheraga, H.A.: *J. Protein Chem.* **13** (1994), 195–215.
15. Muthusamy, R. and Ponnuswamy, P.K.: *J. Theor. Biol.* **153** (1991), 25–40.
16. Karlin, S., Zucker, M. and Brocchieri, L.: *J. Mol. Biol.* **239** (1994), 227–248.
17. Karlin, S. and Zhu, Z.Y.: *Proc. Natl. Acad. Sci. USA* **93** (1996), 8344–8349.

18. Zhu, Z.Y. and Karlin, S.: *Proc. Natl. Acad. Sci. USA* **93** (1996), 8350–8355.
19. Gugolya, Z., Dostanyi, Zs. and Simon, I.: *Proteins* **27** (1997), 360–366.
20. Gromiha, M.M. and Selvaraj, S.: *J. Biol. Phys.* **23** (1997), 151–162.
21. Selvaraj, S. and Gromiha, M.M.: *Protein Engng.* (submitted) (1997).
22. Farber, G.K. and Petsko, G.A.: *Trends Biochem. Sci.* **15** (1990), 228–234.
23. Holm, L. and Sandor, C.: *J. Mol. Biol.* **233** (1993), 123–138.
24. Sergeev, Y. and Lee, B.: *J. Mol. Biol.* **244** (1994), 168–182.
25. Bernstein, F.C., Koetzle, T.F., Williams, J.B., Meyer, E.F. Jr., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M.: *J. Mol. Biol.* **112** (1977), 535–542.
26. Abola, E., Bernstein, F.C., Bryant, S.H., Koetzle, T.F. and Weng, J.: Protein data Bank. In: F.H. Allen, G. Bergerhoff and R. Sievers (eds.), *Crystallographic Databases – Information Content, Software Systems, Scientific Applications*, 1987, pp. 107–132.
27. Protein Data Bank, Brookhaven National Laboratory, Upton, NY, September, 1996.
28. Manavalan, P. and Ponnuswamy, P.K.: *Nature* **275** (1978), 673–674.
29. Ponnuswamy, P.K., Prabakaran, M. and Manavalan, P.: *Biochim. Biophys. Acta.* **623** (1980), 301–306.
30. Ponnuswamy, P.K., Warme, P.K. and Scheraga, H.A.: *Proc. Natl. Acad. Sci.* **70** (1973), 830–833.
31. Argos, P.: *J. Mol. Biol.* **193** (1987), 385–396.
32. Fasman, G.D.: In: G.D. Fasman (ed.), *Prediction of Protein Structure and Principles of Protein Conformation*, 1989, New York: Plenum Press, pp. 193–316.
33. Gromiha, M.M. and Ponnuswamy, P.K.: *Int. J. Peptide Protein Res.* **45** (1995), 225–240.

Address for correspondence: Dr M. Michael Gromiha, Genetic Information Lab, Tsukuba Life Science Center, The Institute of Physical and Chemical Research (RIKEN), 3-1-1 Koyadai, Tsukuba, Ibaraki 305, Japan

Phone: +81 298 36-9082; Fax: +81 298 36-9080; E-mail: gromiha@rtc.riken.go.jp