All-Atom Contact Potential Approach to Protein Thermostability Analysis

Changjun Chen,[1,2] Lin Li,[1,2] Yi Xiao[1,2]

[1] *Biomolecular Physics and Modeling Group, Department of Physics, Huazhong University of Science and Technology, Wuhan 430074, Hubei, China*

[2] *The Abdus Salam International Center for Theoretical Physics, Strada Costiera 11, 34014 Trieste, Italy*

Received 9 October 2005; revised 27 August 2006; accepted 30 August 2006
Published online 8 September 2006 in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/bip.20600

ABSTRACT:

In this paper we use all-atom potential energy to define and analyze the inter-residue contacts in mesophilic and thermophilic proteins. Fifteen families of proteins are selected and each family has two representative proteins with greatly different preferred environmental temperatures. We find that both the number and energy of the contacts defined in this way show stronger correlations with the preferred temperatures of proteins than other factors used before. We also find that the charged-polar and charged-nonpolar residue contacts not only have larger contact numbers but also have lower single contact energies. Furthermore, the most important is that most of the thermophilic proteins have more charged-polar and charged-nonpolar residue contacts than their mesophilic counterparts. This suggests that they may play an important role in the thermostability of proteins, except usual charged–charged and nonpolar–nonpolar residue contacts. Charged residues may exert their profound influence by forming contacts not only with other charged residues but also with polar or nonpolar residues, thus further increasing the strength of contact network and then the thermostability of proteins. © *2006 Wiley Periodicals, Inc. Biopolymers 85: 28–37, 2007.*

Keywords: *protein thermostability; residue–residue contacts; all-atom potential energy*

Correspondence to: Yi Xiao; e-mail: yxiao@mail.hust.edu.cn
Contract grant sponsor: NSFC
Contract grant numbers: 30525037, 30470412
Contract grant sponsor: Foundation of the Ministry of China

This article was originally published online as an accepted preprint. The "Published Online" date corresponds to the preprint version. You can request a copy of the preprint by emailing the Biopolymers editorial office at biopolymers@wiley.com

INTRODUCTION

It is known that every protein has its preferred temperature. When the environmental temperature is too higher or lower than its preferred temperature, a protein will lose its stability and then its function. So studying the relationship between protein structures and their preferred temperatures can really help us understand the mechanism of protein thermostability.

In the past, many works have been done in this field[1,2] and some physical characters were regarded as the keys to the thermostability of proteins, such as hydrogen bonds and fractional polar surface,[3,4] hydrophobicity,[5,6] electrostatic interactions,[7–12] salt bridges,[13–16] and contact density.[17]

Generally it is believed that the electrostatic interaction (include ion pairs and hydrogen bonds) and the hydrophobic interaction (nonpolar residue packing) in the proteins are the most important factors to protein thermostability (as well reviewed in[18–22]). For the electrostatic interaction, Kumar et al. have compared 18 nonredundant protein families of mesophilic and thermophilic proteins. They proposed that salt bridges and side-chain hydrogen bonds greatly stabilize the proteins in high temperatures.[15] As a detailed study, Zhou and Dong checked the thermophilic *Bacillus caldolyticus* cold shock protein (Bc-Csp) and the mesophilic *Bacillus subtilis* cold shock protein B (Bs-CspB), and they are different in the sequence for 11 of the 66 residues. By analyzing electrostatic effects in both the folded and unfolded states, they noted that Bc-Csp has more unfavorable charge–charge interactions in the unfolded state, which would inversely increase its thermostablility.[11]

For the hydrophobic interaction, Gromiha selected a set of proteins from different families which have the highest

and lowest average environmental temperatures.[23] After analyzing the preference of surrounding residues influenced by medium- and long-range contacts, he found that the thermophilic proteins have more long-range contacts at residue distance 31–34 in their sequences than the mesophilic proteins. He also observed that hydrophobic and polar–nonpolar contacts are the preferred residue pairs that enhance the thermostability of thermophilic proteins.

