Journal of Biological Physics **23:** 151–162, 1997.

Influence of Medium and Long Range Interactions in Different Structural Classes of Globular Proteins

M. MICHAEL GROMIHA[1] and S. SELVARAJ[2]
[1]*International Centre for Genetic Engineering and Biotechnology, Trieste, Italy;* [2]*Department of Physics, Bharathidasan University, Tamil Nadu, India*

Accepted in final form 1 May 1997

Abstract. An analysis of the dependence known three dimensional structure of globular proteins on their residue contacts and their interactions provides much information about their folding and stability. In this work, we analyse the residue-residue contacts and the role of medium and long range interactions in globular proteins belonging to different structural classes. The results show that while medium range interactions predominate in all alpha class proteins, long range interactions predominate in all beta class. The residues Pro and Gly are found to have lowest medium range contacts, probably due to their helix breaking tendency. The hydrophobic residues Ile, Val and Tyr have higher long range contacts, and hence may serve as good nucleation centres. Further, the role of charged residues and disulfide bridges in these interactions are also discussed.

Key words: Globular proteins, Interactions, Residue contacts, Structural class, Tertiary structure

1. Introduction

The folding of a polypeptide chain into a unique three dimensional structure involves numerous atomic/group interactions. Various investigations have been performed to analyse the residue contacts and the role of various interactions in the structure and stability of globular proteins [1–8]. The crystal structural data for the available proteins has been used as a major tool for these studies. Based on this, the free energy contacts between pairs of amino acid residues [9], the preferred environment and co-operative behaviour of amino acid residues [10], interactions of each kind of side chain with specified atom type and other side chains [11], residue-residue preference/association potentials [3, 4, 12, 13], effective inter residue contact energies [14], aromatic-aromatic and polar interactions [15, 16] and side chain clusters [17, 18] has been reported. Recently, Muthusamy & Ponnuswamy [19] analysed the side chain structures on the residue-residue associations in globular proteins. Very recently, Karlin et al. [20] proposed a number of distance measures between residues in protein structures based on average, minimum and maximum distances of all the atom co-ordinates or with respect to side chain co-ordinates only.

It has also been shown that short and medium interactions play a dominant role in determining the conformation of amino acid residues in a protein [21, 22]. The importance of long range interactions in protein structure prediction has also

been stressed by many workers [23, 24]. In this article, we relate the medium and long range interactions with the residue-residue contacts computed in a set of 63 globular protein molecules belonging to different structural classes. Finally the role of ion pairs, hydrophobic residues and disulfide bridges has been explored.

2. Material and Methods

Data base. The crystallographic data of 63 protein molecules taken from the Brookhavan Data Base [25], forms the source for our present study. The proteins selected were nonhomologous and their structures was determined to a high resolution (R $<$ 2.5 È). Among these, 14 proteins belong to the structural class all-α, 16 belong to all-β, 15 belong to $\alpha + \beta$ and 18 to α/β. The total number of residues examined were 10877.

Each residue in the protein molecule is represented by its α carbon atom. 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 was calculated for a sphere of radius 8 È, which has been shown to be the required volume of the medium within which a residue in a protein molecule which is known to exert a detectable influence [10, 26]. The procedure was 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.

From the composition of surrounding residues within the sphere of 8 È radius, the contribution due to short ($\pm$ 2 residues along the sequence), medium ($\pm$ 4 residues along the sequence) and long range ($>$ $\pm$ 4 residues) interactions are computed.

For a given residue, the composition of surrounding residues (within a sphere of 8 È radius) was analysed in terms of their location at the sequence level. Those residues that are within the distance of 2 residues from the central residue were considered to be contributions from short range interaction type [27], and those within the distance of 4 residues as the medium range [22] and those $>$ 4 residues were considered as the long range contributions.

The number of short-range, medium-range and long-range contacts for all the residues in all 63 proteins was computed. The total number of contacts obtained for each type of interaction in a protein was divided by the total number of residues in a protein to calculate average contacts per residue. Also, the contacts were computed for the four structural classes. The contribution of each amino acid residues towards short, medium and long range interactions was also computed for the entire database as well as for the four structural classes.

