Ion pairs in non-redundant protein structures

B A Gowri Shankar[1], R Sarani[1], Daliah Michael[1], P Mridula[1], C Vasuki Ranjani[1], G Sowmiya[1], B Vasundhar[1], P Sudha[1], J Jeyakanthan[1], D Velmurugan[1] and K Sekar[1,2,*]

[1]*Bioinformatics Centre (Centre of Excellence in Structural Biology and Bio-computing)*

[2]*Supercomputer Education and Research Centre, Indian Institute of Science, Bangalore 560 012, India*

**Corresponding author (Fax, 91-80-23600683; Email, sekar@physics.iisc.ernet.in)*

Ion pairs contribute to several functions including the activity of catalytic triads, fusion of viral membranes, stability in thermophilic proteins and solvent–protein interactions. Furthermore, they have the ability to affect the stability of protein structures and are also a part of the forces that act to hold monomers together. This paper deals with the possible ion pair combinations and networks in 25% and 90% non-redundant protein chains. Different types of ion pairs present in various secondary structural elements are analysed. The ion pairs existing between different subunits of multisubunit protein structures are also computed and the results of various analyses are presented in detail. The protein structures used in the analysis are solved using X-ray crystallography, whose resolution is better than or equal to 1.5 Å and R-factor better than or equal to 20%. This study can, therefore, be useful for analyses of many protein functions. It also provides insights into the better understanding of the architecture of protein structure.

[Gowri Shankar B A, Sarani R, Michael D, Mridula P, Ranjani C V, Sowmiya G, Vasundhar B, Sudha P, Jeyakanthan J, Velmurugan D and Sekar K 2007 Ion pairs in non-redundant protein structures; *J. Biosci*. **32** 693–704]

1. Introduction

In proteins, ion pairs are electrostatic interactions between the nitrogen atoms of basic residues and the carboxylate oxygen atoms of acidic residues. The basic residues include histidine, arginine and lysine, and the acidic residues are aspartate and glutamate. In the case of basic residues, the ND1 and NE2 atoms of hisitidine, NH1 and NH2 atoms of arginine and NZ atom of lysine are involved in the formation of ion pairs. In acidic residues, the OD1 and OD2 atoms of aspartate and OE1 and OE2 of glutamate take part. Ion pairs play important roles in protein structure and function such as oligomerization, molecular recognition, allosteric regulation, domain motions and α-helix capping (Perutz 1970; Fersht 1972; Barlow and Thornton 1983; Musafia *et al* 1995; Xu *et al* 1997a, b; Kumar *et al* 2000). Most importantly, ion pairs play a vital role in the stabilization of secondary structures such as in case of α-helical 'rod' domains of intermediatory filament proteins (Letai and Fuchs 1995). Different ion pairs are possible due to various residue-wise combinations of acidic and basic amino acids. However, when the atoms involved in ion pairs are considered, two classes of ion pairs (complete and incomplete) are possible (figures 1 and 2). In complete ion pairs, either both the atoms of both residues participate in the ion pair or there could be two atoms from the acidic residue and one atom from the basic residue or vice versa. In incomplete ion pairs, only one atom from each residue is involved in the ion pair. Two atoms from different acidic residues and two atoms of the same basic residue or vice versa, form networks of ion pairs (figures 3 and 4).

As has been reported in the literature (Barril *et al* 1998; Hendsch and Tidor 1994; Persikov *et al* 2005), several groups have addressed various issues related to ion pairs such as electrostatic and energetic contributions of solvent-exposed ion pairs, electrostatic strengths in thermophilic and mesophilic enzymes, and their involvement in thermal stability. However, there is no information available in the literature on the percentage of various possible residues involved in ion pair formation and the nature of ion pairs found in protein structures. Thus, the present study, which has been developed based on the pioneering work of Barlow

Keywords. Charged amino acid residues; electrostatic interactions; protein structures; salt bridges; secondary structures

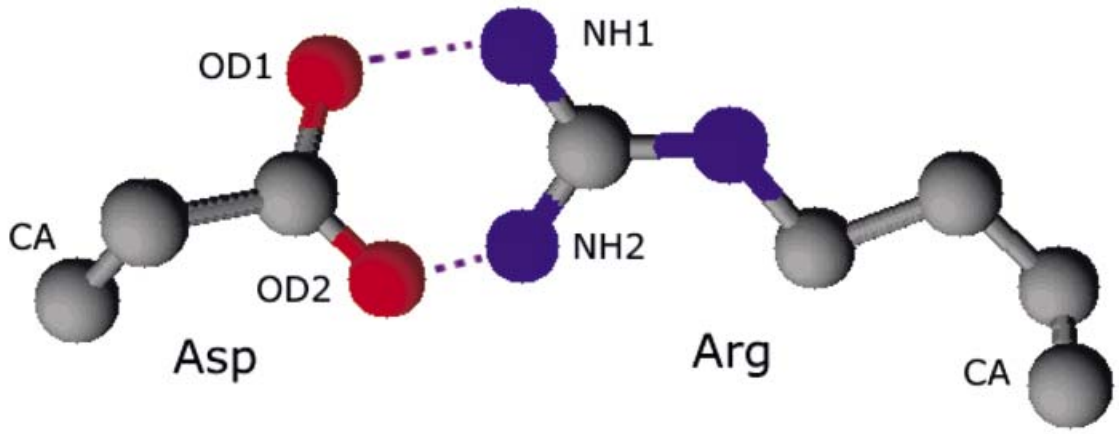

Figure 1. The figure depicts a complete ion pair between the residues aspartate and arginine.