Recently, Berezovsky and Shakhnovich[24] made a detailed analysis on five groups of proteins with different environment temperatures. They assumed that proteins enhance their thermostability mainly based on two strategies: one is structure-based, proteins increase the number of contacts to become more stable. The other is sequence-based, proteins mutate the special amino acids and strengthen some special interactions to become more stable. They also declared that the choice of a certain way for each protein to enhance thermostability is determined under the evolutionary history and environmental conditions. Ancient organisms adapt the structure-based way, and later evolved organisms often adapt the sequence-based way.

In this paper, we use contact analysis to investigate the inter-residue interactions between thermophilic and mesophilic proteins and try to find the main factors and ways that enhance the thermostability by using a different way. Although up to now many factors have been put out to evaluate the thermostability of proteins, such as hydrogen bonds, ion pairs, nonpolar surface area, and contact number, all of them are purely geometry-based, and may lead to greater statistical errors (exemplified in the Discussion section). On the other hand, our criterion, the contacts based on all-atom force field, has more physical characters. Through analyzing the protein data, we find both the normalized contact number and contact energy strongly correlate with the thermostability of proteins. So this criterion would be a good benchmark. Additionally, we also study various kinds of contacts, like hydrophobic contacts, polar contacts, half-charged contacts, and charged contacts (see definition in Discussion section) and extract some important features of special contacts to protein thermostability.

MATERIALS AND METHODS

It is well known that the stability of a protein is determined by the inter-atomic interactions in it. But because of the large number of atom types, the interactions between atoms in protein are complex.[25–27] It is not easy to draw out the main features. So many statistical residue–residue contact potentials have been proposed to describe the characters of protein structures.[28] Among these, the most famous one is MJ contact potentials,[29,30] which are derived from a lot of crystal structures of globular proteins through the quasi-chemical approximation. From these contact energies, some inter-residue characters can be identified: firstly, the formation of Cys-X contacts from Cys-Cys and X-X contacts always cause a relative large energy loss; secondly, the strong attraction exists between the charged residues (positive Lys, Arg and negative Glu, Asp), and thirdly, the segregation of hydrophobic and hydrophilic residues is observed. The MJ matrix was validated by some experiments.[31] They also showed that the contact energies can well discriminate native conformations from other decoy folds. Recently, the studies on contacts go on more detail, such as integrated radial distribution,[32] long range interactions,[33] secondary characters,[34] and solvent influence.[35]

Although statistical contact potential is helpful in determining the crucial features of residue–residue interactions in proteins, but it does have its limitations. Recently Khatun et al.[36] derived a new set of contact potentials by optimizing the parameters with experimental $\Delta\Delta G$ (stability changes upon mutations) using singular value decomposition algorithm. The contact potentials are applied in predicting $\Delta\Delta G$ for other proteins. The results of prediction showed that they are not accurate and transferable within experimental error. So we still need effort to develop the efficiency of contact potentials in predicting the changes of stability of proteins.

In this paper, we try to calculate contact potentials with all-atom force field. We use GB/SA model[37,38] as implicit solvent model to simulate the aqueous environment. GB/SA is a reduced model from the continuum model, which treats the water as continuous medium and there are usually three terms included in the free energy of solvation:

$$\Delta G_{\text{sol}} = \Delta G_{\text{cav}} + \Delta G_{\text{vdw}} + \Delta G_{\text{pol}} \tag{1}$$

where ΔG_{cav} is a solvent-solvent cavity term, corresponding to the free energy of creating a cavity of solute in the solvent continuum; ΔG_{vdw} is the free energy term representing the interactions between the solute and solvent. And the last term ΔG_{pol} denotes electrostatic interactions between the solute and solvent. The advantage of this model is that it does not need treat solvent molecules explicitly and cost much less computation time.

The sum of the first two terms in Eq. (1) is often regarded proportional to the solvent-accessible surface area of the solute.

$$\Delta G_{\text{cav}} + \Delta G_{\text{vdw}} = \sum_{i=1}^{N} \sigma_i A_i \tag{2}$$

where A_i is the solvent-accessible surface area of the atom i and σ_i is an special empirical parameter corresponding to atom i. Generally all σ_i is 4.9 cal/(mol/Å^2) and solvent probe radius is 1.4 Å.