Table I. Average residue contacts in the short, medium and long range interactions in a set of 63 globular proteins

No.	Protein Code	N	Average residue contacts		
			Short	Medium	Long
1	3CPV	108	3.944 (426)	2.722 (294)	2.500 (270)
2	3CYT	103	3.942 (406)	2.039 (210)	3.417 (352)
3	2MHR	118	3.949 (466)	2.983 (352)	1.644 (194)
4	2HCOA	141	3.957 (558)	3.106 (438)	2.113 (298)
5	2HCOB	146	3.959 (578)	3.096 (452)	2.137 (312)
6	2MBN	153	3.961 (606)	3.268 (500)	1.765 (270)
7	156B	100	3.920 (392)	3.380 (338)	2.040 (204)
8	1PPT	36	3.833 (138)	2.278 (82)	1.333 (48)
9	1GCN	29	2.793 (110)	2.897 (84)	0.069 (2)
10	1LH1	153	3.961 (606)	3.150 (482)	2.078 (318)
11	1MLT	26	3.769 (98)	3.462 (90)	0.077 (2)
12	1PP2	122	3.951 (482)	2.197 (268)	3.066 (374)
13	2CTS	437	3.986 (1742)	2.705 (1182)	2.847 (1244)
14	1TGS	56	3.893 (218)	1.107 (62)	3.357 (188)
15	4CHA	239	3.933 (940)	1.038 (248)	5.590 (1336)
16	3CNA	237	3.975 (942)	0.717 (170)	5.637 (1336)
17	3EST	240	3.975 (954)	1.058 (254)	5.433 (1304)
18	3EBX	62	3.919 (243)	0.645 (40)	4.903 (304)
19	3FAB	219	3.973 (870)	0.785 (172)	4.621 (1012)
20	2SOD	151	3.960 (598)	0.821 (124)	5.881 (888)
21	2PAB	114	3.947 (450)	0.807 (92)	4.596 (524)
22	1REI	107	3.944 (422)	0.748 (80)	5.290 (566)
23	1RDG	52	3.885 (202)	1.038 (54)	3.500 (182)
24	1TPA	58	3.897 (226)	1.103 (64)	4.379 (254)
25	1PCY	99	3.939 (390)	0.828 (82)	4.869 (482)
26	2APR	325	3.982 (1294)	1.022 (332)	5.415 (1760)
27	1NXB	62	3.903 (242)	0.548 (34)	4.710 (292)
28	2AZA	129	3.953 (510)	1.163 (150)	4.791 (618)
29	1CTX	71	3.915 (278)	0.817 (58)	4.563 (324)
30	1GCR	174	3.966 (690)	0.862 (150)	5.598 (974)
31	1FDX	54	3.889 (210)	1.148 (62)	3.481 (188)
32	3INS	102	3.765 (384)	1.980 (202)	3.804 (388)
33	2LZM	164	3.963 (650)	2.720 (446)	2.500 (410)
34	3LYZ	129	3.953 (510)	2.124 (274)	3.659 (472)
35	2PAD	108	3.944 (426)	2.722 (294)	2.500 (270)
36	3BP2	122	3.951 (482)	2.328 (284)	2.754 (336)
37	1RNS	124	3.903 (484)	1.387 (172)	4.113 (510)
38	2SNS	141	3.957 (558)	1.759 (248)	4.255 (600)
39	2SSI	107	3.944 (422)	1.252 (134)	4.131 (442)
40	3TLN	316	3.981 (1258)	2.070 (654)	4.608 (1456)
41	2ACT	218	3.972 (866)	1.541 (336)	5.055 (1102)
42	2B5C	85	3.929 (334)	1.953 (166)	2.753 (234)
43	1HIP	85	3.929 (334)	1.341 (114)	3.835 (326)
44	2GAP	208	3.971 (826)	1.962 (408)	3.250 (676)
45	2CDV	107	3.944 (422)	1.720 (184)	2.654 (284)
46	3ADK	194	3.969 (770)	2.536 (492)	2.784 (540)
47	4ADH	374	3.984 (1490)	1.599 (598)	5.107 (1910)

Table I (continued).