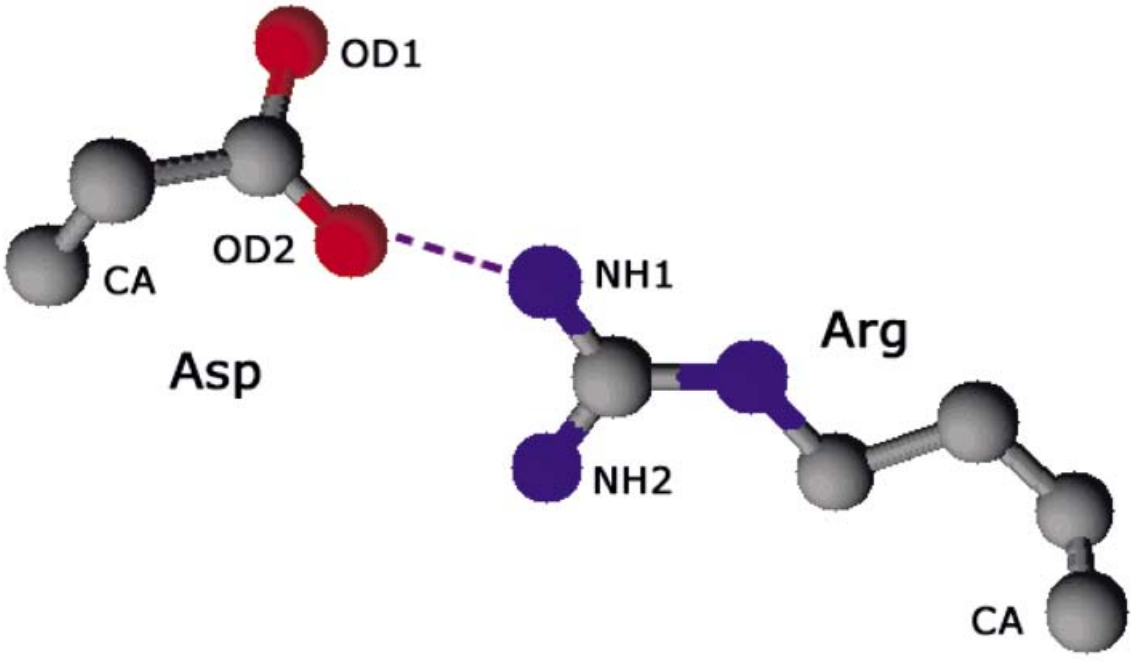

Figure 2. A schematic representation of arginine and aspartate forming an incomplete ion pair.

and Thornton (1983), aims to analyse the various kinds of ion pairs and their distribution in different secondary structures. Further, we study the charged networks formed by them and their role in the interactions between subunits. Furthermore, compared to the previous studies, the analysis of this study is carried out on a larger dataset consisting of 1227 highly resolved protein chains solved using X-ray crystallography available in non-redundant (25% and 90%) databases. The results are presented in the subsequent sections.

2. Materials and methods

2.1 *Database composition*

The non-redundant protein chains used in the present study were chosen from 25% and 90% non-redundant databases (Hobohm and Sander 1994), consisting of 2485 and 8595 protein chains, respectively. Only those X-ray analysed crystal structures whose resolution was better than or equal

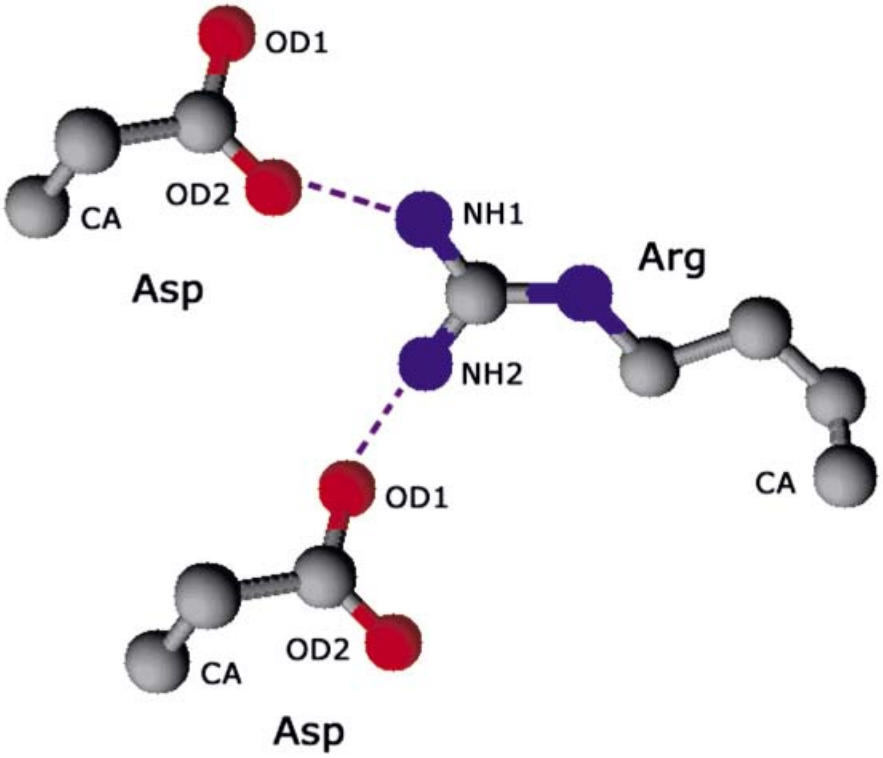

Figure 3. A network of ion pairs formed by two aspartate residues and an arginine residue.

to 1.5 Å and R-factor better than or equal to 20% were selected and included in the calculation. After applying these quality cut-off criteria, the dataset reduced to 419 and 808 chains in 25% and 90% datasets, respectively. Out of 419 chains in the 25% dataset, 11 protein chains have two subunits. In the case of the 90% dataset, 36 protein chains have two subunits, 2 protein chains have three subunits and one protein chain has four subunits.

The three-dimensional atomic coordinates of the corresponding protein chains were downloaded from the locally maintained Protein Data Bank FTP anonymous server (Bioinformatics Centre [Centre of Excellence in Structural biology and Bio-computing], Indian Institute of Science, Bangalore, India). The necessary Perl scripts were developed to perform the required computations.

