Relationship between Ion Pair Geometries and Electrostatic Strengths in Proteins

Sandeep Kumar* and Ruth Nussinov*†

*Laboratory of Experimental and Computational Biology, National Cancer Institute, Frederick Cancer Research and Development Center, Frederick, Maryland 21702 USA; and †Sackler Institute of Molecular Medicine, Department of Human Genetics, Sackler School of Medicine, Tel Aviv University, Tel Aviv 69978, Israel

ABSTRACT The electrostatic free energy contribution of an ion pair in a protein depends on two factors, geometrical orientation of the side-chain charged groups with respect to each other and the structural context of the ion pair in the protein. Conformers in NMR ensembles enable studies of the relationship between geometry and electrostatic strengths of ion pairs, because the protein structural contexts are highly similar across different conformers. We have studied this relationship using a dataset of 22 unique ion pairs in 14 NMR conformer ensembles for 11 nonhomologous proteins. In different NMR conformers, the ion pairs are classified as salt bridges, nitrogen–oxygen (N–O) bridges and longer-range ion pairs on the basis of geometrical criteria. In salt bridges, centroids of the side-chain charged groups and at least a pair of side-chain nitrogen and oxygen atoms of the ion-pairing residues are within a 4 Å distance. In N–O bridges, at least a pair of the side-chain nitrogen and oxygen atoms of the ion-pairing residues are within 4 Å distance, but the distance between the side-chain charged group centroids is greater than 4 Å. In the longer-range ion pairs, the side-chain charged group centroids as well as the side-chain nitrogen and oxygen atoms are more than 4 Å apart. Continuum electrostatic calculations indicate that most of the ion pairs have stabilizing electrostatic contributions when their side-chain charged group centroids are within 5 Å distance. Hence, most (~92%) of the salt bridges and a majority (68%) of the N–O bridges are stabilizing. Most (~89%) of the destabilizing ion pairs are the longer-range ion pairs. In the NMR conformer ensembles, the electrostatic interaction between side-chain charged groups of the ion-pairing residues is the strongest for salt bridges, considerably weaker for N–O bridges, and the weakest for longer-range ion pairs. These results suggest empirical rules for stabilizing electrostatic interactions in proteins.

INTRODUCTION

Oppositely charged residue pairs often form ion pairs in proteins. Ion pairs play important roles in protein structure and function, such as in oligomerization, molecular recognition, domain motions, thermostability, and α-helix capping (Perutz, 1970; Fersht, 1972; Barlow and Thornton, 1983; Musafia et al., 1995; Xu et al., 1997a,b; Kombo et al., 2000; Kumar and Nussinov, 2001a,b; Kumar et al., 2000a,b, 2001a,b). Experimental and theoretical estimates of the electrostatic free energy contribution of ion pairs range from highly stabilizing to highly destabilizing (e.g., Warshel and Russell, 1984; Hwang and Warshel, 1988; Sham et al., 1998; Horovitz and Fersht, 1992; Marqusee and Sauer, 1994; Kumar and Nussinov, 1999; Singh, 1988; Barril et al., 1998; Dao-pin et al., 1991; Waldburger et al., 1995; Grimsley et al., 1999; Sheinerman et al., 2000). Continuum electrostatics-based calculations using computer mutations of the salt bridging residue side chains to their hydrophobic isosteres allow estimation of the electrostatic free energy contribution of an ion pair toward protein stability. Using this method, Hendsch and Tidor (1994) have studied 21 ion pairs from nine proteins. They found that most of the ion pairs are destabilizing. However, similar continuum electrostatic calculations have indicated that salt bridges in cytochrome P450cam are highly stabilizing (Lounnas and Wade, 1997). Xu et al. (1997a,b) have shown that hydrogen bonds and salt bridges play a more significant role in protein binding than in folding. The inter-subunit salt bridges in their analysis had stabilizing electrostatic contributions. Using a greater value for effective dielectric constant for the protein interior and a different method, Schutz and Warshel (2001) have shown that the ion pair His 31–Asp70 in T4 lysozyme is not destabilizing.

Kawamura et al. (1997) have shown that the disruption of the Glu34–Lys38 salt bridge in a DNA-binding protein, HU from *Bacillus stearothermophilus*, reduced its thermal stability. Salt bridges are more frequent in many proteins from thermophiles as compared with their mesophilic homologs (Kumar et al., 2000a; Sterner and Liebl, 2001; Kumar and Nussinov, 2001b). Computational studies have suggested that salt bridges and their networks contribute substantially to protein stability in thermophiles (Elcock, 1998; Xiao and Honig, 1999; de Bakker et al., 1999; Kumar et al., 2000b). Experimentally, surface salt bridges have been shown to stabilize proteins (Spek et al., 1998; Strop and Mayo, 2000).

Recently, we have computed the electrostatic free energy contributions for 222 salt bridges from 36 nonhomologous monomeric proteins (Kumar and Nussinov, 1999) using the continuum electrostatics methodology. Most (~86%) of the salt bridges in these proteins have stabilizing electrostatic free energy contributions, regardless of whether they are

Submitted September 14, 2001, and accepted for publication April 23, 2002.

Address reprint requests to Dr. Ruth Nussinov, SAIC Frederick, NCI-FCRDC, Bldg. 469, Rm. 151, Frederick, MD 21702. Tel.: 301-846-5579; Fax: 301-846-5598; E-mail: ruthn@ncifcrf.gov.

0006-3495/02/09/1595/18 $2.00

buried or solvent exposed, isolated or networked, or contained hydrogen bonds or not. The total electrostatic free energy contribution of a salt bridge depends upon the location of the salt-bridging residues in the protein and the interactions of the salt-bridging residue side-chain charged groups with each other and with other polar as well as ionized groups in the protein. That earlier study indicated that the geometrical orientation of the side-chain charged groups with respect to each other may be a crucial factor in determining the total electrostatic free energy contribution of the salt bridge.

Crystal structures present a static picture of a protein structure. This yields a single side-chain orientation for the salt-bridging residues, with the variability in the position of the side-chain charged group atoms inferred from B-factors. Hence, electrostatic calculations using protein crystal structures classify an ion pair as either stabilizing or destabilizing. The outcome of these calculations also depends on the methodology and the value of dielectric constant used for the protein interior as was recently reviewed by Schutz and Warshel (2001).

Recently, we have performed continuum electrostatic calculations using individual conformers in the NMR conformer ensembles of proteins (Kumar and Nussinov, 2000, 2001a). The calculated electrostatic free energy contribution for a pair of oppositely charged residues often fluctuates and interconverts between being stabilizing and destabilizing. These fluctuations reflect variations in atomic positions of the oppositely charged residue pairs in the protein conformers, even though it is difficult to separate the true protein mobility in solution from artifacts due to NMR data collection and structure calculation protocols (Kumar and Nussinov, 2000, 2001a). Fluctuations in ion pairs and their electrostatic strengths are also seen when different crystal structures for the same protein are compared (Kumar and Nussinov, 2001a). These observations illustrate the sensitivity of the total electrostatic free energy contribution of oppositely charged residue pair to the geometrical positioning of the side-chain charged groups.

Here, we examine the relationship between the geometrical orientation of the side-chain charged groups and the electrostatic free energy contribution of ion pairs. We refer to the total electrostatic free energy contribution by an ion pair to protein stability as the electrostatic strength of the ion pair. The electrostatic strength of an ion pair depends on the ion pair geometry and on the protein structural context. The ion pair geometry denotes the orientation of the side-chain charged groups in the ion pair with respect to each other. It can be characterized by computing the distance between the centroids of the ion-pairing residue side-chain charged groups and the angular orientation of these side-chain charged groups with respect to each other (Kumar and Nussinov, 1999, 2000, 2001a; see also Materials and Methods). The protein structural context of an ion pair refers to the location of the ion-pairing charged residues in the protein along with the identity, number, and spatial distribution of other polar and ionized groups in the protein. In different NMR conformers of a protein, the protein structural contexts are expected to be similar. This facilitates the study of the relationship between geometry and electrostatic strength of the ion pair because the effect due to the variations in the protein structural context on its electrostatic strength is minimized, though not completely removed.

We have studied 22 ion pairs in 14 NMR conformer ensembles of 11 nonhomologous proteins. A total of 1174 computations of ion pair geometry and electrostatic strengths have been performed. Most of these calculations are homologous as they pertain to the ensemble behavior of the 22 unique ion pairs. We classify each ion pair in each conformer as a salt bridge, nitrogen–oxygen (N–O) bridge, or a longer-range ion pair. Salt bridges have the best geometrical orientation of the side-chain charged groups with respect to each other. The side-chain charged group centroids and at least one pair of side-chain nitrogen and oxygen atoms are within 4 Å distance. The ion pair geometry in N–O bridges is worse than that in salt bridges but better than that in the longer-range ion pairs. The N–O bridges have at least a pair of side-chain nitrogen and oxygen atoms within 4 Å distance, but the centroids of their side-chain charged groups are more than 4 Å apart. In the longer-range ion pairs, the side-chain charged group centroids as well as the side-chain nitrogen and oxygen atoms are more than 4 Å apart. In NMR conformer ensembles, the electrostatic interactions among the ion-pairing residues are strongest when they form salt bridges, weaker in N–O bridges, and the weakest in longer-range ion pairs. Most salt bridges stabilize the protein structures. The majority of the N–O bridges are also stabilizing but with a smaller proportion. Most longer-range ion pairs are destabilizing. Most of the ion pairs with side-chain charged group centroids within 5 Å distance are stabilizing toward protein structures.

MATERIALS AND METHODS

Definition of ion pair and ion pair geometry

A pair of oppositely charged residues, Asp or Glu with Arg, His, or Lys, is considered an ion pair. The geometrical orientation of the side-chain charged groups in the ion-pairing residues with respect to each other is characterized in terms of 1) the distance (r) between the side-chain charged group centroids and 2) the angular orientation (θ) of the side-chain charged groups in the two ion-pairing residues. This is the angle between two unit vectors. Each unit vector joins a C^{α} atom and a side-chain charged group centroid in a charged residue. Fig. 1 presents a schematic diagram.

The coordinates for a side-chain charged group centroid are computed by taking the average of the coordinates of the heavy (nonhydrogen) atoms in the side-chain charged group. We have used only the heavy atoms because the protein crystal structures often lack coordinates for hydrogen atom positions. The present study uses protein structures solved both by x-ray crystallography and NMR spectroscopy. The following side-chain charged group atomic coordinates are used for computing the side-chain charged group centroid position: Asp, C^{γ}, $O^{\delta 1}$, $O^{\delta 2}$; Glu, C^{δ}, $O^{\epsilon 1}$, $O^{\epsilon 2}$; Arg, N^{ϵ}, C^{ξ}, $N^{\eta 1}$, $N^{\eta 2}$; His, C^{γ}, $N^{\delta 1}$, $C^{\delta 2}$, $C^{\epsilon 1}$, $N^{\epsilon 2}$; and Lys, N^{ξ}.

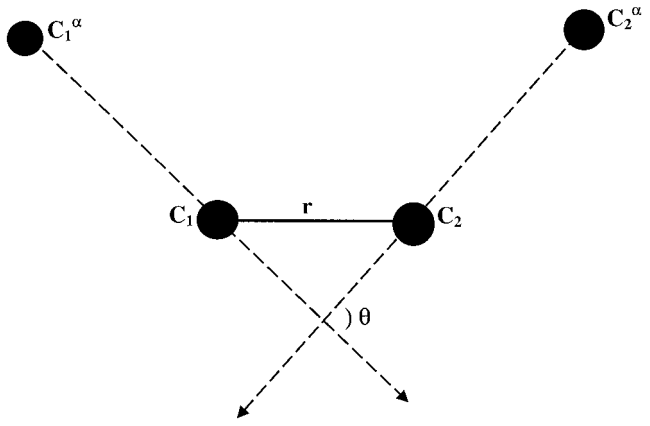

FIGURE 1 A schematic diagram characterizing geometrical orientation of the side-chain charged residues in an ion pair. Ion pair geometry can be measured in terms of r and θ. r is the distance in angstroms between the centroids, C1 and C2, of the side-chain charged groups in the ion-pairing residues 1 and 2. θ, in degrees, is the angle between two unit vectors $C1^{\alpha}C1$ and $C2^{\alpha}C2$. Note that we actually take the supplemental angle.