The occurrence of the last term, in Eq. (1), ΔG_{pol}, is due to the polarization of the solvent which is caused by the solute. The charge distribution of the solute directly determines that of the solvent, which in turn influences the solute reversely.

To obtain ΔG_{pol}, the most precise method is to solve the Poisson–Boltzmann (PB) equation, the result of which is very close to that of explicit water. However, it is still too slow to be applicable in normal molecular dynamics simulations. Recently some numerical methods related to solving the PB equation have been published, which is promising to be used widely.

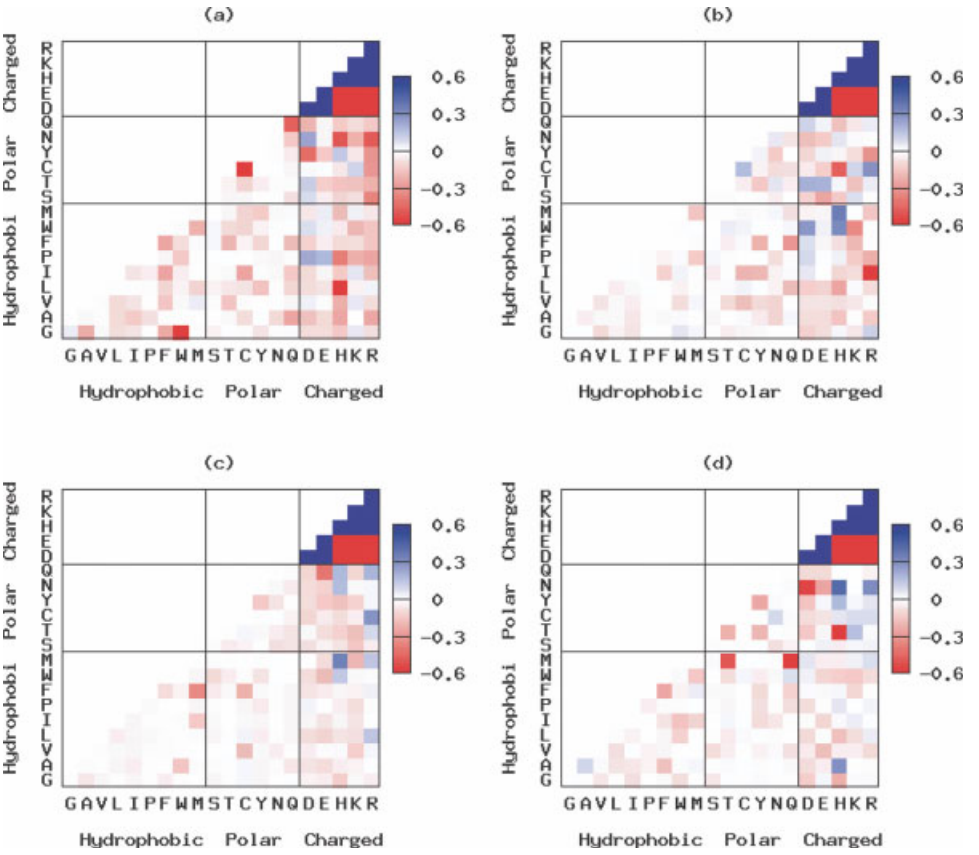

FIGURE 1 Averaged electrostatic energies between all residue pairs for the proteins (a) 1ash, (b) 1ino, (c) 1nar, and (d) 1pex. The residue types from 1 to 20 are sorted in the order: GLY, ALA, VAL, LEU, ILE, PRO, PHE, TRP, MET, SER, THR, CYS, TYR, ASN, GLN, ASP, GLU, HIS, LYS, ARG. The energy unit is kcal/mol.