2.2 *Methods*

Ion pairs are divided into four geometrical categories; salt bridges, N–O bridges, C–C bridges and long-range ion pairs (Barlow and Thornton 1983; Kumar and Nussinov 2002a, b) and our focus in the present report is on salt bridges. An ion pair is classified as a salt bridge when the centroids of the side-chain atoms of the charged residues are within a distance of 2.5 Å and 4.0 Å, and at least one pair of aspartate/glutamate side-chain carboxylate oxygen atoms and the side-chain nitrogen atoms of arginine/lysine/histidine is within a distance of 4.0 Å (Kumar and Nussinov 1999, 2000, 2001, 2002a). The three-dimensional atomic coordinates of the charged atoms participating in ion pairs were extracted from their respective PDB files. From the distances, ion pairs within the range 2.5–4.0 Å were selected and classified based on the nature of atoms involved in the interaction. Of the observed ion pairs, the percentage of individual basic and acidic residues involved in ion pair formation was also calculated and these residues were classified based on their geometrical orientations. The knowledgebase SSEP (Balamurugan *et al* 2005; Shanthi *et al* 2003) was deployed to get the secondary structural elements and the residues involved in ion pairs were analysed.

3. Results and Discussion

3.1 *General*

Of the 419 protein chains in the 25% dataset and 808 chains in the 90% dataset (table 1), only 405 and 779 chains, respectively, contain ion pairs. A total of 12% of the charged residues (both basic and acidic) are present and the details

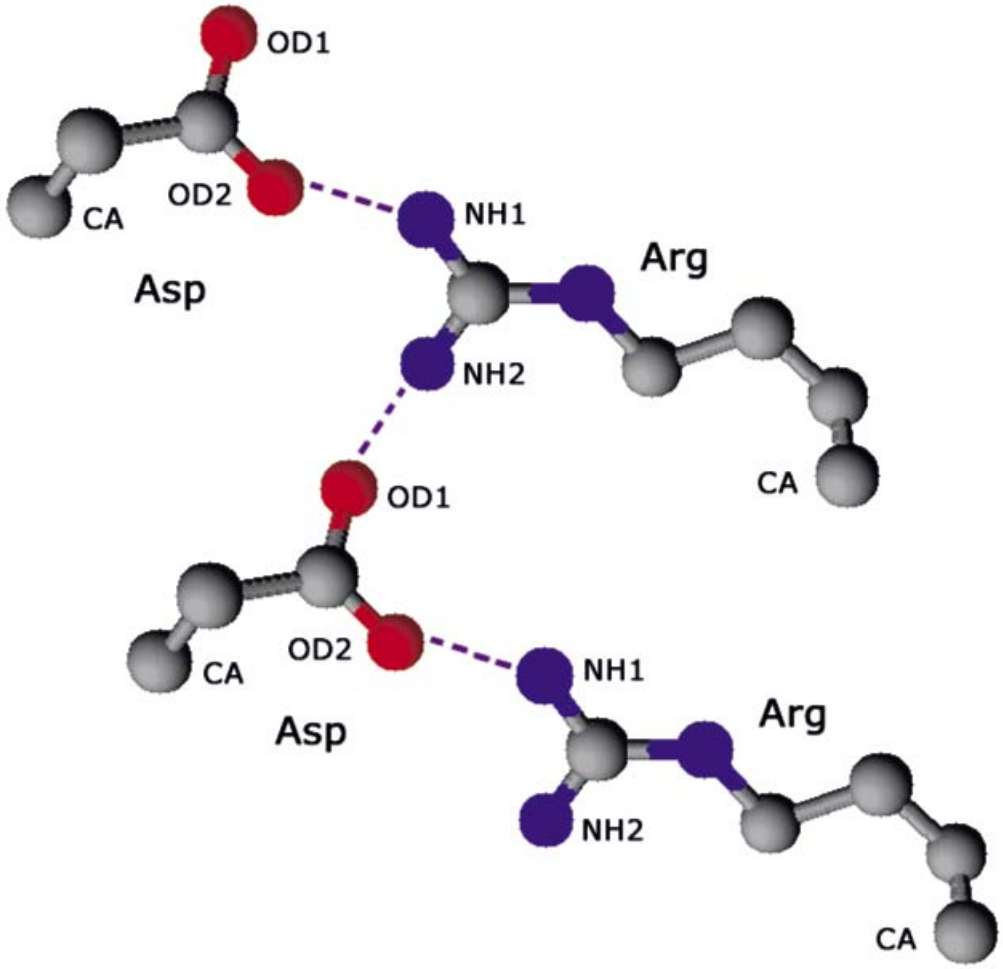

Figure 4. An extended network of ion pairs formed by arginine and aspartate resdidues.

Table 1. Number of charged residues present in non-redundant proteins (25% and 90%) and their involvement in ion pairs

Types of residues	Number of charged residues *	Number of residues involved in ion pairs *
25% (lys/arg/his)	10555 (12.67)	3647 (34.55)
25% (asp/glu)	10020 (12.03)	4215 (42.06)
90% (lys/arg/his)	21892 (11.87)	7621 (34.81)
90% (asp/glu)	20765 (12.52)	8770 (42.23)

* Number within parentheses denotes percentage.

are given in table 1. However, 35% of the basic residues and 42% of the acidic residues are involved in ion pairs.

Thus, the percentage of negatively charged residues that form ion pairs is higher than the corresponding percentage of positively charged residues. Further, the number of glutamate/aspartate in ion pairs is more than the number of lysine/arginine/histidine even though the total number of ion pairs formed by the negatively charged residues is equal to the total number of ion pairs formed by the positive residues. Since the number of positive residues is greater than the number of negative residues found in proteins in both the 25% and the 90% datasets (table 1), this would lead to the argument that the negatively charged residues tend to form networks or the positive residues form a greater proportion of complete ion pairs. Table 2 shows the number of residues involved in ion pairs and it is seen that among the charged residues, histidine occurs in the lowest proportion (figure 5). Note that the acidic residues occur in a higher proportion as compared to the basic residues and among the basic residues, lysine has the highest proportion (figure 5).