48	3CPA	309	3.961 (1224)	1.942 (600)	4.557 (1408)
49	3DFR	162	3.963 (642)	1.271 (206)	3.889 (630)
50	3FXN	138	3.957 (546)	2.043 (282)	3.681 (508)
51	2GPD	333	3.982 (1326)	1.742 (580)	4.547 (1514)
52	3PGK	416	3.986 (1654)	1.899 (788)	4.376 (1816)
53	3LDH	329	3.982 (1310)	2.018 (664)	3.793 (1248)
54	2SBT	275	3.971 (1092)	1.629 (448)	5.702 (1568)
55	2CAB	256	3.977 (1018)	1.117 (286)	5.328 (1364)
56	1RHD	293	3.980 (1166)	1.843 (540)	3.993 (1170)
57	1SRX	108	3.944 (426)	2.093 (226)	3.630 (392)
58	1TIM	494	3.976 (1964)	2.121 (1048)	4.089 (2020)
59	2CPP	405	3.985 (1614)	2.326 (942)	3.575 (1448)
60	1ABP	306	3.980 (1218)	2.118 (648)	4.275 (1308)
61	1CRN	46	3.870 (178)	2.043 (94)	2.913 (134)
62	1OVO	224	3.893 (872)	1.321 (296)	3.813 (854)
63	1CTF	68	3.912 (266)	2.176 (148)	3.441 (234)
All α proteins (1–14)			3.951 (6826)	2.802 (4834)	2.361 (4076)
All β proteins (15–30)			3.950 (9294)	0.921 (2168)	5.201 (12210)
$\alpha + \beta$ proteins (31–45)			3.951 (8166)	1.925 (3978)	3.722 (7694)
α/β proteins (46–63)			3.970 (18776)	1.879 (8886)	4.242 (20066)
Complete set			3.959 (43062)	1.826 (19866)	4.049 (44046)

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

Abbreviations:

3CPV - Parvalbumin; 3CYT - Cytochrome c; 2MHR - Myo hemerythrin; 2HCOA - Hemoglobin A chain; 2HCOB - Hemoglobin B chain; 2MBN - Myoglobin; 156B - Cytochrome b; 1PPT - Avian pancreatic polypeptide; 1GCN - Glucagon; 1LH1 - Leghemoglobin; 1MLT - Mellitin; 1PP2 - Hydrolase; 2CTS - Citrate synthase; 1TGS - Trypsinogen; 4CHA - Chymotrypsin; 3CNA - Concanavalin A; 3EST - Tosyl elastase; 3EBX - Erabutoxin; 3FAB - Immunoglobulin; 2SOD - Superoxide dismutase; 2PAB - Prealbumin; 1REI - Bence Jones immunoglobulin; 1RDG - Rubredoxin; 1TPA - Trypsin; 1PCY - Plastocyanin; 2APR - Acid protease; 1NXB - Neurotoxin B; 2AZA - Azurin; 1CTX - Cobratoxin; 1GCR - Gamma crystallin 1FDX - Ferredoxin; 3INS - Insulin; 2LZM - Lysozyme T4; 3LYZ - Hen lysozyme; 2PAD - Papain; 3BP2 - Phospholipase A2; 1RNS - Ribonuclease S; 2SNS - Stap. nuclease; 2SSI - Strept. subtilisin inhibitor; 3TLN - Thermolysin; 2ACT - Actinidin; 2B5C - Cytochrome b5; 1HIP - High potential iron protein; 2GAP - Gene activator protein; 2CDV - Cytochrome c3; 3ADK - Adenylate kinase; 4ADH - Livel alcohol dehydrogenase; 3CPA - Carboxypeptidase A; 3DFR - Dihydrofolate reductase; 3FXN - Flavodoxin; 2GPD - Glyceraldehade phosphate dehydrogenase; 3PGK - Phosphoglycerate kinase; 3LDH - Lactate dehydrogenase; 2SBT - Subtilisin BPN'; 2CAB - Carbonic anhydrase B; 1RHD - Rhodanase; 1SRX - Thioredoxin; 1TIM - Triose phosphate isomerase; 2CPP - Cytochrome p450; 1ABP - Arabinose binding protein; 1CRN - Crambin; 1OVO - Ovomucoid; 1CTF - Ribosomal protein.

3. Results and Discussions

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 63 protein molecules is given in Table I.

For all the 10877 residues, 43062 short range contacts, 19866 medium range contacts and 44046 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 clearly shows that a significant number of residues far away in the sequence can come together through long range interactions to maintain the tertiary structure.