Another attractive approach to calculate ΔG_{pol} in Eq. (1) is to use generalized Born (GB) model proposed by Still et al.[37] and developed by others.[38] This model calculates ΔG_{pol} as follows:

$$\Delta G_{\text{pol}} = -166.0\left(1-\frac{1}{\varepsilon}\right)\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{q_i q_j}{\sqrt{r_{ij}^2+\alpha_i\alpha_j\, e^{-D_{ij}}}} \qquad (3)$$

where $D_{ij} = r_{ij}^2/4\alpha_i\alpha_j$, r_{ij} is the distance between atom i and atom j. q_i and q_j are the charges of atom i and atom j. ε is the dielectric constant of the solvent. The most important, α_i is the effective Born radius of atom i, which is related to the effective Born free energy of solvation.

In the following discussion we would analyze the interaction energies between residue pairs upon all-atom force field. Although it costs too much time on the computation than previous published methods, such as contacts between C_α atoms, it is precise and can produce some new information which is helpful on the study of the thermostability of the proteins. The software we use is Tinker (See: http://dasher.wustl.edu/tinker/) with Charmm27 force field.[39] Before formal analysis we optimize all the structures with the conjugate-gradient method and the gradient tolerance is 0.2 kcal/(Å mol).

It is necessary to note that the analysis presented in this paper examines the contact and solvation energies of the folded protein conformations only. Strictly, it is important to include an unfolded reference state in our protein stability calculations. However, since an unfolded reference state model is still absent in all-atom level, it is difficult to consider the effect of unfolded states in our calculation at this stage. Fortunately, since mesophilic and thermophilic proteins in the same family have similar sequences, they should have similar unfolded energies. Therefore, at least as a first order approximation, we may assume that these unfolded energies will cancel out of any analysis regarding the relative stability of these homologous proteins. Of course, electrostatic desolvation may be sensitive to these small differences in sequence, particularly since these differences results may result from mutations that substitute charged residues for polar residues. We shall try to consider the unfolded state effect in our future works.

RESULTS AND DISCUSSION

Contact Potential Energies

To get an overall view on the contact potential energies calculated by all-atom force field, we arbitrarily select tens of pro-