3.2 *Types of ion pairs*

3.2.1 *Within the subunit:* In protein structures with only one polypeptide chain, 60 different types of ion pairs are

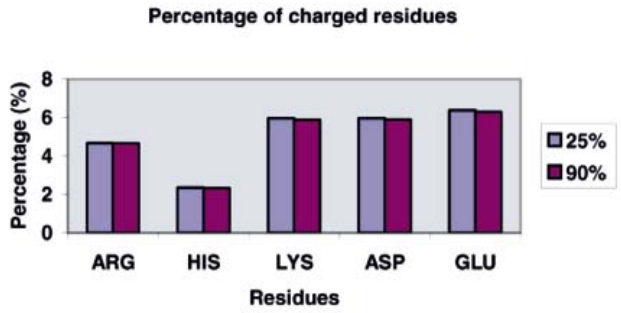

Figure 5. Percentage distribution of different charged residues in 419 and 808, the chains of 25% and 90% non-redundant protein structures, respectively.

Table 2. Number of different charged residues present in proteins and their involvement in ion pairs

Residues	Number of residues in non-redundant proteins		Number of residues involved in ion pairs	
	25%	90%	25%	90%
His	1953	4111	586	1226
Lys	4673	9535	1225	2577
Arg	3929	8246	1836	3818
Asp	4984	10311	2062	4285
Glu	5036	10454	2153	4485

observed (tables 3 and 4) and they are further classified into two broad classes, namely, complete and incomplete ion pairs. A total of 40 different types (table 3) of complete ion pairs and 20 different types of incomplete ion pairs (table 4) are observed. As shown in figure 1, in complete ion pairs, either three (NZ of lysine and OD1, OD2 of aspartate/OE1, OE2 of glutamate) or all the four charged atoms (NH1, NH2 of arginine/ND1, NE2 of histidine and OD1, OD2 of aspartate/OE1, OE2 of glutamate) participate in the formation of ion pairs. In incomplete ion pairs, only two atoms (for example, NH1 of arginine and OD1 of aspartate) are involved, as is shown in figure 2. In the 25% dataset, incomplete ion pairs account for 41.25% of the total ion pairs observed and the analogous number in the 90% dataset is 41.85%. The corresponding values for complete ion pairs are 58.75% and 58.15%, respectively. In both classes, ion pairs between arginine and aspartate/glutamate are the most frequent and they form 53.08% and 53.24% of the observed ion pairs in the 25% and 90% datasets, respectively. Between lysine and aspartate/glutamate, the percentage of ion pairs is 32.44% and 31.96% in the 25% and 90% datasets, respectively, and the corresponding values for histidine are 14.47% and 14.79%, respectively. The plausible reasons for the lower percentage contribution of lysine and histidine as compared to arginine may be due to the position of NZ in the side-chain of the lysine residue, the side-chain movement of the histidine residue or the conformations adopted by charged side-chain atoms.

Table 2 shows the number of each charged residue in the two datasets and the number of those residues that are found to actually participate in the ion pairs. It can be seen that, generally, the number of glutamate residues comprising ion pairs is slightly higher than the number of aspartate residues, just as the total number of glutamate residues in the datasets is marginally higher than the total number of aspartate residues. However, from table 7, it is apparent that the trend is reversed in case of histidine – glutamate and histidine – aspartate interactions (a total of 260:303 and 545: 639 ion pairs in the 25% and 90% datasets, respectively). The reversal may be because of the conservation of the histidine–aspartate pair in catalytic triads, where aspartate stabilizes the activated histidine as suggested earlier by Barlow and Thornton (1983).

3.2.2 *Between different subunits:* Of the 11 proteins with multiple subunits in the 25% dataset, only six proteins have ion pairs between the subunits. There are 24 pairs, of which 50% are between lysine and aspartate/glutamate and 75% of them are complete ion pairs. However, in the 90% dataset, 36 proteins have multiple subunits, of which only 29 form ion pairs between the subunits. A total of 103 ion pairs are found between the subunits, of which 39.8% are complete ion pairs and 40% of the ion pairs are between lysine and aspartate/ glutamate. It is evident that among the basic residues, lysine is most frequently involved in ion pair formation between subunits in both datasets (25% and 90%).

3.2.3 Analysis of the types of ion pairs: Complete ion pairs may be considered biologically significant because they have two interactions between the same (acidic and basic) residues that form the ion pair and hence this would be stronger compared to incomplete ion pairs. Thus, complete ion pairs are likely to contribute to the stability and the folding of the protein, apart from any other specific functions it may have. On comparing tables 3 and 4, it can be seen that there are more types of complete ion pairs. Further, the number of complete ion pairs is greater than the number of incomplete ion pairs. Thus, complete ion pairs are prefered over incomplete ion pairs (except in the case of ion pairs involving histidine).

In contrast, incomplete ion pairs are weaker compared with complete ion pairs and probably contribute to the local flexibility of the protein to some extent. The importance of the incomplete ion pair is its ability to form networks by interacting with a charged atom of another residue. Thus, it has implications for the stability of the protein. On the other hand, a complete ion pair cannot form networks, since all the charged atoms are involved in interactions and none are free to interact with another residue.