Geometrical criteria for various ion pair types

Ion pairs are divided into four geometrical categories, salt bridges, N–O bridges, C–C bridges, and longer-range ion pairs. An ion pair is classified as a salt bridge when 1) the side-chain charged group centroids are within a 4.0 Å distance and 2) at least one pair of Asp/Glu side-chain carbonyl oxygen and Arg/Lys/His side-chain nitrogen atoms are within a 4.0 Å distance (Kumar and Nussinov, 1999, 2000, 2001a). An ion pair is classified as a C–C bridge when it satisfies only the first criterion. The ion pair is a N–O bridge when it violates the first but satisfies the second criterion for salt bridge formation. In a longer-range ion pair both salt bridge criteria are violated.

Database composition

The ion pairs and NMR conformer ensembles used in the present study are shown in Table 1. Our database consists of 22 ion pairs in 14 NMR conformer ensembles for 11 nonhomologous monomeric proteins or single-protein domains that contain ≥50 amino acid residues. All NMR conformer ensembles but one contain ≥40 NMR conformers. The 22 ion pairs were selected because their oppositely charged residues form salt bridges in at least one crystal structure (if available) or in the NMR average energy-minimized structure or in the most representative conformer (Kelly et al., 1996) in the NMR conformer ensemble (if both the crystal structure and the NMR average energy-minimized structures are unavailable) of the proteins. The 11 nonhomologous proteins are β-spectrin pleckstrin homology (PH) domain (Nilges et al., 1997); CheY (Bellsolell et al., 1994; Moy et al., 1994; Volz and Matsumura, 1991); c-Myb DNA-binding domain repeat 1 (R1), repeat 2 (R2), and repeat 3 (R3) (Ogata et al., 1995); cysteine-rich intestinal protein (CRIP) (Perez-Alvarado et al., 1996); CSE-I (Jablonsky et al., 1999); cyanovirin-N (Bewley et al., 1998; Yang et al., 1999); horse heart cytochrome c (reduced form) (Banci et al., 1999); high mobility group 1 (HMG1) box 2 (Read et al., 1993); ISL-1 (Ippel et al., 1999); B1 domain of protein G (Gallagher et al., 1994; Gronenborn et al., 1991; Kuszewski et al., 1999); and U1 SNRP A (Avis et al., 1996). The 14 NMR conformer ensembles involved in this study are taken from the Protein Data Bank (PDB) (Bernstein et al., 1977) entries 1B3C, 1BW5, 1CEY, 1FHT, 1GB1, 1HSN, 1IML, 1MBF, 1MBH, 1MBK, 1 MPH, 2EZN, 2GIW, and 3GB1. For each repeat of c-Myb DNA-binding domain, a separate NMR conformer ensemble containing 50 conformers is available. Hence, there are three NMR conformer ensembles (1MBF, 1MBH, and 1MBK) for c-Myb DNA-binding domain. Two conformer ensembles (1GB1 and 3GB1) are available for the B1 domain of protein G. The details of the selection of protein NMR conformer ensembles for the purpose of our study have been described as well as the advantages and limitations of using NMR conformers for computations of electrostatic strengths of ion pairs (Kumar and Nussinov, 2000, 2001a). The criteria and rationale for selecting the individual ion pairs are summarized in Table 1.

Computation of electrostatic free energy contributions by ion pairs

Free energy contributions due to electrostatic interactions in proteins are frequently computed using continuum electrostatics methodologies. The solute (protein) is represented in atomic detail, but the solvent (water) is represented only in terms of its bulk properties. These methods have found several applications (Honig and Nicholls, 1995). The methodologies to model electrostatic effects in proteins have been recently reviewed by Warshel and Papazyan (1998).

The electrostatic strength of an ion pair is calculated relative to the hydrophobic isosteres of the side chains of the charged residues. The hydrophobic isosteres are the charged residue side chains with their partial atomic charges set to zero (Hendsch and Tidor, 1994). This method has been used extensively (Hendsch and Tidor, 1994; Xu et al., 1997b; Lounnas and Wade, 1997; Kumar and Nussinov, 1999, 2000, 2001a; Kumar et al., 2000b, 2001), and its predictions are consistent with experimental observations (Waldburger et al., 1995, 1996; Spector et al., 2000). At the same time, disagreements between the results obtained by calculations using this method and by experiments have also been noted, and the origin of this inconsistency was discussed by Schutz and Warshel (2001). These include the use of a low value for the protein dielectric constant and the neglect of protein reorganization. The advantages and limitations of this method have been also been discussed by Hendsch and Tidor (1994).

The total electrostatic strength of an ion pair ($\Delta\Delta G_{tot}$) is

$$\Delta\Delta G_{tot} = \Delta\Delta G_{dslv} + \Delta\Delta G_{brd} + \Delta\Delta G_{prt},$$

where $\Delta\Delta G_{dslv}$ is the desolvation energy penalty incurred by the ion-pairing residues because of the desolvation of the charged side chains from the highly polar solvent water to the lower dielectric folded protein interior. This energy penalty depends on the location of the charged residues in the folded protein. $\Delta\Delta G_{brd}$ is the favorable energy term that represents the electrostatic interaction between charged residue side-chain groups. This term is sensitive to the geometrical orientation of the side-chain charged groups with respect to each other. $\Delta\Delta G_{prt}$ represents the electrostatic interaction of the ion-pairing residue side-chain charged groups with all other polar and ionized groups in the protein. For each ion pair, we have also computed an additional term, the association energy ($\Delta\Delta G_{assoc}$). $\Delta\Delta G_{assoc}$ takes into account the desolvation energy penalty for the ion pair and the electrostatic interaction between its side-chain charged groups. However, it ignores the electrostatic interaction between the ion pair and the charges in the rest of the protein (Hendsch and Tidor, 1994). All the energy terms, including $\Delta\Delta G_{tot}$, are sensitive to the value of the dielectric constant used for the protein interior.

$\Delta\Delta G_{dslv}$ depends upon the location (burial) of the charged residues (Hendsch and Tidor, 1994; Kumar and Nussinov, 1999). $\Delta\Delta G_{prt}$ depends upon the distribution of partial atomic charges in the rest of the proteins with respect to the ion pair. Hence, both $\Delta\Delta G_{dslv}$ and $\Delta\Delta G_{prt}$ represent the free energy contributions caused by the environment of the ion pair and are sensitive to the variations in the protein structural context, such as the movement of the side chains with respect to other polar/ionized group(s) and to the protein surface. On the other hand, the bridge energy ($\Delta\Delta G_{brd}$) and association energy ($\Delta\Delta G_{assoc}$) terms are sensitive primarily to the ion pair geometry. The protein structural context also affects these terms. It determines the screening of the electrostatic interactions between the

TABLE 1 Various types of ion pairs and their geometries in NMR conformer ensembles in our database

Ion pair	Protein Name	NMR conformer ensemble	Number of conformers	Ion pair type	N (%)	Average geometry	
						r (Å)	θ(°)
D40–K42	β-spectrin PH domain*	⟨1MPH)	50	Salt bridge	15 (30)	3.3 ± 0.3	111 ± 3
				C–C bridge	0 (0)		
				N–O bridge	0 (0)		
				Longer-range ion pair	35 (70)	6.3 ± 0.7	137 ± 19
K91–E95	β-spectrin PH domain*	⟨1MPH⟩	50	Salt bridge	40 (80)	3.1 ± 0.3	108 ± 13
				C–C bridge	0 (0)		
				N–O bridge	0 (0)		
				Longer-range ion pair	10 (20)	5.2 ± 0.5	109 ± 25
D12–K109	CheY†	⟨1CEY⟩	46	Salt bridge	1 (2.2)	3.6	49
				C–C bridge	0 (0)		
				N–O bridge	2 (4.4)	4.3	39
				Longer-range ion pair	43 (93.6)	8.1 ± 1.6	49 ± 21
D41–K45	CheY†	⟨1CEY⟩	46	Salt bridge	7 (15.2)	3.7 ± 0.2	137 ± 9
				C–C bridge	0 (0)		
				N–O bridge	1 (2.2)	4.1	152
				Longer-range ion pair	38 (82.6)	6.3 ± 0.9	156 ± 13
D57–K109	CheY†	⟨1CEY⟩	46	Salt bridge	3 (6.5)	3.6 ± 0.4	94 ± 9
				C–C bridge	0 (0)		
				N–O bridge	2 (4.4)	4.8	104
				Longer-range ion pair	41 (89.1)	6.7 ± 1.4	132 ± 23
E47–R73	c-Myb DNA binding domain repeat 1 (R1)†	⟨1MBF⟩	50	Salt bridge	27 (54)	3.5 ± 0.3	72 ± 12
				C–C bridge	0 (0)		
				N–O bridge	19 (38)	4.2 ± 0.2	62 ± 13
				Longer-range ion pair	4 (8)	6.0 ± 0.6	47 ± 11
D48–R81	c-Myb DNA binding domain repeat 1 (R1)‡	⟨1MBF⟩	50	Salt bridge	50 (100)	3.6 ± 0.03	104 ± 5
				C–C bridge	0 (0)		
				N–O bridge	0 (0)		
				Longer-range ion pair	0 (0)		
E49–K52	c-Myb DNA binding domain repeat 1 (R1)‡	⟨1MBF⟩	50	Salt bridge	16 (32)	3.4 ± 0.4	155 ± 12
				C–C bridge	0 (0)		
				N–O bridge	2 (4)	4.8	119
				Longer-range ion pair	32 (64)	8.3 ± 1.4	115 ± 24
K92–E99	c-Myb DNA binding domain repeat 2 (R2)‡	⟨1MBH⟩	50	Salt bridge	11 (22)	3.4 ± 0.3	44 ± 18
				C–C bridge	0 (0)		
				N–O bridge	0 (0)		
				Longer-range ion pair	39 (78)	10.1 ± 2.5	76 ± 21
D100–K133	c-Myb DNA binding domain repeat 2 (R2)‡	⟨1MBH⟩	50	Salt bridge	50 (100)	3.4 ± 0.05	96 ± 4
				C–C bridge	0 (0)		
				N–O bridge	0 (0)		
				Longer-range ion pair	0 (0)		
E150–R153	c-Myb DNA binding domain repeat 3 (R3)‡	⟨1MBK⟩	50	Salt bridge	16 (32)	3.6 ± 0.2	132 ± 14
				C–C bridge	0 (0)		
				N–O bridge	11 (22)	4.6 ± 0.4	140 ± 28
				Longer-range ion pair	23 (46)	8.6 ± 1.8	95 ± 29

(*Continued*)

TABLE 1 (*Continued*)

Ion pair	Protein Name	NMR conformer ensemble	Number of conformers	Ion pair type	*N* (%)	Average geometry	
						r (Å)	θ(°)
E151–R176	c-Myb DNA binding domain repeat 3 (R3)[‡]	⟨1MBK⟩	50	Salt bridge	36 (72)	3.6 ± 0.2	82 ± 12
				C–C bridge	0 (0)		
				N–O bridge	11 (22)	4.2 ± 0.1	74 ± 21
				Longer-range ion pair	3 (6)	6.4 ± 0.1	114 ± 5
K2–D7	CRIP[§]	⟨1IML⟩	48	Salt bridge	2 (4.2)	4.0	141
				C–C bridge	0 (0)		
				N–O bridge	1 (2.1)	5.0	138
				Longer-range ion pair	45 (93.7	8.0 ± 2.4	130 ± 31
K13–E27	CSE-I[¶]	⟨1B3C⟩	40	Salt bridge	12 (30)	3.4 ± 0.3	57 ± 29
				C–C bridge	0 (0)		
				N–O bridge	14 (35)	4.3 ± 0.3	79 ± 22
				Longer-range ion pair	14 (35)	6.1 ± 1.3	82 ± 26
E68–K84	Cyanovirin-N[‖]	⟨2EZN⟩	40	Salt bridge	5 (12.5)	3.8 ± 0.1	132 ± 6
				C–C bridge	0 (0)		
				N–O bridge	1 (2.5)	4.1	142
				Longer-range ion pair	34 (85)	6.3 ± 1.0	149 ± 21
E61–K99	Cytochrome c**	⟨2GIW⟩	40	Salt bridge	26 (65)	3.2 ± 0.4	111 ± 11
				C–C bridge	0 (0)		
				N–O bridge	1 (2.5)	4.3	121
				Longer-range ion pair	13 (32.5)	6.0 ± 0.9	115 ± 31
K62–E66	HMG1 Box2[††]	⟨1HSN⟩	49	Salt bridge	6 (12.2)	3.6 ± 0.3	139 ± 4
				C–C bridge	0 (0)		
				N–O bridge	2 (4.1)	4.5	144
				Longer-range ion pair	41 (83.7)	6.6 ± 1.1	152 ± 14
K54–D58	ISL-1[‡‡]	⟨1BW5⟩	50	Salt bridge	9 (18)	3.6 ± 0.3	111 ± 16
				C–C bridge	0 (0)		
				N–O bridge	6 (12)	4.3 ± 0.3	108 ± 11
				Longer-range ion pair	35 (70)	5.8 ± 1.1	115 ± 30
K4–E15	Protein G B1 domain[§§]	⟨1GB1⟩	60	Salt bridge	12 (20)	3.5 ± 0.4	135 ± 9
				C–C bridge	0 (0)		
				N–O bridge	8 (13.3)	4.3 ± 0.2	142 ± 8
				Longer-range ion pair	40 (66.7)	6.4 ± 1.2	155 ± 13
K4–E15	Protein G B1 domain[§§]	⟨3GB1⟩	32	Salt bridge	10 (31.2)	3.6 ± 0.3	128 ± 11
				C–C bridge	0 (0)		
				N–O bridge	3 (9.4)	4.4 ± 0.5	130 ± 10
				Longer-range ion pair	19 (59.4)	6.6 ± 1.5	139 ± 24
E27–K28	Protein G B1 domain[§§]	⟨1GB1⟩	60	Salt bridge	0 (0)		
				C–C bridge	0 (0)		
				N–O bridge	0 (0)		
				Longer-range ion pair	60 (100)	9.5 ± 1.6	97 ± 24
E27–K28	Protein G B1 domain[§§]	⟨3GB1⟩	32	Salt bridge	0 (0)		
				C–C bridge	0 (0)		
				N–O bridge	1 (3.1)	4.4	163
				Longer-range ion pair	31 (96.9)	9.1 ± 1.2	107 ± 18