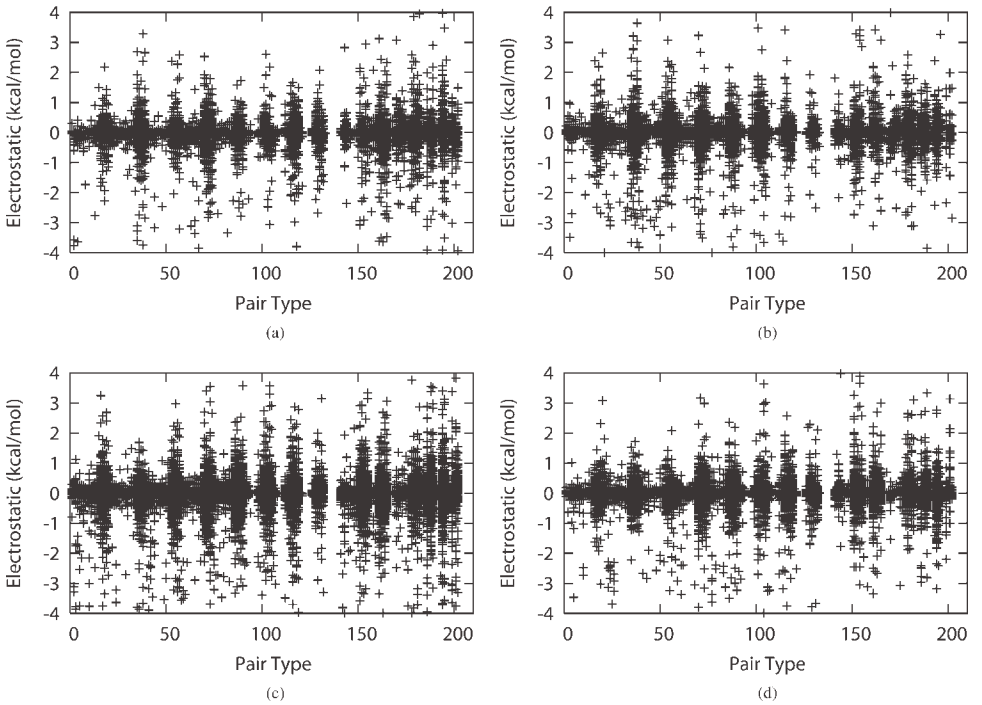

FIGURE 2 A one-dimensional view of electrostatic energies (−10 kcal/mol to + 10 kcal/mol) between all residues pairs for the proteins: (a) 1ash, (b) 1ino, (c) 1nar, and (d) 1pex. To evaluate the residue–residue interaction, we first sort the 20 amino acids from 1 to 20 as follows: GLY, ALA, VAL, LEU, ILE, PRO, PHE, TRP, MET, SER, THR, CYS, TYR, ASN, GLN, ASP, GLU, HIS, LYS, ARG. Then we get 210 pairs one by one in the order: GLY-ALA, GLY-VAL, . . . , ALA-VAL, ALA-LEU, ... The energy unit is kcal/mol.

teins from Brookhaven Data Bank[40,41] which belong to four different structure classes: all α, $\alpha + \beta$, α/β, and all β. For simplicity, in the following we only show results of one protein for each class. They are hemoglobin (1ash), inorganic pyrophosphatase (1ino), narbonin (1nar), and collagenase-3(1pex), respectively.

We first calculated the averaged electrostatic energies of all residue pairs (Figure 1). We found that, the strongest attractive and repulsive interactions are between the charged residues (positive LYS, HIS, ARG and negative GLU, ASP), the second strongest interactions are those corresponding to the charged-polar pairs or charged-hydrophobic pairs. We call both of them half-charged pairs. This can be seen more clearly from Figure 2, which shows the detailed inter-residue electrostatic energies in one dimension (all the 20 amino acids are sorted as follows: GLY, ALA, VAL, LEU, ILE, PRO, PHE, TRP, MET, SER, THR, CYS, TYR, ASN, GLN, ASP, GLU, HIS, LYS, ARG. Then we get 210 pairs one by one in the order: GLY-ALA, GLY-VAL, . . . , ALA-VAL, ALA-LEU, . . .).

The bands in Figure 2 correspond to the half-charged pairs. To understand this phenomenon, we plot in Figure 3 the dependence of the electrostatic energy on the centroid distance between charged residue GLU and polar residue THR. It can be seen that the energies are close to zero when the distance between them is far away, but they diverge from zero when they approach to each other. So unlike charged residue pairs, their energies are not a single-value function of their distances; for example, residue pair GLU7-THR227 and THR201-GLU207 in the protein 1nar: their electrostatic energies are −1.60 and 1.84 kcal/mol respectively, while their centroid distances are 9.67 and 9.41 Å (almost at the same distance). This phenomenon may be due to two reasons. The first is the orientation of the side chains of residue pairs. Different orientations of side chains cause different distances between detailed atom pairs and then certainly affects the electrostatic energy between residues. The second factor is the burial degree and therefore solvation effect. The closer the residue pairs near the protein surface, the stronger the solvent would influence the inter-residue interaction. Because of the large number, this half-charged pairs are more important than the charged-charged pairs on the whole. We'll discuss it in the next section.

Figure 4 gives the averaged potential energies (including three energy terms: Van der Waals energy, electrostatic energy, and solvent polarized energy) of all residue–residue pairs for