Table 3. Types of complete ion pairs, the nature of interactions and their corresponding number in 25% and 90% datasets

S. No.	Type	90% dataset		25% dataset	
		Within subunit	Between subunits	Within subunit	Between subunits
1	NH2 … OD1 NH2 … OD2	341	3	694	8
2	NH2 … OE1 NH2 … OE2	306	0	635	7
3	NH1 … OE1 NH1 … OE2	144	0	262	5
4	NE2 … OE1 NE2 … OE2	47	0	94	0
5	ND1 … OD2 NE2 … OD2	5	0	18	0
6	NH1 … OD1 NH1 … OD2	104	0	217	1
7	NH1 … OE1 NH2 … OE1	56	0	119	0
8	NZ … OE1 NZ … OE2	286	4	572	5
9	NZ … OD1 NZ … OD2	303	4	585	14
10	NH1 … OD1 NH2 … OD1	55	0	107	4
11	NH1 … OD2 NH2 … OD2	50	0	98	1
12	ND1 … OD1 ND1 … OD2	74	0	172	1
13	NH1 … OE2 NH2 … OE2	65	0	130	1
14	ND1 … OE1 ND1 … OE2	48	0	89	2
15	NE2 … OD1 NE2 … OD2	58	0	120	1
16	ND1 … OE1 NE2 … OE1	5	0	21	0
17	NH1 … OD1 NH2 … OD2	2	0	3	0
18	ND1 … OD1 NE2 … OD1	7	0	12	0
19	NH1 … OE2 NH2 … OE1	10	0	23	0
20	ND1 … OE1 NE2 … OE2	1	0	2	0
21	ND1 … OE2 NE2 … OE2	13	0	21	0
22	ND1 … OE2 NE2 … OE1	4	0	12	0
23	ND1 … OD1 NE2 … OD2	1	0	2	0

Table 3. (*Continued*)

S. No.	Type	90% dataset		25% dataset	
		Within subunit	Between subunits	Within subunit	Between subunits
24	NH1 … OE1 NH2 … OE2	4	0	5	1
25	NH1 … OD2 NH2 … OD1	18	0	32	0
26	ND1 … OD2 NE2 … OD1	3	0	5	0
27	NH1 … OD1 NH1 … OD2 NH2 … OD1	10	0	27	0
28	NH1 … OE1 NH1 … OE2 NH2 … OE2	17	0	26	0
29	NH1 … OE1 NH1 … OE2 NH2 … OE1	15	0	27	1
30	ND1 … OE1 ND1 … OE2 NE2 … OE1	2	0	5	0
31	NH1 … OD1 NH2 … OD1 NH2 … OD2	16	1	39	1
32	NH1 … OD1 NH1 … OD2 NH2 … OD2	12	0	22	1
33	NH1 … OD2 NH2 … OD1 NH2 … OD2	16	0	29	1
34	ND1 … OD1 NE2 … OD1 NE2 … OD2	1	0	4	0
35	NH1 … OE1 NH2 … OE1 NH2 … OE2	10	0	24	0
36	ND1 … OE1 NE2 … OE1 NE2 … OE2	1	0	4	0
37	ND1 … OE1 ND1 … OE2 NE2 … OE2	1	0	2	0
38	NH1 … OE2 NH2 … OE1 NH2 … OE2	7	0	17	0
39	NH1 … OD1 NH1 … OD2 NH2 … OD1 NH2 … OD2	88	1	179	2
40	NH1 … OE1 NH1 … OE2 NH2 … OE1 NH2 … OE2	64	1	160	5

Table 4. Types of incomplete ion pairs, the nature of interactions and their corresponding number in 25% and 90% datasets

S. No.	Type	25% dataset		90% dataset	
		Within subunit	Between subunits	Within subunit	Between subunits
1	NZ … OE1	176	3	363	5
2	NH1 … OD1	93	0	217	1
3	NZ … OD2	175	1	325	4
4	NH2 … OE1	125	0	206	3
5	NH1 … OE1	40	1	114	2
6	NH2 … OE2	103	0	228	5
7	NH2 … OD2	101	0	198	1
8	NH1 … OE2	52	0	109	2
9	ND1 … OE2	26	0	41	0
10	NE2 … OE2	48	0	118	0
11	NH1 … OD2	33	0	88	3
12	NZ … OE2	184	0	421	5
13	NH2 … OD1	90	0	198	0
14	NE2 … OD1	42	1	85	1
15	NE2 … OE1	41	2	88	4
16	NE2 … OD2	48	1	108	2
17	NZ … OD1	127	0	272	2
18	ND1 … OD1	32	0	53	0
19	ND1 … OD2	29	1	57	1
20	ND1 … OE2	21	0	42	0

3.3 *Ion pairs in networks (charged networks)*

Networks of ion pairs are defined as interactions between all the interacting atoms of a single basic residue (one in the case of lysine and two in the case of histidine and arginine) and atoms of two different acidic residues (Barlow and Thornton 1983) or vice versa (figure 3). Most of the charged residues participate in more than one ion pair, forming networks of ion pairs (figure 3). The network can extend to a large number of residues; a representation is shown in figure 4. The present study reveals 42 different types of networks (table 5), of which only 32 types occur in the 25% dataset and all 42 types occur in the 90% dataset. In both the 25% and 90% datasets, the networks between atoms of same arginine residue and different aspartate/glutamate residues are prominently seen. They form 59.6% and 49.18% of the total networks of ion pairs observed in the 25% and 90% datasets, respectively. In contrast, lysine and aspartate/glutamate constitute 16.55% and 32.24% in the 25% and 90% datasets, respectively, and networks of ion pairs are not seen between subunits in the 25% dataset. However, three types of networks [NH2 (248:G, 445:G) and OD1 (44:Y), OD2 (44:Y) of protein with PDB-id: 1uky, NH1 (386:B), NZ (386:B) and OD2 (463:A), OD1 (463:A) of protein with PDB-id: 1hbn, and NZ (386:B) and OD1 (463:A), OD2 (463:A) of protein with PDB-id: 1hbn] are observed in the 90% dataset. The data in parentheses denote residue number and chain ID. Since the network of ion pairs seen on the surface of the protein comprise about 70% of intersubunit interactions (Musafia *et al* 1995), they are likely to contribute to the protein–protein interactions or in the stabilization of monomers in multi-subunit protein structures. The distribution of charges in the hydrophobic core of the protein by the formation of ion pairs may also affect the stability of the protein.