(*Continued*)

TABLE 1 (*Continued*)

Ion pair	Protein Name	NMR conformer ensemble	Number of conformers	Ion pair type	*N* (%)	Average geometry	
						r (Å)	θ(°)
D47–K50	Protein G B1 domain§§	⟨1GB1⟩	60	Salt bridge	1 (1.7)	3.4	165
				C–C bridge	0 (0)		
				N–O bridge	4 (6.7)	4.6 ± 0.3	171 ± 6
				Longer-range ion pair	55 (91.6)	7.7 ± 1.2	136 ± 16
D47–K50	Protein G B1 domain§§	⟨3GB1⟩	32	Salt bridge	1 (3.1)	3.7	172
				C–C bridge	0 (0)		
				N–O bridge	6 (18.8)	4.5 ± 0.2	161 ± 10
				Longer-range ion pair	25 (78.1)	7.1 ± 1.3	139 ± 20
E108–R110	U1-SNRP A¶¶	⟨1FHT⟩	43	Salt bridge	2 (4.7)	3.9	149
				C–C bridge	0 (0)		
				N–O bridge	1 (2.3)	5.2	123
				Longer-range ion pair	40 (93)	10.8 ± 2.4	82 ± 47
Overall averages and range	All proteins	All NMR ensembles	1174	Salt bridge	358 (30.5)	3.5 ± 0.3	104 ± 28
				Range		2.5−3.99	6−172
				C–C bridge	0 (0)		
				N–O bridge	96 (8.2)	4.4 ± 0.3	105 ± 42
				Range		4.01−5.2	27−178
				Longer-range ion pair	720 (61.3)	7.6 ± 2.1	118 ± 39
				Range		4.2−14.8	4−175
Overall ion pair geometry				Average		6.1 ± 2.6	112 ± 36
				Range		2.5−14.8	4−178

The NMR conformer ensemble for each protein is shown by its PDB (Bernstein et al., 1977) entry. *N* and % denote the number and percentage of conformers in which a given ion pair forms a salt bridge, C–C bridge, N–O bridge, or longer-range ion pair. For each ion pair type, mean and SD values are presented for the side-chain charged group centroid distance, *r* (Å), and for the angular orientation of the side-chain charged groups (θ in degrees) with respect to each other. Standard deviations about the average values are presented only in those cases where three or more data points are available. The procedure for the computation of ion pair geometries and the criteria for various ion pairs types are in Materials and Methods. The ion pairs shown in this table were selected because they form salt bridges either in the crystal structure(s) or in the average energy-minimized NMR structure. In the cases where both of these are not available, the indicated ion pairs found to form salt bridges in the most representative conformer of the NMR ensemble are shown. This evidence is presented below for each protein.

*Ion pairs in β-spectrin pleckstrin homology (PH) domain (Nilges et al., 1997). These ion pairs form salt bridges in conformer 12 in NMR ensemble. Conformer 12 is the most representative for this ensemble.

†Ion pairs in bacterial chemotaxis protein, CheY. Two high-resolution crystal structures (PDB entries 1CHN and 3CHY (Bellsolell et al., 1994; Volz et al., 1991) are available along with the NMR structure (Moy et al., 1994). The ion pairs form salt bridges in one or the other crystal structure.

‡Ion pairs in c-Myb DNA-binding domain repeats R1, R2, and R3 (Ogata et al., 1995). The three repeats share 31–46% sequence homologies. Each repeat consists of 52 residues and has its own NMR conformer ensemble. The ion pairs form salt bridges in the corresponding average energy-minimized structures (PDB entries 1MBE, 1MBG, and 1MBJ for repeats R1, R2, and R3, respectively) for the NMR conformer ensembles.

§Ion pair in cysteine-rich intestinal protein (CRIP) (Perez-Alvarado et al., 1996). The ion pair forms a salt bridge in conformer 16, the most representative for the NMR ensemble.

¶Ion pair in CSE-I, a β-neurotoxin from scorpion (Jablonsky et al., 1999). The ion pair forms a salt bridge in the average energy-minimized structure (PDB entry 2B3C).

‖Ion pair in cyanovirin-N, a potent HIV-inactivating protein. Cyanovirin-N forms a domain-swapped dimer in its crystal form (PDB entry 3EZM, (Yang et al., 1999). However, it exists as a monomer in solution (Bewley et al., 1998). The ion pair forms a salt bridge in the crystal structure of cyanovirin-N. An average energy-minimized structure (PDB entry 2EZM) is also available for cyanovirin-N.

**Ion pair in reduced from of horse heart cytochrome c. The NMR structure for this protein was solved by Banci et al. (1999). The ion pair forms a salt bridge in the average energy-minimized structure (PDB entry 1GIW).

††Ion pair in HMG1 box 2 (Read et al., 1993). The ion pair forms a salt bridge in the average energy minimized structure (PDB entry 1HSM).

‡‡Ion pair in homeodomain of rat insulin-gene enhancer protein, ISL-1 (Ippel et al., 1999). The ion pair forms a salt bridge in conformer 6, the most representative for this conformer ensemble.

§§Ion pairs in the B1 domain of immunoglobulin-binding protein G. Two crystal structures (PDB entries 1PGA and 1PGB, (Gallagher et al., 1994)) and two NMR conformer ensembles (Gronenborn et al., 1991; Kuszewski et al., 1999) are available. In addition, an average energy-minimized structure (PDB entry 2GB1) is also available. The three ion pairs form salt bridges in one or the other crystal structure.

¶¶Ion pair in the N-terminal Rnp domain of U1A protein (Avis et al., 1996). The ion pair forms a salt bridge in conformer 36, the most representative for the NMR ensemble.

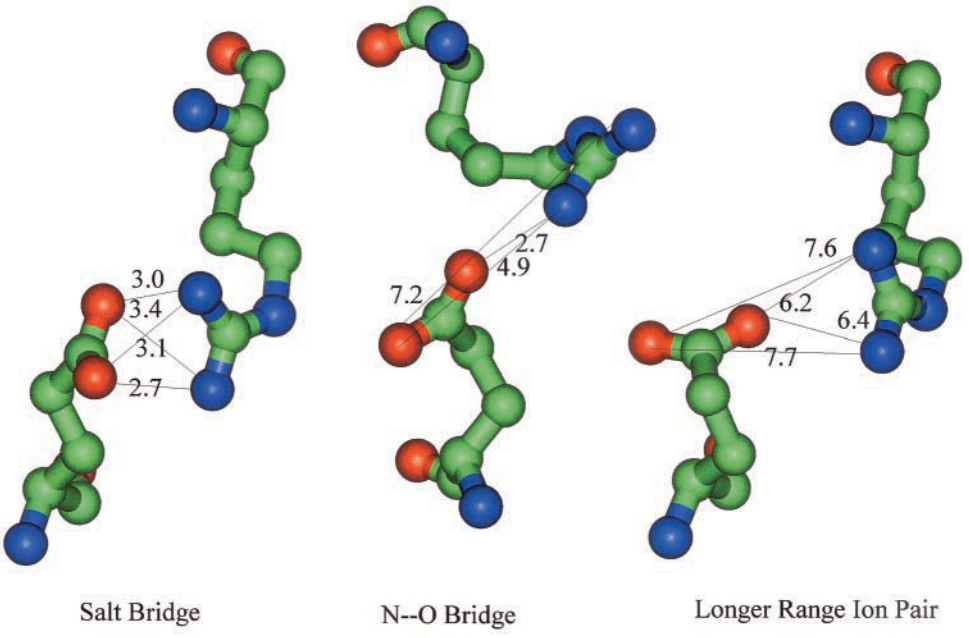

FIGURE 2 Different types of ion pairs in proteins defined in Materials and Methods. Salt bridge, N–O bridge, and longer-range ion pair types are shown for Glu 47–Arg 73 in the NMR conformer ensemble (⟨1MBF⟩) of c-Myb DNA-binding domain repeat 1 (R1). The side-chain charged groups are close and optimally oriented toward each other in a salt bridge. The separation of side-chain charged groups increases in N–O bridges and in longer-range ion pairs. Atoms are color coded. Oxygen atoms are in red, nitrogen in blue, and carbon in green. In general, ion pairs are strongest when they form salt bridges, considerably weaker when they form N–O bridges, and the weakest when they form longer-range ion pairs.

ion-pairing residue side-chain charged groups. All the free energy terms are sensitive to variations in side-chain conformations of the charged residues, explaining the relationship between the fluctuations in ion pairs and their geometries and electrostatic strengths in NMR conformer ensembles of proteins (Kumar and Nussinov, 2000, 2001a).

The detailed protocol for the computation of the electrostatic free energy contribution of an ion pair has been described previously (Hendsch and Tidor, 1994; Kumar and Nussinov, 1999, 2000, 2001a; Kumar et al., 2000b). We have used this protocol except with a finer grid spacing (0.5 Å per grid step). For the crystal structures, we have fixed the hydrogen atoms using the BIOPOLYMER module of INSIGHT II. We have energy minimized the crystal structure, keeping the nonhydrogen atom positions fixed. The minimization process consisted of 100 steps of steepest descent followed by 500 steps of conjugate gradient. Energy minimization is carried out using the CFF91 force field in the DISCOVER module of INSIGHT II. This procedure improves the accuracy of the continuum electrostatic calculations (Nielsen et al., 1999). The calculations are carried out for all ion pairs listed in Table 1, in the NMR conformer ensembles, energy-minimized average NMR structures, and crystal structures. All calculations have been carried out at pH 7.0 and at zero ionic strength. All energy values correspond to 25°C. All the calculations, except those shown in Table 5, have been carried out using a protein dielectric constant of 4.

RESULTS

Variations in ion pair geometries in NMR conformer ensembles

Table 1 shows the frequencies with which the ion pairs in our database form salt bridges, N–O bridges, and longer-range ion pairs. It also presents the average ion pair geometry in each class. Most ion pairs form salt bridges only in a minority of the conformers. The incidence of N–O bridge formation by the ion pairs is even more rare, and there are no C–C bridges in our database. Fig. 2 shows an example of the change in the interaction between the side-chain charged groups when an ion pair forms a salt bridge, a N–O bridge, and a longer-range ion pair. Overall, there are 1174 homologous observations for the 22 unique ion pairs in Table 1. These form 358 (30.5%) salt bridges, 96 (8.2%) N–O bridges, and 720 (61.3%) longer-range ion pairs. The ion pairs D48–R81 in c-Myb R1 (c-Myb DNA-binding domain repeat 1) and D100–R123 in c-Myb R2 occupy structurally equivalent positions. The same holds for E47–R73 in c-Myb R1 and E151–R176 in c-Myb R3 (Kumar and Nussinov, 2001a). However, their structural contexts are different because the repeats share only 31–46% sequence identity. These ion pairs are not grouped here. The observations on ion pairs in two NMR conformer ensembles of the B1 domain of protein G are also treated separately.