In the medium range, the highest value of 3.46 contacts/residue was found for the protein mellitin (1MLT) which belongs to the all α structural class. As more than two neighbouring residues come within the 8 Å radius sphere in an α helix (due to intra-chain hydrogen bonding) this effect is appreciable. The lowest value of 0.55 was found for neurotoxin (1NXB), an all-β protein, due to the extended nature of the backbone. It is interesting to note that both of the above proteins are small in size.

In our earlier work [7] we computed the radius for a set of globular proteins and found that the radius is directly related to the total number of residues in a protein. The total number of contacts was computed and then the values were normalized by dividing with the total number of residues. Hence, it is fair to compare the average contacts per residue between proteins of different size.

The reverse was observed for the long range contacts. The lowest value of 0.07 was observed for glucagon (1GCN), another small α type protein, having only two long range contacts. In this protein, 50 % of the residues are in helical conformation. Hence these residues prefer local interaction (short and medium) and the number of long range contacts are very minimal. The highest value of 5.88 was observed for superoxide dismutase (2SOD), an all-β type protein. These results reveal the dominance of medium range interactions in stabilizing alpha helices and the long range interactions the beta strands. The conformational studies on peptides and proteins also showed that the alpha helical structures are stabilized by short and medium range interactions (28). In the present study, we excluded the effect of short range interactions of alpha helical proteins due to (a) the abundance of such interactions arising from the helical backbone, and (b) those interactions due to nearest neighbour residues and hence the number of contacts are almost equal ($\sim$ 4) for all the proteins (except for 1GCN).

Variations Among Structural Class

From Table I, we also observe that average medium range contacts are highest (2.80) for all-α type proteins, and lowest (0.92) for all-β type proteins. This indicates the

vital influence of medium range contacts in all-α type proteins. In the $\alpha + \beta$ and α/β classes the medium and long range contacts lie in the range between all-α and all-β type proteins.

In the all-α type highest (3.46) medium range contacts are found in 1MLT and lowest (1.1) in 1TGS. It has also been shown that 1MLT acquires more electrostatic free energy due to high charge-helix dipole interactions and ion pairs [7]. In the all-β type highest (1.163) medium range contacts are found in 2AZA, and lowest (0.55) in 1NXB. The free energy studies on these proteins [7] showed that 2AZA acquires more hydrophobic free energy (0.73 kcal/mol per residue) than 1NXB (0.5 kcal/mol per residue). In the $\alpha + \beta$ type, highest medium range contacts is 2.72 in 2LZM and lowest (1.15) in 1FDX. In the α/β type highest (2.53) medium range contacts is found in 3ADK and lowest in (1.12) 2CAB. It is interesting to note that the lowest number of contacts per residue in all-α proteins is nearly equal to that of the highest number of contacts/residue in all-β proteins. This may be due to the presence of a minimum number of residues in helical segments of 1TGS (all-α type) and the occurrence of some residues in the helical conformation of 2AZA (all-β type).

Conversely, long range contacts are highest in the all-β type proteins and lowest in the all-α type proteins. This indicates the importance of long range effects in predicting the beta strands. Among the all-α proteins highest (3.42) long range contacts are observed for cytochrome C (3CYT) and lowest (0.07) for glucagon (1GCN). In the all-β type, highest (5.88) long range contacts are found in superoxide dismutase and lowest (3.5) in 1RDG. Interestingly, the lowest value in all-β proteins is greater than the highest value observed for the all-α type proteins. In the $\alpha + \beta$ type, highest (5.06) long range contacts are found in 2ACT and lowest (2.5) in 2LZM. In the α/β type, highest (5.7) long range contacts are found in 2SBT and lowest (2.78) in 3ADK. It is interesting to note that all the proteins with higher long range contacts acquire more hydrophobic free energy [7].

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 the four structural classes are given in Table II. In the all-α type, among the 20 amino acid residues, the residue Leu has the highest medium range contacts. Also the residues Phe, Met, Ala and His have values higher than 3.0. The residues Pro and Gly have lowest medium range contacts, due to the fact that these residues are helix breakers [29].

In the all-β class the highest medium range contacts was observed for Met (1.2) than other residues. Also the residues Trp, Asp, Gln, Cys and Ser have values greater than 1.0. The residue His has the lowest medium range contact.