3.4 *Ion pairs in secondary structures*

3.4.1 *Within the subunit:* As has been stated earlier, most of the ion pair forming residues are present in the secondary structural elements (Persikov *et al* 2005; Lyu *et al* 1992; Keskin *et al* 2005; Marqusee and Sauer 1994) and the number of residues involved are listed in table 6. Of all the secondary structures, α-helical regions contain most of the ion pair forming residues. A total of 50.89% of the basic and 48.26% of the acidic residues that form ion pairs occur in the α-helical regions. In the case of the 90% dataset, the corresponding values are 50.59% and 46.09% respectively. The analysis implies that most of the ion pairs are involved

Table 5. Ion pair networks observed in non-redundant protein chains. The characters, a, b and c in parentheses refer to different residues involved in the formation of ion pair networks

S. No.	Type	25% single subunit	90% single subunit
1	NZ (a) … OE1 (c) NZ (b) … OE2 (c)	3	5
2	NZ (a) … OE1 (c) NH1 (b) … OE2 (c)	3	6
3	NH1 (a) … OE1 (c) NZ (b) … OE2 (c)	0	1
4	NH2 (a) … OE1 (c) NZ (b) … OE2 (c)	1	2
5	NZ (a) … OE1 (c) NH2 (b) … OE2 (c)	3	4
6	NZ (a) … OE1 (c) ND1 (b) … OE2 (c)	1	1
7	NZ (a) … OE1 (c) NE2 (b) … OE2 (c)	0	2
8	NE2 (a) … OE1 (c) NZ (b) … OE2 (c)	0	1
9	NZ (a) … OD1 (c) NE2 (b) … OD2 (c)	1	1
10	NH1 (a) … OE1 (c) NE2 (b) … OE2 (c)	1	1
11	NH2 (a) … OE1 (c) NE2 (b) … OE2 (c)	0	1
12	NH2 (a) … OD1 (c) ND1 (b) … OD2 (c)	1	4
13	NH2 (a) … OE1 (c) NE2 (b) … OE2 (c)	0	1
14	ND1 (c) … OD1 (a) NE2 (c) … OD2 (b)	1	2
15	NH2 (c) … OD1 (a) NH1 (c) … OD2 (b)	2	7
16	NH1 (c) … OE1 (a) NH2 (c) … OE2 (b)	19	19
17	ND1 (c) … OE1 (a) NE2 (c) … OE2 (b)	3	5
18	NH1 (c) … OE1 (a) NH1 (c) … OE2 (b)	5	5
19	NH2 (c) … OE1 (a) NH2 (c) … OE2 (b)	6	18
20	ND1 (c) … OE1 (a) ND1 (c) … OE2 (b)	1	1
21	NE2 (c) … OE1 (a) NE2 (c) … OE2 (b)	4	5
22	NE2 (c) … OD1 (a) NE2 (c) … OD2 (b)	1	1
23	ND1 (c) … OD1 (a) ND1 (c) … OD2 (b)	0	1
24	NH2 (c) … OD1 (a) NH2 (c) … OD2 (b)	5	10
25	NH1 (c) … OD1 (a) NH1 (c) … OD2 (b)	2	8
26	NH1 (a) … OD1 (c) NH1 (b) … OD2 (c)	0	1
27	NH2 (a) … OD1 (c) NH2 (b) … OD2 (c)	3	5
28	NH2 (a) … OE1 (c) NH2 (b) … OE2 (c)	1	1
29	NH1 (a) … OE1 (c) NH1 (b) … OE2 (c)	1	1
30	ND1 (a) … OE1 (c) NE2 (b) … OE2 (c)	0	1
31	NH1 (c) … OE1 (a) NH2 (c) … OE1 (b)	13	29
32	ND1 (c) … OE1 (a) NE2 (c) … OE1 (b)	1	3
33	ND1 (c) … OE2 (a) NE2 (c) … OE2 (b)	3	7
34	NH1 (c) … OE2 (a) NH2 (c) … OE2 (b)	8	31
35	NH1 (c) … OD2 (a) NH2 (c) … OD2 (b)	11	19
36	NH1 (c) … OD2 (a) NH2 (c) … OD2 (b)	7	9
37	ND1 (c) … OD1 (a) NE2 (c) … OD1 (b)	3	4
38	NH1 (c) … OD1 (a) NH2 (c) … OD1 (b)	12	25
39	NZ (c) … OE1 (a) NZ (c) … OE1 (b)	2	39
40	NZ (c) … OD1 (a) NZ (c) … OD1 (b)	7	20
41	NZ (c) … OD2 (a) NZ (c) … OD2 (b)	8	34
42	NZ (c) … OE2 (a) NZ (c) … OE2 (b)	8	25

in stabilization of the α-helix and thus help in anchoring the secondary structural elements in tertiary structures. In the case of β-turns, 22.48% and 23.03% ion pair forming residues are present in the 25% and 90% datasets, respectively. Here, the β-turns that are formed between the sheets are stabilized by the ion pairs. In β-sheets, the corresponding values are 14.38% and 13.15%, which is comparatively lower than that of α-helix and β-turns.