The distribution of ion pair geometries is shown in Fig. 3. The geometries of salt bridges, N–O bridges, and longer-range ion pairs are shown in blue, green, and red, respectively. The side-chain charged group centroid distances fall within 5 Å for almost all N–O bridges. In contrast, in the majority of the longer-range ion pairs, they are >5 Å apart. Overall, the spatial orientation of the side-chain charged groups with respect to each other is most favorable in salt bridges with the average geometric parameters being r_{av} =

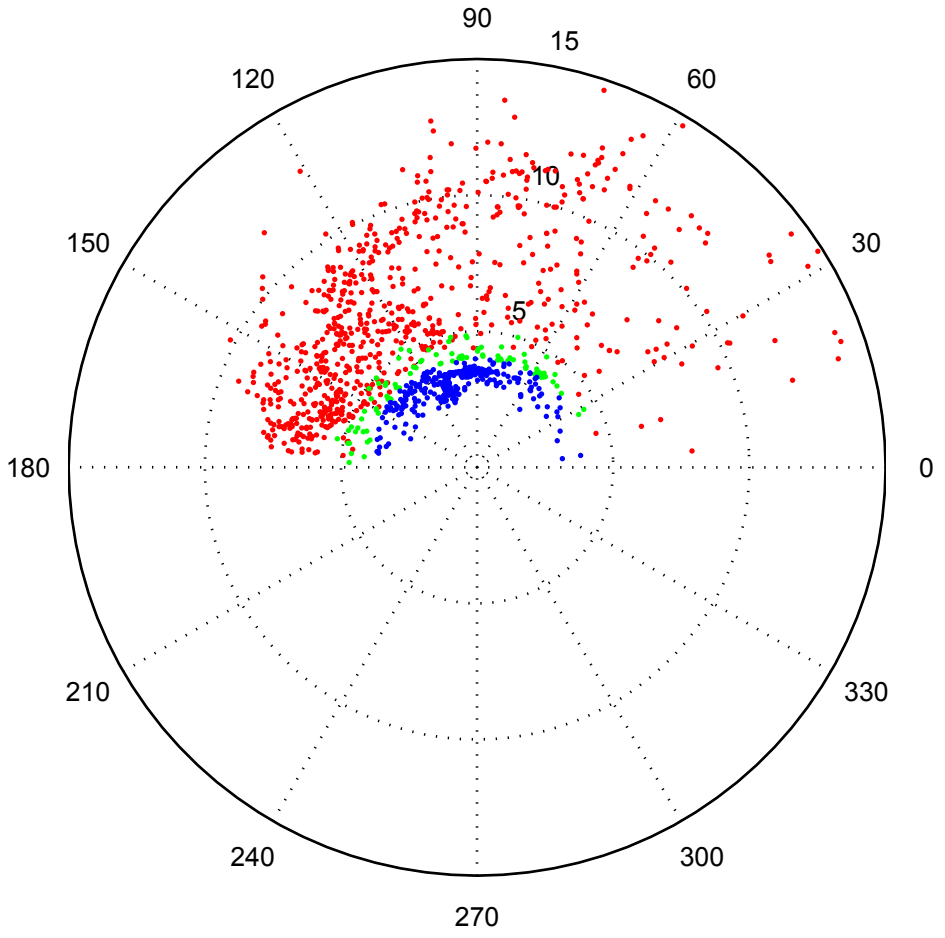

FIGURE 3 The spatial orientations of side-chain charged groups in 1174 ion pairs in 14 NMR conformer ensembles. The number of unique ion pairs is 22. The ion pair types are color coded. Salt bridges are in blue, N–O bridges in green, and longer range ion pairs in red. The geometry of an ion pair can be characterized in terms of distance between its side-chain charged group centroids (r (Å)) and angular orientation of its side-chain charged groups (θ (°)). In this polar plot, radii of the concentric circles represent different r (Å) values. The dotted lines connecting the circles denote different θ (°) values. Additional details are given in Materials and Methods. Most of the ion pairs with $r \leq 5$ Å are stabilizing. In contrast, most of the ion pairs with $r > 5$ Å are destabilizing.

3.5 ± 0.3 Å and θ_{av} = 104 ± 27°. N–O bridges also have good geometries (r_{av} = 4.4 ± 0.3 Å; θ_{av} = 105 ± 42°). In longer-range ion pairs, the geometry is most unfavorable with r_{av} = 7.6 ± 2.1 Å and θ_{av} = 118 ± 39° (Table 1). The geometries for longer-range ion pairs vary to a greater extent (Fig. 3 and Table 1). The average distance between the side-chain charged group centroids in the longer-range ion pair increases by 4.1 Å with respect to the salt bridges and by 3.2 Å with respect to the N–O bridges.

Electrostatic free energy contribution by various ion pair types

Table 2 presents the electrostatic strengths of each ion pair in the ensembles, classified by the ion pair types. In all NMR conformer ensembles the free energy terms that are sensitive to variations in ion pair geometry, $\Delta\Delta G_{brd}$ and $\Delta\Delta G_{assoc}$, are most favorable for salt bridges followed by N–O bridges (Table 2). They are the weakest for the longer-range ion pairs. Because $\Delta\Delta G_{dslv}$ and $\Delta\Delta G_{prt}$ values do not depend upon ion pair geometry, these free energy terms do not show a consistent variation for ion pair types.

The favorable interaction ($\Delta\Delta G_{brd}$) between oppositely charged side-chain groups in an ion pair is mainly responsible for overcoming the unfavorable desolvation energy penalty, $\Delta\Delta G_{dslv}$ (Kumar and Nussinov, 1999). Consistently, we find that salt bridges have better overall electrostatic strengths ($\Delta\Delta G_{tot}$) than N–O bridges and longer-range ion pairs (Table 2). The only notable exception is that of ion pair E151–R176 in the NMR conformer ensemble (⟨1MBK⟩) of c-Myb DNA-binding domain repeat 3 (R3). In this ensemble, the average $\Delta\Delta G_{tot}$ (−6.4 ± 3.9 kcal/mol) of the three longer-range ion pairs formed by Glu 151 and Arg 176 is slightly stronger than that (−5.8 ± 2.8 kcal/mol) for

TABLE 2 Electrostatic strengths of ion pairs in different categories

Ion pair and NMR conformer ensemble	Ion pair type	$\Delta\Delta G_{dslv}$ (kcal/mol)	$\Delta\Delta G_{brd}$ (kcal/mol)	$\Delta\Delta G_{prt}$ (kcal/mol)	$\Delta\Delta G_{tot}$ (kcal/mol)	$\Delta\Delta G_{assoc}$ (kcal/mol)
D40–K42	Salt bridge	2.7 ± 0.7	−4.9 ± 0.8	−11.0 ± 2.9	−13.3 ± 2.7	−4.3 ± 0.6
⟨1MPH⟩	Longer-range ion pair	2.6 ± 0.9	−0.5 ± 0.2	−12.5 ± 3.0	−10.4 ± 2.9	0.3 ± 0.2
K91–E95	Salt bridge	6.6 ± 2.8	−3.6 ± 1.3	−6.2 ± 3.2	−3.3 ± 2.1	−0.9 ± 0.6
⟨1MPH⟩	Longer-range ion pair	4.5 ± 1.3	−0.9 ± 0.3	−6.5 ± 2.0	−2.9 ± 1.7	−0.1 ± 0.1
D12–K109	Salt bridge	10.9	−5.4	−3.1	+2.4	−1.6
⟨1CEY⟩	N–O bridge	11.6	−4.0	+2.8	+10.4	−0.9
	Longer-range ion pair	11.2 ± 3.4	−1.1 ± 0.6	−2.8 ± 4.6	+7.3 ± 4.1	−0.3 ± 0.1
D41–K45	Salt bridge	4.0 ± 0.5	−2.9 ± 0.7	−2.8 ± 1.8	−1.8 ± 1.3	−1.8 ± 0.5
⟨1CEY⟩	N–O bridge	4.9	−2.6	+2.2	+4.5	−1.4
	Longer-range ion pair	4.0 ± 1.5	−0.8 ± 0.4	−2.7 ± 2.1	+0.5 ± 1.8	−0.4 ± 0.2
D57–K109	Salt bridge	8.9 ± 1.3	−5.3 ± 2.1	−3.3 ± 6.6	+0.3 ± 3.2	−2.4 ± 1.7
⟨1CEY⟩	N–O bridge	7.8	−2.2	−0.9	+4.6	−0.8
	Longer-range ion pair	8.3 ± 2.3	−1.3 ± 0.5	+0.7 ± 4.6	+7.7 ± 4.4	−0.5 ± 0.2
E47–R73	Salt bridge	8.5 ± 1.8	−8.0 ± 2.8	−4.9 ± 3.5	−4.3 ± 4.1	−4.5 ± 1.9
⟨1MBF⟩	N–O bridge	8.0 ± 2.6	−5.6 ± 2.0	−6.2 ± 3.1	−3.8 ± 3.1	−3.4 ± 1.6
	Longer-range ion pair	13.2 ± 3.4	−2.0 ± 1	−7.1 ± 3.6	+4.0 ± 0.5	−0.5 ± 0.2
D48–R81 ⟨1MBF⟩	Salt bridge	9.6 ± 1.2	−11.4 ± 0.9	−2.3 ± 1.2	−4.1 ± 1.0	−5.8 ± 0.3
E49–K52	Salt bridge	2.6 ± 2.0	−4.1 ± 0.9	−3.4 ± 2.1	−4.9 ± 2.0	−3.3 ± 0.6
⟨1MBF⟩	N–O bridge	3.9	−1.5	−4.4	−2.1	−0.9
	Longer-range ion pair	4 ± 1.5	−0.4 ± 0.2	−6.9 ± 2.9	−3.3 ± 2.4	−0.2 ± 0.1
K92–E99	Salt bridge	8.8 ± 2.7	−6.3 ± 1.7	−8.5 ± 2.7	−6.0 ± 2.4	−3.3 ± 0.6
⟨1MBH⟩	Longer-range ion pair	7.8 ± 2.5	−0.5 ± 0.7	−10.0 ± 3.2	−2.7 ± 3.3	−0.2 ± 0.3
D100–R133 ⟨1MBH⟩	Salt bridge	10.1 ± 1.2	−12.4 ± 0.7	−2.1 ± 1.1	−4.4 ± 0.9	−6.4 ± 0.3
E150–R153	Salt bridge	1.6 ± 0.8	−3.8 ± 1.0	−2.6 ± 2.5	−4.9 ± 1.6	−3.5 ± 0.9
⟨1MBK⟩	N–O bridge	1.3 ± 1.0	−2.9 ± 1.3	−1.6 ± 2.2	−3.2 ± 1.7	−2.6 ± 1.1
	Longer-range ion pair	1.6 ± 1.2	−0.4 ± 0.3	−1.4 ± 1.5	−0.3 ± 1.5	−0.3 ± 0.2
E151–R176	Salt bridge	6.6 ± 2.1	−7.6 ± 2.0	−4.8 ± 3.1	−5.8 ± 2.8	−4.8 ± 0.9
⟨1MBK⟩	N–O bridge	6.5 ± 1.8	−5.9 ± 1.4	−4.6 ± 3.5	−3.9 ± 2.6	−3.6 ± 0.7
	Longer-range ion pair	6.3 ± 3.1	−0.9 ± 0.2	−11.8 ± 6.8	−6.4 ± 3.9	−0.4
K2–D7	Salt bridge	0.4	−1.6	−0.2	−1.4	−1.4
⟨1IML⟩	N–O bridge	0.9	−1.2	+0.4	+0.1	−0.9
	Longer-range ion pair	1.2 ± 1.0	−0.5 ± 0.3	−0.1 ± 0.7	+0.6 ± 1	−0.4 ± 0.2
K13–E27	Salt bridge	5.8 ± 3.2	−4.3 ± 1.2	+0.2 ± 1.6	+1.8 ± 3.4	−2.4 ± 0.7
⟨1B3C⟩	N–O bridge	3.7 ± 2.1	−2.0 ± 0.6	−0.8 ± 1.7	+0.9 ± 2.4	−1.2 ± 0.3
	Longer-range ion pair	5.0 ± 2.2	−1.0 ± 0.5	−0.6 ± 2.7	+3.4 ± 2.8	−0.5 ± 0.2
E68–K84	Salt bridge	1.8 ± 0.8	−2.2 ± 0.6	+0 ± 0.2	−0.4 ± 0.5	−1.6 ± 0.4
⟨2EZN⟩	N–O bridge	0.8	−1.3	−0.1	−0.6	−1.0
	Longer-range ion pair	1.7 ± 1.6	−0.8 ± 0.2	−0.3 ± 0.9	+0.7 ± 1.7	−0.5 ± 0.2
E61–K99	Salt bridge	1.4 ± 0.9	−4.9 ± 1.0	−1.7 ± 1.8	−5.2 ± 1.4	−4.3 ± 1.0
⟨2GIW⟩	N–O bridge	0.6	−1.5	−0.6	−1.6	−1.3
	Longer-range ion pair	1.8 ± 1.4	−0.9 ± 0.3	−3.8 ± 3.2	−2.9 ± 2.2	−0.6 ± 0.2