In the $\alpha + \beta$ class, highest medium range contact was observed for the residue Cys (2.3). Also the residues Ala, Leu, Met, Gln, Phe, Glu and His had higher values

Table II. Average medium range contacts for the 20 amino acid residues in 4 structural classes of globular Proteins

Residue	Average medium range contacts			
	all α	all β	$\alpha+\beta$	α/β
Ala	3.04	0.98	2.26	2.08
Asp	2.63	1.16	1.90	1.76
Cys	2.84	1.04	2.30	1.92
Glu	2.67	0.80	2.15	2.24
Phe	3.26	0.78	2.17	1.99
Gly	2.11	0.84	1.71	1.67
His	3.02	0.67	2.04	1.88
Ile	2.93	0.91	1.88	1.80
Lys	2.83	0.90	1.91	1.96
Leu	3.38	0.86	2.23	2.17
Met	3.19	1.20	2.21	2.34
Asn	2.88	1.12	1.90	1.84
Pro	1.88	0.91	1.20	1.38
Gln	2.88	1.10	2.19	2.19
Arg	2.77	0.83	2.25	2.11
Ser	2.43	1.02	1.66	1.68
Thr	2.62	0.79	1.60	1.75
Val	2.88	0.77	1.70	1.61
Trp	2.85	1.17	1.93	1.96
Tyr	2.78	0.91	1.52	1.82

greater than 2.0. The residue Pro had the lowest medium range contact (1.2) among all the residues.

In the α/β type class, highest medium range contact was observed for the residue Met (2.34). Also the residues Glu, Gln, Leu, Arg and Ala had values greater than 2.0. The residue Pro had the lowest medium range contacts.

In all the 63 considered proteins, irrespective of the structural class the residue Met had the highest medium range contact (2.27). Also the residues Leu, Ala, Glu and Gln had more than 2 contacts per residue. From the analysis of the helix forming tendency of residues in a set of 63 proteins we observed that the residues involving more medium range contacts such as Met, Leu, Ala and Glu are very strong helix formers and the residue Gln is also a helix former [29, 30]. It is also interesting to note that in most of the classes (all-$\alpha, \alpha+\beta$ and α/β) Pro had the lowest medium range contact, indicating the fact that it is not a favoured residue in alpha helical conformation [3, 29, 30]. Also it may explain why prediction of alpha helices is more accurate than that of beta strands [30–32].