The plausible reason behind the lower percentage of ion pairs in β-sheets is due to the existence of the involvement

Table 6. Number of positively and negatively charged residues forming ion pairs, observed (within and between subunits) in different secondary structural elements

Type of secondary structure	25% dataset (same subunit)		25% dataset (different subunits)		90% dataset (same subunit)		90% dataset (different subunits)	
	A	B	A	B	A	B	A	B
α-helix	908	1028	12	11	1485	1617	47	45
3_{10} helix	55	67	0	0	85	110	1	1
Sheet type 1	207	203	2	3	322	304	3	5
Sheet type 2	79	74	2	0	140	120	5	2
γ-turn	23	44	0	0	29	71	2	1
hair pin loop	160	186	1	3	303	373	6	2
β-turn i	128	199	3	4	214	340	5	7
β-turn i1	11	22	0	1	18	45	0	1
β-turn ii	40	44	0	1	60	67	0	1
β-turn ii1	6	14	0	0	9	27	0	0
β-turn iv	132	195	0	2	207	346	3	7
β-turn via	3	3	0	0	2	4	0	0
β-turn vib	2	1	0	0	1	2	0	0
β-turn vii	30	50	1	0	60	82	2	0

A denotes positively charged residues and B denotes negatively charged residues.

Table 7. Number of ion pairs formed by the charged residues (within the subunit and between subunits)

Residue	ASP				GLU			
	25% dataset		90% dataset		25% dataset		90% dataset	
	Same subunit	Different subunits	Same subunit	Different subunits	Same subunit	Different subunits	Same subunit	Different subunits
LYS	605	5	1183	21	646	7	1356	16
ARG	1029	5	2145	25	1018	2	2085	32
HIS	300	3	636	6	258	2	539	6

of inter β-strand hydrogen bonds in a β-sheet. However, in a helix, bonds are formed by proximal residues and thus it is likely that helices contain intrahelix ion pairs and β-sheets have interstrand ion pairs.

3.4.2 Between different subunits: In multisubunit protein chains available in the 25% dataset, the number of basic and acidic residues (in different secondary structures) forming ion pairs is 21 and 25, respectively. Out of these, 57% of basic residues and 44% of acidic residues are present in the α-helical regions. In the 90% dataset, 74 and 72 basic and acidic residues, respectively, involved in ion pairs are found in different types of secondary structural elements. Among them, 63.5% of basic residues and 62.5% of acidic residues are present in the α-helical regions and the corresponding numbers are 13.5% and 22.2% in the β-turns. In contrast, a very low percentage is observed in the case of β-sheets.

In agreement with earlier studies (Mitchell *et al* 1992), the present study showed that arginine and aspartate/glutamate form most of the ion pairs that constitute 30 different types, whereas lysine and aspartate/glutamate form only six different types. It is interesting to note that the latter has a higher frequency of occurrence compared to the former, though the residues lysine and arginine take part in equal proportions in ion pair formation between subunits in multisubunit proteins (table 7). Because of the position of the charged atoms and the side-chain movement of histidine, the ion pairs formed between histidine and the acidic residues are much lower compared to those formed by the extended side-chain residues arginine and lysine.

3.4.3 Analysis of ion pairs in secondary structures: From tables 1 and 6, it is clear that less than half the residues forming ion pairs have been found in a regular secondary

structure implying that the remaining half of the residues can be a part of random coils. Further, ion pairs are found in disparate secondary structural elements. This corresponds to the results of Barlow and Thornton (1983) and confirms that ion pairs have a greater effect in anchoring the tertiary structure than the secondary structure. It is evident from table 6 that a simple majority of those residues comprising an ion pair are found in α-helices (approximately 50% and 66% in the 25% and 90% datasets, respectively). This suggests that the ion pairs are involved in intra- and interhelical interactions which could act as stabilizing or de-stabilizing forces. Intrahelical interactions would affect the formation and stability of the α-helix, while the monomeric structure of the protein would be influenced more by the interhelical interaction. The second highest occurence of ion pairs is in the β-turns (25% and 20% of the residues in the 25% and 90% datasets, respectively). The role of these ion pairs might be to stabilize the β-turns that are found between the β-sheets.

The ion pairs found in β-sheets are the next highest in number and usually form ion pairs between the strands of the sheet. The ion pairs may also be found between different β-sheets. In either case, the residues may be distal in sequence, which is not the case with intrahelical interactions. γ-turns normally comprise very few residues. Since these residues are more likely to be proline than a charged residue, not many ion pairs are formed in γ-turns.

Subunit interactions: It is evident from table 7 that the distribution of ion pairs found in intersubunits is scarce compared to those found in intrasubunits, being only 24 in 11 multimeric proteins of the 25% dataset and 29 in 36 mutlimeric proteins of the 90% dataset. These interactions are probably present in hydrophilic pockets or solvent-accessible surfaces rather than pockets where hydrophobic interactions dominate. Since hydrophobic interactions are the primary interactions holding the subunits together, ion pairs serve specific roles by forming electrostatic interactions (Kumar and Nussinov 2002b).

Although several α-helices and β-turns do not have ion pairs, they affect the stability of the secondary structural elements to a degree. In most cases, these ion pairs have a biological significance, and are thus seen in the active sites or structurally important regions of the protein molecule.

3.5 *Implications of ion pairs*

Ion pairs have the following specific functions, for example, ion pairs significantly affect thermostability. This is evident from the conservation of functionally important domains of psychrophilic proteins and disruption of all other ion pairs when compared with mesophilic proteins (Papaleo *et al* 2006). Histidine – aspartate is a functionally significant ion pair found in catalytic dyads or triads and is essential for efficient enzyme hydrolysis (Barlow and Thornton 1983). Furthermore, formation of ion pairs stabilizes viral membrane fusion in viral infection of cells (Kampmann *et al* 2006).

In all the above cases, the frequency and occurrence of ion pairs in the protein are central to the problem being addressed and thus the classification and analysis in the present study could be of use. Further, knowledge of ion pairs can be effectively used in the study of solvent–protein interactions, as in the case of high-density lipoproteins which are disrupted in a high salt environment due to the perturbation of ion pairs on the surface of the protein (Jayaraman *et al* 2006). In addition, ion pairs become prime candidates for interface-disrupting mutations when they are responsible for holding the monomers together in a protein (Brinda and Vishveshwara 2005). It is noteworthy that the results presented here can be used to understand the formation of ion pairs and to provide insights where the rational design of protein structures is attempted.