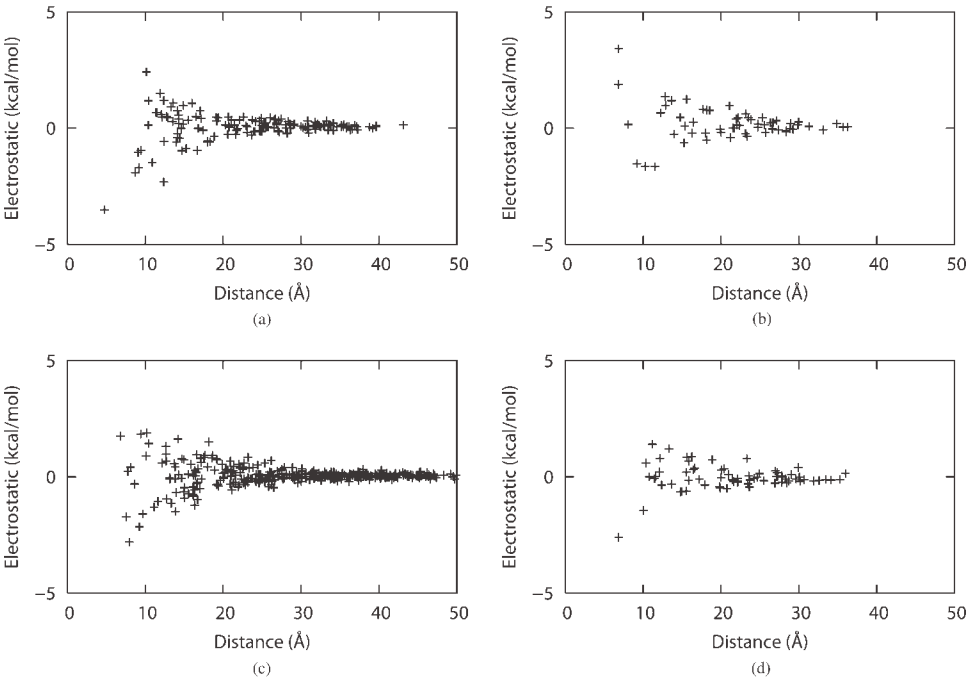

FIGURE 3 Electrostatic energies vs. centroid distance between polar residue THR and charged residue GLU for protein (a) 1ash, (b) 1ino, (c) 1nar, and (d) 1pex. The energy unit is kcal/mol.