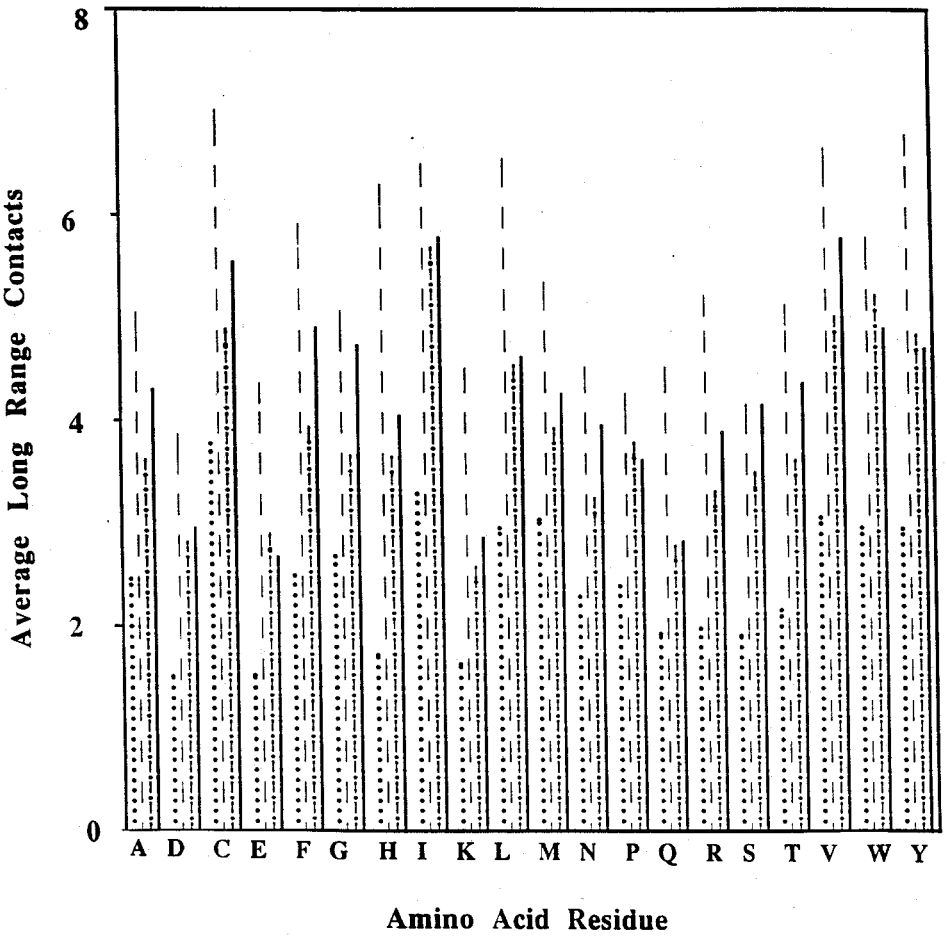

Figure 1. Variation of long range contacts for the 20 amino acid residues in different structural classes of globular proteins.
..... all α; - - - - - all β; –·–·– $\alpha + \beta$; — α/β.

Long Range Contacts. The variation of average long range contacts for the 20 types of amino acid residues are displayed in Figure 1. In the all-α class the residue Cys had the highest long range contact (3.78). This indicated that the disulfide bridge forming tendency of Cys is far apart in their sequence level [33]. In most of the proteins used in this study, a similar pattern was observed and hence Cys had a higher long range contact. Also the residues Ile, Val and Met had values greater than 3.0. It is interesting to note that all the four residues are hydrophobic in nature. This shows the formation of hydrophobic cores in proteins with residues which are far down in the sequence. The lowest value was found for Asp (1.51). Also the residues Glu, Lys and His had low values. It is noteworthy that all these residues

Table III. Comparison of medium and long range contacts in globular proteins

Residue	Medium	Long	Ratio*
Ala	2.11	3.92	1.86
Asp	1.80	2.85	1.58
Cys	1.88	5.55	2.95
Glu	2.09	2.72	1.30
Phe	1.98	4.53	2.29
Gly	1.53	4.31	2.82
His	1.98	3.77	1.90
Ile	1.77	5.58	3.15
Lys	1.96	2.79	1.42
Leu	2.19	4.59	2.10
Met	2.27	4.14	1.82
Asn	1.84	3.64	1.98
Pro	1.32	3.57	2.70
Gln	2.03	3.06	1.51
Arg	1.94	3.78	1.95
Ser	1.57	3.75	2.39
Thr	1.57	4.09	2.61
Val	1.63	5.43	3.33
Trp	1.90	4.83	2.54
Tyr	1.67	4.93	2.95

*Ratio $= \frac{\text{Average number of long range contacts}}{\text{Average number of medium range contacts}}$

are charged residues. This is consistent with the study of Matthews [34] that ion pairs prefer a local environment.

In the all-β class, highest long range contacts was observed for Cys (6.95). Corresponding values in decreasing order were for Tyr, Val, Leu, Ile and His. The lowest value was observed for Asp (3.79). In the $\alpha + \beta$ category, highest long range contact was observed for Ile (5.61). The residues Trp, Val, Cys and Leu have comparable values. Lowest long range contact was observed for Lys (2.5). In the α/β class, highest long range contact was observed for Ile (5.79). The residue Val also had an almost equivalent value (5.78). The residue Cys had a value of 5.5. Lowest long range contact was observed for Glu (2.68).

Overall, Ile had the highest (5.58) long range contact followed by Cys and Val. Lowest value was observed for Glu (2.72). Other residues having lowest long range contacts are Lys (2.79), and Asp (2.85). Again, residues having minimum number of long range contacts are charged residues. This indicates that the probability of forming ion pairs in the sequence level in this range ($> \pm 4$ residues) is minimum and most of the ion-pairs are formed between neighbouring residues [34, 35].