4. Conclusion

The present study on high-resolution non-redundant protein crystal structures revealed 60 types of ion pairs and different charged networks. Furthermore, most of the ion pairs present in secondary structures such as α-helix and β-turns confirm their participation in the stability of secondary structures to a great extent. Also, it can be seen from our analysis that different types of ion pairs have a role to play in maintaining the structural integrity of the protein. The ion pairs between subunits play an important role in holding the subunits together, thus helping in the formation of multisubunit protein structures. Lastly, the classification used in the present study makes it easier to understand the formation and function of various kinds of ion pairs.

Acknowledgements

The authors gratefully acknowledge the use of the Bioinformatics Centre (DIC), the Interactive Graphics Based Molecular Modeling facility (IGBMM) and the Supercomputer Education and Research Centre (SERC). The facilities DIC and IGBMM are supported by the Department of Biotechnology, Government of India. One of the authors (KS) thanks the Department of Biotechnology for financial support. The work is also supported by the National Bioscience award granted to Dr K Sekar. The authors thank the anonymous referees for their suggestions and comments.

References

Balamurugan B, Samaya M K, Ramesh J, Roshan M N A, Sumathi K and Sekar K 2005 SSEP - 2.0: secondary structural elements of proteins; *Acta Cryst.* **D61** 634–636

Barlow D J and Thornton J M 1983 Ion pairs in proteins; *J. Mol. Biol.* **168** 867–885

Barril X, Aleman C, Orozco M and Luque F J 1998 Salt bridge interactions: stability of the ionic and neutral complexes in the gas phase, in solution, and in proteins; *Proteins* **32** 67–79

Brinda K V and Vishveshwara S 2005 Oligomeric protein structure networks: insights into protein–protein interactions; *BMC Bioinformatics* **6** 296 01–15

Fersht A R 1972 Conformational equilibria in α- and δ-chymotripsin: the energetic and importance of the salt bridge; *J. Mol. Biol.* **64** 497–509

Hendsch Z S and Tidor B 1994 Do salt bridges stabilize proteins? A continuum electrostatic analysis; *Protein Sci.* **3** 211–226

Hobohm U and Sander C 1994 Enlarged representative set protein structures; *Protein. Sci*. **3** 522–524

Jayaraman S, Gantz D L and Gursky O 2006 Effects of salt on the thermal stability of human plasma high-density lipoprotein; *Biochemistry* **45** 4620–4628

Kampmann T, Mueller D S, Mark A E, Young P R and Kobe B 2006 The role of histidine residues in low-pH-mediated viral membrane fusion; *Structure* **14** 1481–1487

Keskin O, Ma B and Nussinov R 2005 Hot regions in protein–protein interactions: the organization and contribution of structurally conserved hot spot residues; *J. Mol. Biol.* **145** 1281–1294

Kumar S, Ma B, Tsai C J and Nussinov R 2000 Electrostatic strengths of salt bridges in thermophilic and mesophilic glutamate dehydrogenase monomers; *Proteins* **38** 368–383

Kumar S and Nussinov R 1999 Salt bridge stability in monomeric proteins; *J. Mol. Biol.* **293** 1241–1255

Kumar S and Nussinov R 2000 Fluctuations between stabilizing and destabilizing electrostatic contributions of ion pairs in conformers of the c-Myc-Max leucine zipper; *Proteins* **41** 485-497

Kumar S and Nussinov R 2001 Fluctuations in ion pairs and their stabilities in proteins; *Proteins* **43** 433–454

Kumar S and Nussinov R 2002a Relationship between ion pair geometries and electrostatic strengths in proteins; *Biophys. J.* **83** 1595–1612

Kumar S and Nussinov R 2002b Close-range electrostatic interactions in proteins; *Chem. Biol. Chem.* **3** 604–617

Letai A and Fuchs E 1995 The importance of intramolecular ion pairing in intermediate filaments; *Proc. Natl. Acad. Sci. USA* **92** 92–96

Lyu P C, Gans P J and Kallenbach N R 1992 Energetic contribution of solvent-exposed ion pairs to alpha-helix structure; *J. Mol. Biol.* **223** 343–350

Marqusee S and Sauer R T 1994 Contributions of a hydrogen bond/salt bridge network to the stability of secondary and tertiary structure in X repressor; *Protein Sci.* **3** 2217–2225

Mitchell J B, Thornton J M, Singh J and Price S L 1992 Towards an understanding of the arginine–apartate interaction; *J. Mol. Biol.* **226** 251–262

Musafia B, Buchner V and Arad D 1995 Complex salt bridges in proteins: statistical analysis of structure and function; *J. Mol. Biol.* **254** 761–770

Papaleo E, Olufsen M, De Gioia L, Brandsdal B O 2006 Optimization of electrostatics as a strategy for cold-adaptation: a case study of cold- and warm-active elastases; *J. Mol. Graph. Model*; Epub ahead of print

Persikov A V, Ramshaw J A M, Kirkpatrick A and Brodsky B 2005 Electrostatic interactions involving lysine make major contributions to collagen triple-helix stability; *Biochemistry* **44** 1414–1422

Perutz M F 1970 Stereochemistry of cooperative effects in haemoglobin; *Nature* **228** 726–739

Shanthi V, Selvarani P, Ch. Kiran Kumar, Mohire C S and Sekar K 2003 Secondary structural element of proteins; *Nucl. Acids Res.* **31** 3404–3405

Xu D, Lin S L and Nussinov R 1997a Protein binding versus protein folding: the role of hydrophobic bridges in protein associations; *J. Mol. Biol.* **205** 68–84

Xu D, Tsai C J and Nussinov R 1997b Hydrogen bonds and salt bridges across protein–protein interfaces; *Protein Eng*. **10** 999–1012

MS received 21 September 2006; accepted 6 February 2007

*e*Publication: 2 April 2007

Corresponding editor: Somdatta Sinha