(*Continued*)

TABLE 2 (*Continued*)

Ion pair and NMR conformer ensemble	Ion pair type	$\Delta\Delta G_{dslv}$ (kcal/mol)	$\Delta\Delta G_{brd}$ (kcal/mol)	$\Delta\Delta G_{prt}$ (kcal/mol)	$\Delta\Delta G_{tot}$ (kcal/mol)	$\Delta\Delta G_{assoc}$ (kcal/mol)
K62–E66	Salt bridge	1.7 ± 0.7	−2.4 ± 0.7	−0.3 ± 0.2	−1.0 ± 0.4	−1.9 ± 0.5
⟨1HSN⟩	N–O bridge	0.8	−1.6	−0.6	−1.4	−1.4
	Longer-range ion pair	1.7 ± 1.4	−0.7 ± 0.2	−0.1 ± 1.0	+1.0 ± 2.1	−0.5 ± 0.2
K54–D58	Salt bridge	5.4 ± 3.7	−4.2 ± 1.7	+0.4 ± 0.8	+1.6 ± 2.6	−2.4 ± 0.7
⟨1BW5⟩	N–O bridge	4.7 ± 1.7	−2.7 ± 0.6	+0.6 ± 1.0	+2.7 ± 2.2	−1.6 ± 0.6
	Longer-range ion pair	4.6 ± 1.8	−1.3 ± 0.6	+0.5 ± 1.7	+3.7 ± 2.4	−0.7 ± 0.3
K4–E15	Salt bridge	1.5 ± 0.8	−2.8 ± 0.8	−0.2 ± 0.5	−1.4 ± 0.9	−2.3 ± 0.7
⟨1GB1⟩	N–O bridge	1.2 ± 0.9	−1.6 ± 0.3	−0.1 ± 0.5	−0.4 ± 0.6	−1.3 ± 0.2
	Longer-range ion pair	1.5 ± 1.0	−0.6 ± 0.2	−0.1 ± 0.7	+0.8 ± 1.2	−0.5 ± 0.2
K4–E15	Salt bridge	2.5 ± 1.6	−3.1 ± 1.0	−0.2 ± 0.2	−0.8 ± 1.7	−2.4 ± 0.8
⟨3GB1⟩	N–O bridge	2.6 ± 1.7	−1.4 ± 0.1	−0.3 ± 0.2	+0.9 ± 1.6	−1.0 ± 0.2
	Longer-range ion pair	3.1 ± 1.6	−0.7 ± 0.4	−0.6 ± 0.9	+1.8 ± 1.7	−0.4 ± 0.2
E27–K28 ⟨1GB1⟩	Longer-range ion pair	1.6 ± 1.1	−0.3 ± 0.1	0.0 ± 0.6	+1.4 ± 1.1	−0.2 ± 0.1
E27–K28	N–O bridge	2.8	−1.8	−1.9	−0.9	−1.1
⟨3GB1⟩	Longer-range ion pair	1.4 ± 0.8	−0.3 ± 0.1	−0.4 ± 0.5	+0.7 ± 0.7	−0.2 ± 0.1
D47–K50	Salt bridge	0.3	−2.2	−0.2	−2.1	−1.9
⟨1GB1⟩	N–O bridge	1.3 ± 0.4	−1.2 ± 0.3	+0.3 ± 0.3	+0.4 ± 0.3	−0.8 ± 0.2
	Longer-range ion pair	1.2 ± 1.1	−0.3 ± 0.2	−0.5 ± 0.5	+0.8 ± 1.0	−0.2 ± 0.1
D47–K50	Salt bridge	1.7	−3.0	+0.1	−1.3	−2.3
⟨3GB1⟩	N–O bridge	1.9 ± 0.7	−1.5 ± 0.3	−1.2 ± 0.7	−0.8 ± 1.4	−1.0 ± 0.3
	Longer-range ion pair	3.5 ± 1.7	−0.6 ± 0.3	−1.9 ± 1.5	+1.1 ± 0.9	−0.4 ± 0.2
E108–R110	Salt bridge	1.0	−2.0	−0.8	−1.8	−1.8
⟨1FHT⟩	N–O bridge	0.5	−0.8	−1.4	−1.7	−0.7
	Longer-range ion pair	1.0 ± 0.8	−0.3 ± 0.2	−0.9 ± 1.1	−0.1 ± 1.2	−0.2 ± 0.1

The data presented in this table summarize the results of 1174 sets of electrostatic free energy calculations using the continuum electrostatics methodology. For each free energy term in different ion pair types, the average and SD values are given. The SD was computed only for those cases where at least three data points were available. The ion pairs and the NMR conformer ensembles are presented in the same order as Table 1. Additional details are given in Materials and Methods.

the 36 salt bridges formed by these residues in different NMR conformers. The average $\Delta\Delta G_{brd}$ weakens by ~7 kcal/mol from being −7.6 ± 2.0 kcal/mol for salt bridges to −0.9 ± 0.2 for longer-range ion pairs. Interestingly, $\Delta\Delta G_{prt}$ becomes stronger by ~7 kcal/mol from −4.8 ± 3.1 kcal/mol for salt bridges to −11.8 ± 6.8 kcal/mol for longer-range ion pairs. Hence, the loss in stability due to deterioration in ion pair geometry appears to have been compensated by the change in the structural context for the longer-range ion pairs. We observe this compensation type between $\Delta\Delta G_{brd}$ and $\Delta\Delta G_{prt}$ for other ion pairs, too.

A closer examination suggests that the longer-range ion pair in conformer 12 may be responsible for this exception. In this conformer, the interaction between Glu 151 and Arg 176 incurs a desolvation penalty ($\Delta\Delta G_{dslv}$) of 9.9 kcal/mol. $\Delta\Delta G_{brd}$ is only −1.1 kcal/mol. However, the interaction is particularly strong with the charges in the rest of the protein. $\Delta\Delta G_{prt}$ is −19.6 kcal/mol, and $\Delta\Delta G_{tot}$ is −10.8 kcal/mol. $\Delta\Delta G_{assoc}$, which indicates the stability of the ion pair in the absence of the charges in the rest of the protein, is only −0.4 kcal/mol. These free energy values differ from the average values for the free energy terms for E151–R176 in the whole NMR ensemble ($\Delta\Delta G_{dslv\text{-}av}$ = 6.6 ± 2.1 kcal/mol, $\Delta\Delta G_{brd\text{-}av}$ = −6.9 ± 2.5 kcal/mol, $\Delta\Delta G_{prt\text{-}av}$ = −5.2 ± 3.8 kcal/mol, $\Delta\Delta G_{tot\text{-}av}$ = −5.4 ± 2.9 kcal/mol, and $\Delta\Delta G_{assoc\text{-}av}$ = −4.3 ± 1.4 kcal/mol). In conformer 12, Glu 151 forms a salt bridge with Lys 144, with a very favorable geometry (r = 3.0 Å; θ = 34°). This alternative salt bridge (Kumar and Nussinov, 2001a), which is absent in the average energy-minimized structure of c-Myb DNA-binding domain repeat 3 (R3, PDB: 1MBJ) and other conformers in 1MBK, may be responsible for the large $\Delta\Delta G_{prt}$ value in conformer 12. This type of compensation between $\Delta\Delta G_{brd}$ and $\Delta\Delta G_{prt}$ contributions toward $\Delta\Delta G_{tot}$

has been reported earlier (Warshel and Russell, 1984; Cutler et al., 1989). In our own analysis, we have found that charged residue pairs constituting ion pairs in protein crystal structures and in NMR average energy-minimized structures often move farther apart and form alternative salt bridges with the other charged residues in the individual conformers in the ensembles (Kumar and Nussinov, 2000, 2001a). The average values for the ion pair geometric parameters (Table 1) and for the free energy terms (Table 2) are simple arithmetic averages computed over conformers in the ensembles. They do not reflect the experimentally measurable values of these parameters in solution.

Stabilizing and destabilizing electrostatic free energy contributions by ion pairs

Table 3 enumerates the times each ion pair has stabilizing and destabilizing electrostatic free energy contributions in the NMR conformer ensembles. In most cases, ion pairs are stabilizing when they form either salt or N–O bridges. Overall, ion pairs have stabilizing contributions in ~92.5% (331 of 358) of the salt bridges. This proportion is similar (86%) in a dataset of 222 salt bridges in 36 crystal structures of nonhomologous protein monomers (Kumar and Nussinov, 1999). Ion pairs have stabilizing contributions in ~68% (65 of 96) of the N–O bridges. Taken together, ion pairs have stabilizing electrostatic contributions in 396 (~87%) of 454 incidents where they form either salt bridges or N–O bridges. Only 237 (~33%) of 720 longer-range ion pairs are stabilizing. Overall, the ion pairs are stabilizing in 633 cases (salt bridges, 52.3%; N–O bridges, 10.3%; and longer-range ion pairs, 27.4%). In contrast, 483 of 541 (89.3%) destabilizing ion pairs are longer range. The changes in proportions of the stabilizing and destabilizing contributions of the ion pair types are significantly different at 95% level of confidence as indicated by a change in proportion test (Kumar and Bansal, 1998). These statistics relate to highly homologous observations.

Taking together, the results presented in this and previous sections, it appears that geometry plays a crucial role in determining the electrostatic strengths of ion pairs. From Tables 1–3 and Fig. 3, a distance of 5 Å between side-chain charged group centroids appears to be a natural cutoff between (largely stabilizing) close-range and (largely destabilizing) longer-range electrostatic interactions in proteins.

Role of protein structural context in ion pair stability

While ion pair geometry is crucial for electrostatic strength of an ion pair, our results also show the importance of the protein structural context. In 27 (of 358, 7.5%) cases, the ion pairs form salt bridges with destabilizing electrostatic contributions. These are distributed over 11 unique ion pairs (Table 3). In 13 of these, the electrostatic interaction between the salt bridges and the other polar and ionized groups in the protein is unfavorable. In the remaining 14 cases, the favorable bridge and protein energy terms are unable to overcome the unfavorable desolvation energy penalty. In 10 of the 27 cases, the salt bridges are only marginally destabilizing (0 kcal/mol $< \Delta\Delta G_{tot} <$ 1 kcal/mol). All 27 destabilizing salt bridges have favorable association free energies (Table 2). Thirty-one N–O bridges with destabilizing contributions show similar features (Tables 2 and 3). Furthermore, we cannot rule out the possibility of uncertainties in the NMR conformer ensembles being a contributing factor in these observations. In our previous analysis (Kumar and Nussinov, 1999) of 222 unique salt bridges from 36 nonhomologous high-resolution (1.6 Å or better) crystal structures, 32 salt bridges also had destabilizing electrostatic contributions because of similar reasons as those cited above.

On the other hand, 237 of the longer-range ion pairs are stabilizing even though the electrostatic interactions among their side-chain charged groups are weak. Because of high homology among the protein structural contexts across the conformers in NMR ensembles, a majority of these longer-range ion pairs are for ion pairs D40–K42 and K91–E95 in β-spectrin (⟨1MPH⟩), E49–K52 in c-Myb DNA-binding domain repeat 1 (R1) (⟨1MBF⟩), K92–E99 c-Myb DNA-binding domain repeat 2 (R2) (⟨1MBH⟩), E151–R176 in c-Myb DNA-binding domain repeat 3 (R3) (⟨1MBK⟩), E68–K84 in cyanovirin-N (⟨2EZN⟩), E61–K99 in horse heart cytochrome c (⟨2GIW⟩), and E108–R110 in U1 SNRP A (⟨1FHT⟩) (Table 2). These NMR conformer ensembles account for 188 (~79%) of 237 stabilizing longer-range ion pairs.