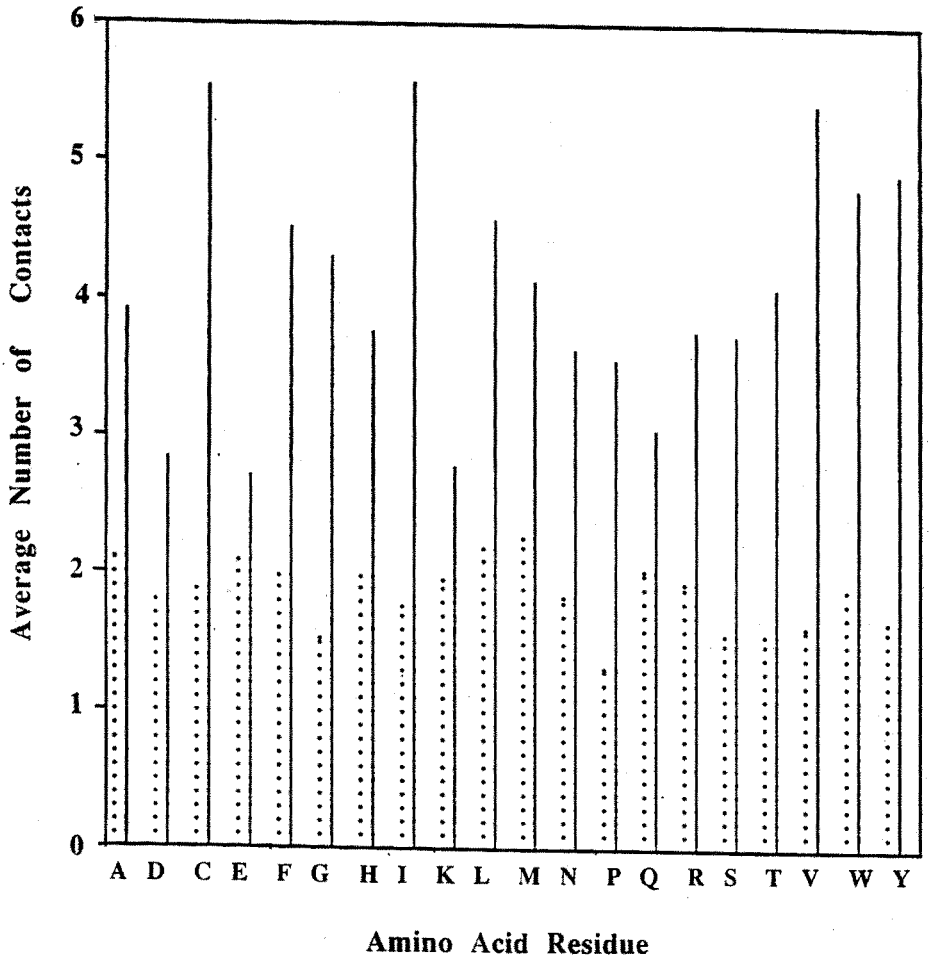

Figure 2. Average medium and long range contacts for the 20 amino acid residues in globular proteins.
..... medium range contacts; —— long range contacts.

Comparison Between Medium and Long Range Contacts. In Table III, the average medium and long range contacts per residue and the ratio between these contacts are presented. In Figure 2, the average medium and long range contacts per residue for the 20 types of amino acid residues are displayed. A perusal of the figure reveals that for all the residues long range contacts are higher than the medium range contacts. However, apart from the residue Cys which is involved in disulfide bonding, the hydrophobic residues Val, Ile and Tyr have relatively higher long range contacts (nearly three times higher than the medium range contacts; Table III). It suggests that these residues can serve as nucleation centres during the process of folding and get buried in the interior. The residues Glu and Lys contribute nearly

equally for medium and long range contacts (although all the charged residues have a similar behaviour).

For now, we have applied this information to predict interactions and inter-residue contacts in crystal structures of proteins.

4. Conclusions

The local side chain interactions are not sufficient for predicting the secondary structures of proteins. The present study reveals the importance of medium range interaction in the formation of α helices and the importance of long range interactions in the case of β strands. The helix forming residues Met, Leu, Ala and Glu have more medium range contacts in the tertiary structure of proteins. The residues in the α class of proteins have higher medium range contacts whereas in β class of proteins have higher long range contacts. The behaviour of Cys residues have been explicitly known from the higher contacts in long range level due to the formation of disulfide bridges. The hydrophobic residues Ile, Cys, Val and Met in all-α proteins have a larger number of long range contacts, as well in other classes, showing their strong effect in the formation of a hydrophobic core in the interior of globular proteins with residues far apart at the level of sequence.

Acknowledgments

We thank Professor P.K. Ponnuswamy for his helpful discussions and Shubha Anand for critical reading of the manuscript. MMG is a recipient of an ICGEB fellowship.