In 227 (~96%) of the 237 longer-range ion pairs the protein energy terms are favorable, compensating for the weak interaction between the charged residues. $\Delta\Delta G_{prt}$ are stronger than the $\Delta\Delta G_{brd}$ for the longer-range ion pairs in β-spectrin, CheY (ion pairs: D12–K109 and D41–K45), c-Myb DNA-binding domain repeat 1 (R1) (E47–R73 and E49–K52), repeat 2 (R2) (K92–E99), repeat 3 (R3) (E150–R153 and E151–R176), horse heart cytochrome c, B1 domain of protein G (E27–K28 in ⟨3GB1⟩ and D47–K50 in ⟨1GB1⟩ and ⟨3GB1⟩), and U1 SNRP A (Table 2). $\Delta\Delta G_{prt}$ terms are particularly strong for longer-range ion pairs in β-spectrin, c-Myb DNA-binding domain repeats (R1, R2, and R3), and horse heart cytochrome c (Table 2). These proteins account for 147 (~62%) stabilizing longer-range ion pairs (Tables 2 and 3). The average desolvation energy penalties for longer-range ion pairs are also small for E68–K84 in cyanovirin-N, E61–K99 in horse heart cytochrome c, and E108–R110 in U1 SNRP A (⟨1FHT⟩) (Table 2). Of the 237, 103 (~43%) are only weakly stabilizing (-1 kcal/mol $< \Delta\Delta G_{tot} <$ 0 kcal/mol). The $\Delta\Delta G_{assoc}$ values for the longer-range ion pairs are weak. Of the 237, 231 (~97%) have $\Delta\Delta G_{assoc}$ values ranging between -1 and 0

TABLE 3 Stabilizing and destabilizing electrostatic contributions of ion pairs in various categories

Ion pair	NMR conformer ensemble	Ion pair type	Total number	Stabilizing number (%)	Destabilizing number (%)
D40–K42	⟨1MPH⟩	All	50	49 (98)	1 (2)
		Salt bridge	15	15 (100)	0 (0)
		Longer-range ion pair	35	34 (97.1)	1 (2.9)
K91–E95	⟨1MPH⟩	All	50	46 (92)	4 (8)
		Salt bridge	40	37 (92.5)	3 (7.5)
		Longer-range ion pair	10	9 (90)	1 (10)
D12–K109	⟨1CEY⟩	All	46	3 (6.5)	43 (93.5)
		Salt bridge	1	0 (0)	1 (100)
		N–O bridge	2	0 (0)	2 (100)
		Longer-range ion pair	43	3 (7)	40 (93)
D41–K45	⟨1CEY⟩	All	46	21 (45.7)	25 (54.3)
		Salt bridge	7	7 (100)	0 (0)
		N–O bridge	1	0 (0)	1 (100)
		Longer-range ion pair	38	14 (36.8)	24 (63.2)
D57–K109	⟨1CEY⟩	All	46	1 (2.2)	45 (97.8)
		Salt bridge	3	1 (33.3)	2 (66.7)
		N–O bridge	2	0 (0)	2 (100)
		Longer-range ion pair	41	0 (0)	41 (100)
E47–R73	⟨1MBF⟩	All	50	39 (78)	11 (22)
		Salt bridge	27	23 (85.2)	4 (14.8)
		N–O bridge	19	16 (84.2)	3 (15.8)
		Longer-range ion pair	4	0 (0)	4 (100)
D48–R81	⟨1MBF⟩	All	50	50 (100)	0 (0)
		Salt bridge	50	50 (100)	0 (0)
E49–K52	⟨1MBF⟩	All	50	48 (96)	2 (4)
		Salt bridge	16	15 (93.8)	1 (6.2)
		N–O bridge	2	2 (100)	0 (0)
		Longer-range ion pair	32	31 (96.9)	1 (3.1)
K92–E99	⟨1MBH⟩	All	50	42 (84)	8 (16)
		Salt bridge	11	11 (100)	0 (0)
		Longer-range ion pair	39	31 (79.5)	8 (20.5)
D100–R133	⟨1MBH⟩	All	50	50 (100)	0 (0)
		Salt bridge	50	50 (100)	0 (0)
E150–R153	⟨1MBK⟩	All	50	38 (76)	12 (24)
		Salt bridge	16	16 (100)	0 (0)
		N–O bridge	11	11 (100)	0 (0)
		Longer-range ion pair	23	11 (47.8)	12 (52.2)
E151–R176	⟨1MBK⟩	All	50	48 (96)	2 (4)
		Salt bridge	36	34 (94.4)	2 (5.6)
		N–O bridge	11	11 (100)	0 (0)
		Longer-range ion pair	3	3 (100)	0 (0)
K2–D7	⟨1IML⟩	All	48	16 (33.3)	32 (66.7)
		Salt bridge	2	2 (100)	0 (0)
		N–O bridge	1	0 (0)	1 (100)
		Longer-range ion pair	45	14 (31.1)	31 (68.9)

(*Continued*)

TABLE 3 (*Continued*)

Ion pair	NMR conformer ensemble	Ion pair type	Total number	Stabilizing number (%)	Destabilizing number (%)
K13–E27	⟨1B3C⟩	All	40	14 (35)	26 (65)
		Salt bridge	12	7 (58.3)	5 (41.7)
		N–O bridge	14	7 (50)	7 (50)
		Longer-range ion pair	14	0 (0)	14 (100)
E68–K84	⟨2EZN⟩	All	40	25 (60)	15 (40)
		Salt bridge	5	4 (80)	1 (20)
		N–O bridge	1	1 (100)	0 (0)
		Longer-range ion pair	34	20 (58.8)	14 (41.2)
E61–K99	⟨2GIW⟩	All	40	40 (100)	0 (0)
		Salt bridge	26	26 (100)	0 (0)
		N–O bridge	1	1 (100)	0 (0)
		Longer-range ion pair	13	13 (100)	0 (0)
K62–E66	⟨1HSN⟩	All	49	18 (36.7)	31 (63.3)
		Salt bridge	6	6 (100)	0 (0)
		N–O bridge	2	2 (100)	0 (0)
		Longer-range ion pair	41	10 (24.4)	31 (75.6)
K54–D58	⟨1BW5⟩	All	50	5 (10)	45 (90)
		Salt bridge	9	4 (44.4)	5 (55.6)
		N–O bridge	6	1 (16.7)	5 (83.3)
		Longer-range ion pair	35	0 (0)	35 (100)
K4–E15	⟨1GB1⟩	All	60	28 (46.7)	32 (53.3)
		Salt bridge	12	11 (91.7)	1 (8.3)
		N–O bridge	8	5 (62.5)	3 (37.5)
		Longer-range ion pair	40	12 (30)	28 (70)
K4–E15	⟨3GB1⟩	All	32	12 (37.5)	20 (62.5)
		Salt bridge	10	8 (80)	2 (20)
		N–O bridge	3	1 (33.3)	2 (66.7)
		Longer-range ion pair	19	3 (15.8)	16 (84.2)
E27–K28	⟨1GB1⟩	All	60	0 (0)	60 (100)
		Longer-range ion pair	60	0 (0)	60 (100)
E27–K28	⟨3GB1⟩	All	32	3 (9.4)	29 (90.6)
		N–O bridge	1	1 (100)	0 (100)
		Longer-range ion pair	31	2 (9.5)	29 (90.5)
D47–K50	⟨1GB1⟩	All	60	6 (10)	54 (90)
		Salt bridge	1	1 (100)	0 (0)
		N–O bridge	4	1 (25)	3 (75)
		Longer-range ion pair	55	4 (7.3)	51 (92.7)
D47–K50	⟨3GB1⟩	All	32	7 (21.9)	25 (78.1)
		Salt bridge	1	1 (100)	0 (0)
		N–O bridge	6	4 (66.7)	2 (33.3)
		Longer-range ion pair	25	2 (8)	23 (92)
E108–R110	⟨1FHT⟩	All	43	24 (55.8)	19 (44.2)
		Salt bridge	2	2 (100)	0 (0)
		N–O bridge	1	1 (100)	0 (0)
		Longer-range ion pair	40	21 (52.5)	19 (47.5)

(*Continued*)

TABLE 3 (*Continued*)

Ion pair	NMR conformer ensemble	Ion pair type	Total number	Stabilizing number (%)	Destabilizing number (%)
Total		All	1174	633 (53.9)	541 (46.1)
		Salt bridge	358	331 (92.5)	27 (7.5)
		N–O bridge	96	65 (67.7)	31 (32.3)
		Longer-range ion pair	720	237 (32.9)	483 (67.1)

All stands for all the ion pairs. Number and % denote number and percentage of conformers in which the ion pairs are stabilizing or destabilizing, respectively. Total number indicates the number of conformers in the NMR conformer ensemble.

kcal/mol. Hence, the association of the side-chain charged groups in these longer-range ion pairs would be only marginally favorable in the absence of the stabilizing effect of the neighboring charged residues. These conformer ensembles are also rich in alternative salt bridges that are not observed in the crystal structures, NMR average energy-minimized structures, or most representative conformers of the ensemble (Kumar and Nussinov, 2001a). Hence, the stronger $\Delta\Delta G_{prt}$ contributions for many of these longer-range ion pairs could be caused by the presence of charged residue(s) closer to the charged residue(s) in the original ion pair. When such a situation occurs in a NMR conformer, the original ion pair breaks and alternative salt bridge(s) form. An original ion pair is the one formed in the reference structure. As suggested by Warshel and coworkers (Warshel and Russell, 1984; Cutler et al., 1989), this type of compensation between $\Delta\Delta G_{brd}$ and $\Delta\Delta G_{prt}$ may be an integral part of the ion pair energetics. The effects of other polar and ionized groups on the stability of a given ion pair has already been studied earlier (Hwang and Warshel, 1988).

Ion pair geometry and electrostatic strength relationship in protein crystal structures and NMR average structures

We have also analyzed ion pair geometry and stability in crystal structures and NMR average energy-minimized structures of the proteins in our database, where two or more sets of atomic coordinates are available. Crystal structures are available for CheY (PDB entries 1CHN and 3CHY), cyanovirin-N (PDB entry 3EZM), and B1 domain of protein G (PDB entries 1PGA and 1PGB). For cyanovirin-N and B1 domain of protein G, the NMR average energy-minimized structures are also available (PDB entries 2EZM and 2GB1, respectively). These proteins contain seven ion pairs with 17 sets of calculations (Table 4). The results are consistent with those on NMR conformer ensembles. In all seven cases, the ion pairs have stronger $\Delta\Delta G_{brd}$ and $\Delta\Delta G_{assoc}$ terms when they have better geometries. For six of the seven ion pairs, the better geometries result in better overall stability. The only exception is E68–K84 in cyanovirin-N. Interestingly, this protein forms domain-swapped dimers in crystals but is monomeric in solution (Bewely et al., 1998; Yang et al., 1999).

DISCUSSION

Here we study the relationship between geometrical orientation and electrostatic strengths of charged residue pairs in proteins. We categorize our observations into different geometrical types. We characterize the geometrical orientation of the charged residues in an ion pair through measurements of the distance and the angular orientation of the side-chain charged groups. We find that the electrostatic interaction between the two charged residues is mostly stabilizing when the side-chain charged groups (as in salt bridges) or at least a pair of side-chain nitrogen and oxygen atoms (as in N–O bridges) are close. Although both salt and N–O bridges may contain hydrogen bonds, they are not the focus of our study. Furthermore, here we use whole side-chain charged groups rather than just acceptor and donor atoms.

Electrostatic interaction is a cooperative phenomenon. When the distance between the charged residues is small, their interaction (proportional to $1/r$) dominates. The interaction with other, more distant charges is proportional to $1/r^3$. However, when the residues in the ion pair have greater distances, this might not be the case. This situation occurs in some longer-range ion pairs. Here, to some extent we address this problem by recomputing alternative salt bridges as the original ones are broken. The NMR ensembles of β-spectrin, c-Myb DNA-binding domain repeats (R1, R2, and R3), and cytochrome c contain several such alternative salt bridges (Kumar and Nussinov, 2001a).

Our analysis facilitates the formulation of empirical rules for detection of stabilizing electrostatic interactions in proteins with known three-dimensional structures. These may be useful in identifying appropriate sites for inclusion of electrostatic interactions in de novo designed proteins to potentially enhance their thermal stability. An interesting example is that of glutamate dehydrogenase. It has been shown that *Pyrococcus furiosus* glutamate dehydrogenase (PfGDH) derives its high thermal stability (Tm = 113°C)

TABLE 4 Ion pairs in protein crystal structures and NMR average energy-minimized structures for which at least two sets of calculations are available

	Ion pair	PDB entry	r (Å)	θ (°)	$\Delta\Delta G_{dslv}$ (kcal/mol)	$\Delta\Delta G_{brd}$ (kcal/mol)	$\Delta\Delta G_{prt}$ (kcal/mol)	$\Delta\Delta G_{tot}$ (kcal/mol)	$\Delta\Delta G_{assoc}$ (kcal/mol)
CheY									
	D12–K109	$1CHN_{x\text{-}ray}$	3.2	54	14.3	−10.2	−4.8	−0.6	−3.4
		$3CHY_{x\text{-}ray}$	4.7	57	10.9	−2.8	−8.3	−0.2	−0.8
	D57–K109	$1CHN_{x\text{-}ray}$	3.8	118	11.9	−4.6	−4.5	+2.7	−1.2
		$3CHY_{x\text{-}ray}$	3.2	90	8.8	−7.6	−4.3	−3.1	−3.6
	D41–K45	$1CHN_{x\text{-}ray}$	3.7	139	4.3	−3.1	−4.8	−3.6	−1.7
		$3CHY_{x\text{-}ray}$	4.6	151	4.1	−1.8	−4.9	−2.6	−0.9
Cyanovirin-N									
	E68–K84	$2EZM_{NMR}$	5.2	156	0.4	−0.9	0.0	−0.5	−0.7
		$3EZM_{x\text{-}ray}$	3.9	118	3.3	−1.7	−1.1	+0.6	−1.3
B1 domain of protein G									
	K4–E15	$2GB1_{NMR}$	5.0	156	0.6	−1.5	+0.1	−0.4	−0.8
		$1PGA_{x\text{-}ray}$	3.3	134	1.6	−2.6	−0.2	−1.2	−1.9
		$1PGB_{x\text{-}ray}$	2.6	119	1.8	−5.9	−0.3	−4.3	−5.2
	E27–K28	$2GB1_{NMR}$	9.8	98	1.1	−0.2	+0.2	+1.1	−0.2
		$1PGA_{x\text{-}ray}$	3.7	135	4.1	−2.9	−4.6	−3.4	−2.0
		$1PGB_{x\text{-}ray}$	8.9	92	3.2	−0.3	−3.8	−0.9	−0.2
	D47–K50	$2GB1_{NMR}$	7.8	141	0.5	−0.3	+0.1	+0.3	−0.3
		$1PGA_{x\text{-}ray}$	3.5	169	2.6	−3.1	−4.7	−5.1	−2.3
		$1PGB_{x\text{-}ray}$	4.0	138	2.0	−2.0	−4.6	−4.5	−1.4

Geometries and electrostatic strengths of ion pairs in protein crystal structures and energy-minimized average NMR structures are shown. The subscripts to PDB entries indicate whether the file contains a crystal structure (x-ray) or a NMR average energy-minimized structure (NMR).

from an increased occurrence of ion pairs and their networks within and across its six subunits (Kumar et al., 2000a,b; Yip et al., 1995). Incorporation of an ion pair network, found in the hinge region of PfGDH, at structurally equivalent positions in homologous (55% sequence identity) and less stable (Tm = 89°C) *Thermotoga maritima* glutamate dehydrogenase (TmGDH) failed to improve the stability of TmGDH (Lebbink et al., 1998). However, a similar attempt to engineer a 16-residue ion pair network across subunit interfaces in TmGDH succeeded in marginally improving its stability (Lebbink et al., 1999). Our results show that incorporation of close-range electrostatic interactions in designed proteins is more likely to improve protein stability if the side-chain charged group centroids are within 5.0 Å. Because most salt bridges are between sequentially close charged residues (Kumar and Nussinov, 1999), this may help in keeping the charged residues close in space, although side-chain motions would nevertheless present a problem.

Protein conformer ensembles can be obtained via conformational sampling around the native state by molecular dynamics simulations and NMR spectroscopy. Both methods have advantages and disadvantages (Kumar and Nussinov, 2000, 2001a). The quality of the experimental data used here is high (Kumar and Nussinov, 2001a). The issue of whether NMR ensembles reflect the protein motion in solution has been controversial but does not bear directly on the analysis performed here. Although it is not possible to completely separate the effects of geometry and the protein structural context, we have attempted to minimize this difficulty through calculations on ion pairs in different conformers of the same protein. The protein structural contexts for an ion pair are expected to be homologous in different NMR conformers for same protein.

Computation of experimentally measurable values for the electrostatic strengths of the ion pairs in solution is not feasible using the NMR conformer ensemble data. NMR conformer ensembles available from PDB cannot be treated as true ensembles in the statistical-mechanical sense. The NMR conformers are obtained by optimizing an energy function consisting of force-field energy terms and experimental restraints. This function is not a Hamiltonian, and probabilities of the individual conformers do not follow a Boltzmann distribution. Additionally, the number of conformers obtained by typical NMR experiments is quite small ($\sim 10^2$ to 10^3). Of these, only a limited number (10–50) of the conformers that have energy function values below an arbitrary threshold are presented in the PDB files. Even for these conformers, the PDB files do not contain data on relative populations of different conformers. Hence, our analysis does not involve comparison with the experimental results. The values of various electrostatic energy terms computed for the ion pairs serve the qualitative purpose of comparison among various ion pair types. These should not be taken as the quantitative estimates of the ion pair stabilities in proteins in aqueous solution. The experimental support of this analysis is implicit and limited to the use of experimental protein structural (NMR and x-ray crystal) data.

TABLE 5 Ion pairs in protein crystal structures and NMR average energy-minimized structures for which at least two sets of calculations are available, with $\epsilon_p = 20$

	Ion pair	PDB entry	r (Å)	θ (°)	$\Delta\Delta G_{dslv}$ (kcal/mol)	$\Delta\Delta G_{brd}$ (kcal/mol)	$\Delta\Delta G_{prt}$ (kcal/mol)	$\Delta\Delta G_{tot}$ (kcal/mol)	$\Delta\Delta G_{assoc}$ (kcal/mol)
Chey									
	D12–K109	$1CHN_{x\text{-}ray}$	3.2	54	2.6	−3.0	−1.0	−1.4	−1.6
		$3CHY_{x\text{-}ray}$	4.7	57	2.2	−1.4	−1.7	−0.9	−0.8
	D57–K109	$1CHN_{x\text{-}ray}$	3.8	118	2.4	−1.9	−0.7	−0.2	−1.0
		$3CHY_{x\text{-}ray}$	3.2	90	1.9	−2.7	−1.0	−1.8	−1.6
	D41–K45	$1CHN_{x\text{-}ray}$	3.7	139	1.0	−1.5	−0.9	−1.5	−1.1
		$3CHY_{x\text{-}ray}$	4.6	151	1.0	−1.1	−1.1	−1.2	−0.7
Cyanovirin-N									
	E68–K84	$2EZM_{NMR}$	5.2	156	0.2	−0.8	0.0	−0.6	−0.7
		$3EZM_{x\text{-}ray}$	3.9	118	0.7	−1.1	−0.2	−0.6	−0.9
B1 domain of protein G									
	K4–E15	$2GB1_{NMR}$	5.0	156	0.2	−0.8	0.0	−0.6	−0.7
		$1PGA_{x\text{-}ray}$	3.3	134	0.4	−1.3	−0.2	−1.1	−1.1
		$1PGB_{x\text{-}ray}$	2.6	119	0.4	−2.1	−0.2	−1.9	−1.9
	E27–K28	$2GB1_{NMR}$	9.8	98	0.3	−0.2	+0.1	+0.2	−0.2
		$1PGA_{x\text{-}ray}$	3.7	135	0.9	−1.4	−1.5	−2.0	−1.1
		$1PGB_{x\text{-}ray}$	8.9	92	0.8	−0.3	−1.6	−1.1	−0.2
	D47–K50	$2GB1_{NMR}$	7.8	141	0.2	−0.3	+0.1	0.0	−0.3
		$1PGA_{x\text{-}ray}$	3.5	169	0.6	−1.4	−1.3	−2.0	−1.1
		$1PGB_{x\text{-}ray}$	4.0	138	0.5	−1.0	−1.4	−1.9	−0.8

The ion pairs shown in this table are same as those in Table 4. This time, however, their electrostatic strengths have been calculated using a higher value (20) for the protein dielectric constant.

The continuum electrostatics methodology has been widely used. Like any other technique, this method also has its drawbacks. Hendsch and Tidor (1994) have discussed the limitations and advantages of computing the electrostatic strengths of ion pairs with respect to their hydrophobic isosteres. More recently, Schutz and Warshel (2001) have highlighted the limitations of this methodology. Estimates of desolvation free energy penalty paid by the charged residues and screening of electrostatic interactions between the ion-pairing residues and between the ion pair and the rest of the charges in the protein depend upon the dielectric constant (ϵ_p) used for the protein. The interior of a protein is often considered largely apolar, and the use of a low dielectric constant for the protein has been quite common. In our calculations, we have used a value of 4 for ϵ_p (Kumar and Nussinov, 1999, 2000, 2001a; Kumar et al., 2000b, 2001a). Such a value for protein dielectric constant has also been used by others (e.g. Hendsch and Tidor, 1994; Lounnas and Wade, 1997; Xu et al., 1997b; Xiao and Honig, 1999). However, the effective dielectric constant experienced by an ion pair in a protein depends on the protein relaxation and reorganization of the other polar and ionized groups in the protein (Sham et al., 1998; Schutz and Warshel, 2001). Recently, Schutz and Warshel (2001) have recommended the use of higher values (e.g., 20) for ϵ_p when computing the electrostatic strengths of ion pairs. In our more recent continuum electrostatic calculations, a value of 20 for ϵ_p also yields reasonable estimates for electrostatic strengths of the salt bridges in citrate synthase (Kumar and Nussinov, unpublished results).

To examine the effect of a higher ϵ_p on our results in the present study, we have recalculated the electrostatic strengths of the ion pairs shown in Table 4 with a protein dielectric constant $\epsilon_p = 20$. All other parameters were kept the same. The results of the new calculations are shown in Table 5. We have also computed the root mean square deviation (rmsd) values between various corresponding energy terms for the ion pairs in Tables 4 and 5. Due to the higher ϵ_p, the electrostatic energy terms $\Delta\Delta G_{dslv}$, $\Delta\Delta G_{brd}$, and $\Delta\Delta G_{prt}$ have smaller magnitudes. The rmsd values for $\Delta\Delta G_{dslv}$, $\Delta\Delta G_{brd}$, and $\Delta\Delta G_{prt}$ are 2.8 kcal/mol, 1.7 kcal/mol, and 0.9 kcal/mol, respectively. The differences in these three energy terms largely cancel according to Eq. 1 (Materials and Methods). Although the values of $\Delta\Delta G_{tot}$ still differ in the two calculation sets, the differences are smaller. The rmsd value for $\Delta\Delta G_{tot}$ is 0.2 kcal/mol. The rmsd value for $\Delta\Delta G_{aasoc}$ is 0.4 kcal/mol. In three of the four incidents where the ion pairs in Table 4 had destabilizing electrostatic contributions, they now become marginally stabilizing (Table 5). In the remaining case, the ion pair is still destabilizing, but only marginally. Not withstanding these differences, the trend with respect to the ion pair geometries remains the same. When the ion pair geometries are better, the energy terms $\Delta\Delta G_{brd}$ and $\Delta\Delta G_{assoc}$ are stronger and, in all but one (E68–K84 in cyanovirin-N) case, the overall electrostatic strengths, $\Delta\Delta G_{tot}$, are stronger as well (Table 5).

The protein structure environment, i.e., the presence of the other polar and ionized groups, also critically affects the electrostatic strength of the ion pair. There appears to be a

certain degree of compensation between the $\Delta\Delta G_{brd}$ and $\Delta\Delta G_{prt}$ terms for ion pairs across different conformers of the NMR ensemble. This compensation mechanism was first recognized by Warshel and coworkers (Warshel and Russell, 1984; Cutler et al., 1989; Hwang and Warshel, 1988; Sham et al., 1998; Schutz and Warshel, 2001). They have used this concept to explain why a simple reversal of charges on residues in enzyme-binding sites and substrates using genetic engineering will not succeed in altering the enzymes-binding specificities (Hwang and Warshel, 1988). We note the consistency between our results and those by Warshel and coworkers, even though the two groups use different methodologies to compute the electrostatic strengths of the ion pairs.

CONCLUSIONS

We have characterized the relationship between geometry and electrostatic strength of ion pairs in proteins. Most ion pairs with side-chain charged group centroids within 5 Å distance are likely to be stabilizing to the protein structure. These results may be useful in formulating guidelines for detecting stabilizing electrostatic interactions in proteins with known three-dimensional structures and for incorporating stabilizing electrostatic interactions in de novo protein design.

We thank Drs. Buyong Ma, Chung-jung Tsai, Neeti Sinha, Gunasekaran Kannan, David Zanuy, and in particular, Jacob V. Maizel for numerous helpful discussions. We also thank the two anonymous reviewers of our manuscript for their excellent suggestions. The research of R. Nussinov in Israel has been supported in part by a grant from the Israel Science Foundation administered by the Israel Academy of Sciences, by the Magnet grant, by the Ministry of Science grant, and by the Tel Aviv University Basic Research grants and by the Center of Excellence, administered by the Israel Academy of Sciences. S.K. dedicates this work to the loving memories of his grandfather, Shri I. C. Mangla.