References

1. Nagano, K. and Ponnuswamy, P.K.: *Adv. Biophys.* **18** (1984), 115–148.
2. Barlow, D.J. and Thornton, J.M.: *Biopolymers* **25** (1986), 1717–1733.
3. MacArthur, M.W. and Thornton, J.M.: *J. Mol. Biol.* **218** (1991), 397–412.
4. Karpeisky, M.Ya and Ilyin, V.A.: *J. Mol. Biol.* **224** (1992), 629–638.
5. Zhu, B-Y., Zhou, N.E., Kay, C.M. and Hodges, R.S.: *Protein Sci.* **2** (1993), 383–394.
6. Vtyurin, N.: *Proteins* **15** (1993), 62–70.
7. Ponnuswamy, P.K. and Gromiha, M.M.: *J. Theor. Biol.* **166** (1994), 63–74.
8. Gobel, U., Sander, C., Schneider, R. and Valencia, A.: *Proteins* **18** (1994), 309–317.
9. Tanaka, S. and Scheraga, H.A.: *Macromolecules* **9** (1976), 945–950.
10. Manavalan, P. and Ponnuswamy, P.K.: *Arch. Biochem. Biophys.* **184** (1977), 476–487.
11. Warme, P.K. and Morgan, R.S.: *J. Mol. Biol.* **118** (1978), 289–304.
12. Narayana, S.V.L. and Argos, P.: *Int. J. Peptide Protein Res.* **24** (1984), 25–39.
13. Viswanathan, V.N.: *Int. J. Biol. Macromol.* **9** (1987), 39–48.
14. Miyazawa, S. and Jernigan, R.L.: *Macromolecules* **18** (1985), 534–552.
15. Burlety, S.K. and Petsko, G.A.: *Science* **229** (1985), 23–29.
16. Burtley, S.K. and Petsko, G.A.: *Adv. Protein Chem.* **39** (1988), 125–189.
17. Heringa, J. and Argos, P.: *J. Mol. Biol.* **220** (1991), 151–171.
18. Magalhaes, A., Margret, B., Hoflack, J., Gomes, J.N. and Scheraga, H.A.: *J. Protein Chem.* **13** (1994), 195–215.
19. Muthusamy, R. and Ponnuswamy, P.K. *J. Theor. Biol.* **153** (1991), 25–40.

20. Karlin, S., Zucker, M. and Brocchieri, L.: *J. Mol. Biol*. **239** (1994), 227–248.
21. Tanaka, S. and Scheraga, H.A.: *Proc. Natl. Acad. Sci*. **72** (1975), 3802–3806.
22. Ponnuswamy, P.K., Warme, P.K. and Scheraga, H.A.: *Proc. Natl. Acad. Sci. USA* **70** (1973), 830–833.
23. Allewell, N.: *Trends. Biochem. Sci*. **16** (1991), 239–240.
24. Tudos, E., Fiser, A. and Simon, I.: *Int. J. Peptide Protein Res.* **43** (1994), 205–208.
25. Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F. Jr, Brice, M.O., Rodgers, J.K., Kennard, O., Shimanouchi, T. and Tasumi, M.: *J. Mol. Biol*. **112** (1977), 535–542.
26. Manavalan, P. and Ponnuswamy, P.K.: *Nature* **275** (1978), 673–674.
27. Ponnuswamy, P.K., Prabakaran, M. and Manavalan, P. *Biochim. Biophys. Acta* **623** (1980), 301–316.
28. Shin, Y.A. and Yoo, S.E.: *Biopolymers* **38** (1996), 183–190.
29. Fasman, G.D. In *Prediction of Protein Structure and Principles of Protein Conformation* (G.D. Fasman, ed.), pp. 193–316. New York: Plenum Press, 1989.
30. Gromiha, M.M. and Ponnuswamy, P.K.: *Int. J. Peptide Protein Res.* **45** (1995), 225–240.
31. Levin, J.M. and Garnier, J. *Biochim. Biophys. Acta* **955** (1988), 283–295.
32. Rost, B., Schneider, R. and Sander, C.: *Trends. Biochem. Sci*. **18** (1993), 120–123.
33. Thornton, J.M.: *J. Mol. Biol*. **151** (1981), 261–287.
34. Matthews, B.W.: *Ann. Rev. Biochem.* **62** (1993), 139–160.
35. Barlow, D.J. and Thornton, J.M.: *J. Mol. Biol.* **168** (1983), 867–885.

Address for correspondence: Dr M. Michael Gromiha, 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