This project has been funded in whole or in part with Federal funds from the National Cancer Institute, National Institutes of Health, under contract NO1-CO-12400. The content of this publication does not necessarily reflect the view or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organization imply endorsement by the U.S. Government.

REFERENCES

Avis, J. M., F. H. T. Allain, P. Howe, G. Varani, K. Nagai, and D. Neuhaus. 1996. Solution structure of the N-terminal Rnp domain of U1A protein: the role of C-terminal residues in structure stability and RNA Binding. *J. Mol. Biol.* 257:398–411.

de Bakker, P. I. W., P. H. Hunenberger, and J. A. McCammon. 1999. Molecular dynamics simulations of the hyperthermophilic protein Sac7d from *Sulfolobus acidocaldarius*: contribution of salt bridges to thermostability. *J. Mol. Biol.* 285:1811–1830.

Banci, L., I. Bertini, J. G. Huber, G. A. Spyroulias, and P. Turano. 1999. Solution structure of reduced horse heart cytochrome c. *J. Biol. Inorg. Chem.* 4:21–31.

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

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

Bellsolell, L., J. Prieto, L. Serrano, and M. Coll. 1994. Magnesium binding to the bacterial chemotaxis protein CheY results in large conformational changes involving its functional surface. *J. Mol. Biol.* 238:489–495.

Bernstein, F. C., T. F. Koetzle, G. J. B. Williams, E. F. Meyer, Jr., M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi. 1977. The protein data bank: a computer based archival file for macromolecular structures. *J. Mol. Biol.* 112:535–542.

Bewley, C. A., K. R. Gustafson, M. R. Boyd, D. G. Covell, A. Bax, G. M. Clore, and A. M. Gronenborn. 1998. Solution structure of cyanovirin-N, a potent HIV-inactivating protein. *Nat. Struct. Biol.* 5:571–578.

Cutler, R. L., A. M. Davies, S. Creighton, A. Warshel, G. R. Moore, M. Smith, and A. G. Mauk. 1989. Role of arginine-38 in regulation of the cytochrome c oxidation-reduction equilibrium. *Biochemistry.* 28: 3188–3197.

Dao-pin, S., D. E. Anderson, W. A. Baase, F. W. Dahlquist, and B. W. Matthews. 1991. Structural and thermodynamic consequences of burying a charged residue within the hydrophobic core of T4 lysozyme. *Biochemistry.* 30:11521–11529.

Elcock, A. H. 1998. The stability of salt bridges at high temperatures: implications for hyperthermophilic proteins. *J. Mol. Biol.* 284:489–502.

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

Gallagher, T., P. Alexander, P. Bryan, and G. L. Gilliland. 1994. Two crystal structures of the B1 immunoglobulin-binding domain of streptococcal protein G and comparison with NMR. *Biochemistry.* 33: 4721–4729.

Grimsley, G. R., K. L. Shaw, L. R. Fee, R. W. Alston, B. M. Huyghues-Despointes, R. L. Thurlkill, J. M. Scholtz, and C. N. Pace. 1999. Increasing protein stability by altering long-range coulombic interactions. *Protein Sci.* 8:1843–1849.

Gronenborn, A. M., D. R. Filpula, N. Z. Essig, A. Achari, M. Whitlow, P. T. Wingfield, and G. M. Clore. 1991. A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein G. *Science.* 253:657–661.

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

Honig, B., and A. Nicholls. 1995. Classical electrostatics in biology and chemistry. *Science.* 268:1144–1149.

Horovitz, A., and A. R. Fersht. 1992. Co-operative interactions during protein folding. *J. Mol. Biol.* 224:733–740.

Hwang, J. K., and A. Warshel. 1988. Why ion pair reversal by protein engineering is unlikely to succeed. *Nature.* 334:270–272.

Ippel, J. H., G. Larsson, G. Behravan, J. Zdunek, M. Schleucher, J. Lundqvist, P. O. Lycksell, and S. S. Wijmenga. 1999. The solution structure of the homeodomain of the rat insulin-gene enhancer protein Isl-1: comparison with other homeodomains. *J. Mol. Biol.* 288:689–703.

Jablonsky, M. J., P. L. Jackson, J. O. Trent, D. D. Watt, and N. R. Krishna. 1999. Solution structure of a β-neurotoxin from the new world scorpion *Centruroides sculpturatus* Ewing. *Biochem. Biophys. Res. Commun.* 254:406–412.

Kawamura, S., I. Tanaka, and M. Kimura. 1997. Contribution of a salt bridge to the thermostability of DNA binding protein HU from *Bacillus stearothermophilus* determined by site directed mutagenesis. *J. Biochem.* 121:448–455.

Kelly, L. A., S. P. Gardner, and M. J. Sutcliffe. 1996. An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally related subfamilies. *Protein Eng.* 9:1063–1065.

Kombo, D. C., M. A. Young, and D. L. Beveridge. 2000. One nanosecond molecular dynamics simulation of the N-terminal domain of the lambda repressor protein. *Biopolymers.* 53:596–605.

Kumar, S., and M. Bansal. 1998. Dissecting α-helices: position-specific analysis of α-helices in globular proteins. *Proteins.* 31:460–476.

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

Kumar, S., C. J. Tsai, and R. Nussinov. 2000a. Factors enhancing protein thermostability. *Protein Eng.* 13:179–191.

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

Kumar, S., and R. Nussinov. 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 R. Nussinov. 2001a. Ion pairs and their stabilities fluctuate in NMR conformer ensembles of proteins. *Proteins.* 43:433–454.

Kumar, S., and R. Nussinov. 2001b. How do thermophilic proteins deal with heat? *Cell. Mol. Life Sci.* 58:1216–1233.

Kumar, S., H. Wolfson, and R. Nussinov. 2001a. Protein flexibility and electrostatic interactions. *IBM J. Res. Dev.* 45:499–512.

Kumar, S., C. J. Tsai, and R. Nussinov. 2001b. Thermodynamic differences among homologous thermophilic and mesophilic proteins. *Biochemistry.* 40:14152–14165.

Kuszewski, J., A. M. Gronenborn, and G. M. Clore. 1999. Improving the packing and accuracy of NMR structures with a pseudopotential for the radius of gyration. *J. Am. Chem. Soc.* 121:2337–2338.

Lebbink, J. H. G., S. Knapp, J. van der Oost, D. Rice, R. Ladenstein, and W. M. de Vos. 1998. Engineering activity and stability of *Thermotoga maritima* glutamate dehydrogenase. I. Introduction of a six residue ion pair network in hinge region. *J. Mol. Biol.* 280:287–296.

Lebbink, J. H. G., S. Knapp, J. van der Oost, D. Rice, R. Ladenstein, and W. M. de Vos. 1999. Engineering activity and stability of *Thermotoga maritima* glutamate dehydrogenase. II. Construction of a 16 residue ion pair network at the subunit interface. *J. Mol. Biol.* 289:357–369.

Lounnas, V., and R. C. Wade. 1997. Exceptionally stable salt bridges in cytochrome P450cam have functional roles. *Biochemistry.* 36: 5402–5417.

Marqusee, S., and R. T. Sauer. 1994. Contribution of a hydrogen bond/salt bridge network to the stability of secondary and tertiary structures in lambda repressor. *Protein Sci.* 3:2217–2225.

Moy, F. J., D. F. Lowry, P. Matsumura, F. W. Dahlquist, J. E. Krywko, and P. J. Domaille. 1994. Assignments, secondary structure, global fold, and dynamics of chemotaxis Y protein using three-and four-dimensional heteronuclear 13C, 15N NMR spectroscopy. *Biochemistry.* 33: 10731–10742.

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

Nielsen, J. E., K. V. Anderson, B. Honig, R. W. W. Hooft, G. Klebe, G. Vriend, and R. C. Wade. 1999. Improving macromolecular electrostatics calculations. *Protein Eng.* 12:657–662.

Nilges, M., M. J. Macias, S. I. O'Donoghue, and H. Oschkinat. 1997. Automated NOESY interpretation with ambiguous distance restraints: the refined NMR solution structure of the pleckstrin homology domain from β-spectrin. *J. Mol. Biol.* 269:408–422.

Ogata, K., S. Morikawa, H. Nakamura, H. Hojo, S. Yoshimura, R. Zhang, S. Aimoto, Y. Ametani, Z. Hirata, A. Sarai, S. Ishii, and Y. Nishimura. 1995. Comparison of the free and DNA-complexed forms of the DNA-binding domain from C-Myb. *Nat. Struct. Biol.* 2:309–319.

Perez-Alvarado, G. C., J. L. Kosa, H. A. Louis, M. C. Beckerle, D. R. Winge, and M. F. Summers. 1996. Structure of the cysteine-rich intestinal protein, CRIP. *J. Mol. Biol.* 257:153–174.

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

Read, C. M., P. D. Cary, C. Crane-Robinson, P. C. Driscoll, and D. G. Norman. 1993. Solution structure of a DNA-binding domain from HMG1. *Nucleic Acids Res.* 21:3427–3436.

Schutz, C. N., and A. Warshel. 2001. What are the dielectric constants of proteins and how to validate electrostatic models? *Proteins.* 44: 400–417.

Singh, U. C. 1988. Probing the salt bridge in the dihydrofolate reductase-methotrexate complex by using the coordinate-coupled free energy perturbation method. *Proc. Natl. Acad. Sci. U.S.A.* 85:4280–4284.

Sham, Y. Y., I. Muegge, and A. Warshel. 1998. The effect of protein relaxation on charge-charge interactions and dielectric constants of proteins. *Biophys. J.* 74:1744–1753.

Sheinerman, F. B., R. Norel, and B. Honig. 2000. Electrostatic aspects of protein-protein interactions. *Curr. Opin. Struct. Biol.* 10:153–159.

Spector, S., M. Wang, S. A. Carp, J. Robblee, Z. S. Hendsch, R. Fairman, B. Tidor, and D. P. Raleigh. 2000. Rational modification of protein stability by mutation of charged surface residues. *Biochemistry.* 39: 872–879.

Spek, E. J., A. H. Bui, M. Lu, and N. R. Kallenbach. 1998. Surface salt bridges stabilize the GCN4 leucine zipper. *Protein Sci.* 11:2431–2437.

Sterner, R., and W. Liebl. 2001. Thermophilic adaptation of proteins. *Crit. Rev. Biochem. Mol. Biol.* 36:39–106.

Strop, P., and S. L. Mayo. 2000. Contribution of surface salt bridges to protein stability. *Biochemistry.* 39:1251–1255.

Volz, K., and P. Matsumura. 1991. Crystal structure of *Escherichia coli* CheY refined at 1.7-Å resolution. *J. Biol. Chem.* 266:15511–15519.

Waldburger, C. D., J. F. Schildbach, and R. T. Sauer. 1995. Are buried salt bridges important for protein stability and conformational specificity? *Nat. Struct. Biol.* 2:122–128.

Waldburger, C. D., T. Jonsson, and R. T. Sauer. 1996. Barriers to protein folding: formation of buried polar interactions is a slow step in acquisition of structure. *Biochemistry.* 93:2629–2634.

Warshel, A., and A. Papazyan. 1998. Electrostatic effects in macromolecules: fundamental concepts and practical modeling. *Curr. Opin. Struct. Biol.* 8:211–217.

Warshel, A., and S. T. Russell. 1984. Calculations of electrostatic interactions in biological systems and in solutions. *Q. Rev. Biophys.* 17: 283–422.

Xiao, L., and B. Honig. 1999. Electrostatic contributions to the stability of hyperthermophilic proteins. *J. Mol. Biol.* 289:1435–1444.

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

Xu, D., S. L. Lin, and R. Nussinov. 1997b. Protein binding versus protein folding: the role of hydrophilic bridges in protein associations. *J. Mol. Biol.* 265:68–84.

Yang, F., C. A. Bewley, J. M. Louis, K. R. Gustafson, M. R. Boyd, A. M. Gronenborn, G. M. Clore, and A. Wlodawer. 1999. Crystal structure of cyanovirin-N, a potent HIV-inactivating protein, shows unexpected domain swapping. *J. Mol. Biol.* 288:403–412.

Yip, K. S., T. J. Stillman, K. L. Britton, P. J. Artymiuk, P. J. Baker, S. E. Sedelnikova, P. C. Engel, A. Pasquo, R. Chiaraluce, V. Consalvi, R. Scandurra, and D. W. Rice. 1995. The structure of *Pyrococcus furiosus* glutamate dehydrogenase reveals a key role for ion-pair networks in maintaining enzyme stability at enzyme stability at extreme temperatures. *Structure.* 3:1147–1158.