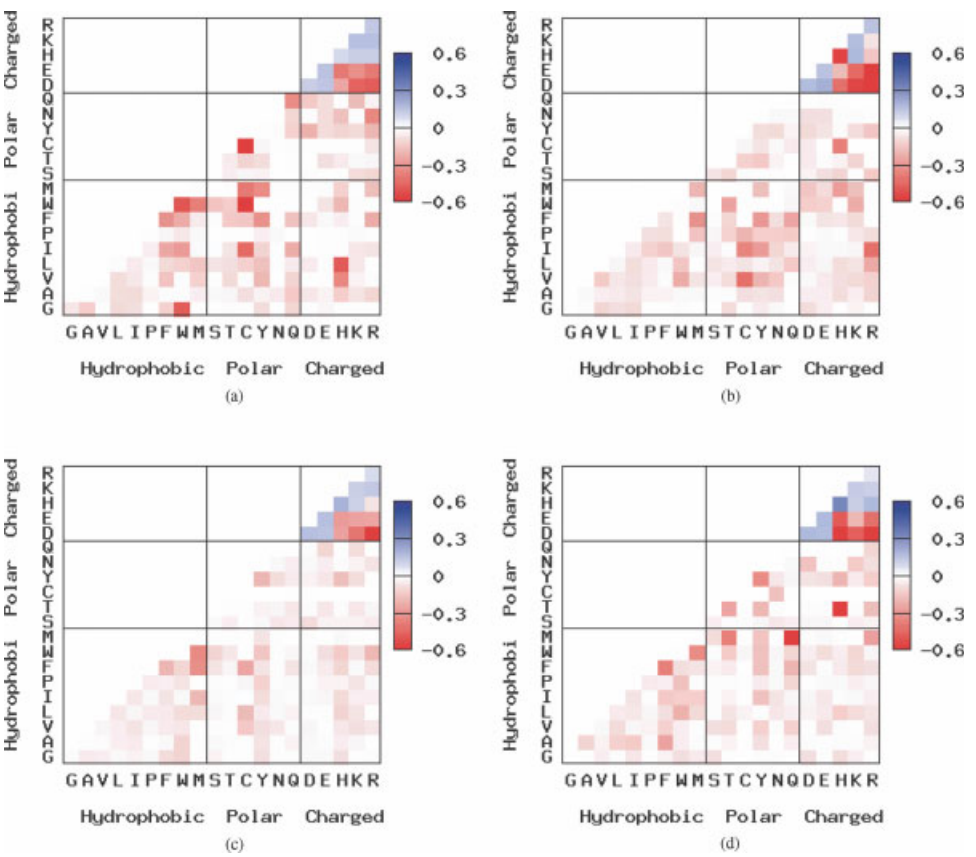

FIGURE 4 Averaged potential energies between all residues pairs for the proteins: (a) 1ash, (b) 1ino, (c) 1nar, and (d) 1pex. See detailed description for residue pairs in Figure 1.

Table I The Calculated Contact Parameters for the Mesophilic and Thermophilic Proteins

			Hydrophobic Contacts		Polar Contacts		Half-charged Contacts		Charged Contacts		Total			
Family	PDB id (L)	T_{env} (°C)	N	E	N	E	N	E	N	E	N_T	E_T	E_{av}	C.D. (%)
Malate dehydrogenase	4MDH (334)	37.0	254	−432.30	205	−435.15	221	−608.87	71	−309.44	751	−1785.76	−2.38	1.39
	1BMD (327)	72.5	311	−519.00	189	−407.93	209	−600.90	66	−270.61	775	−1798.44	−2.32	1.50
Lactate dehydrogenase	6LDH (330)	20.0	242	−409.98	207	−438.41	198	−596.46	62	−275.70	709	−1720.54	−2.43	1.35
	1LDN (316)	52.5	269	−455.79	176	−372.57	207	−590.36	60	−390.09	712	−1727.81	−2.43	1.48
Ribonuclease H	2RN2 (155)	37.0	102	−192.32	116	−237.38	93	−277.57	34	−192.76	345	−900.02	−2.61	3.09
	1RIL (147)	72.5	117	−224.18	67	−122.46	98	−274.85	44	−248.43	326	−869.93	−2.67	3.26
Subtilisin	1ST3 (269)	30.0	272	−417.54	265	−595.68	163	−520.05	18	−118.73	718	−1652.02	−2.30	2.07
	1THM (279)	60.0	291	−438.02	291	−636.23	161	−496.19	25	−200.97	768	−1771.42	−2.31	2.05
Superoxide dismutase	3SDP (183)	27.5	102	−172.75	132	−230.92	74	−177.76	18	−93.02	326	−674.46	−2.07	2.07
	3MDS (200)	72.5	145	−250.78	105	−245.82	126	−361.47	35	−130.73	411	−988.80	−2.41	2.17
Phosphofructokinase	2PFK (305)	37.0	276	−486.43	181	−369.27	186	−529.77	64	−310.04	707	−1695.52	−2.40	1.58
	3PFK (319)	52.5	282	−471.43	192	−367.91	197	−575.80	86	−378.09	757	−1793.22	−2.37	1.54
Phosphoglycerate kinase	3PGK (416)	27.5	284	−426.39	188	−320.60	243	−565.50	85	−270.43	800	−1582.92	−1.98	0.95
	1PHP (394)	52.5	390	−627.83	186	−395.02	276	−708.40	115	−624.20	967	−2355.45	−2.44	1.28
Triose phosphate isomerase	1YPI (247)	27.5	200	−321.76	135	−260.97	147	−407.70	58	−300.15	540	−1290.58	−2.39	1.85
	1BTM (251)	52.5	224	−362.71	155	−320.68	157	−436.15	42	−239.28	578	−1358.82	−2.35	1.92
Hydrolase	1INO (175)	37.0	133	−210.14	106	−226.54	129	−325.68	48	−232.53	416	−994.88	−2.39	2.90
	2PRD (174)	72.5	149	−255.66	68	−142.07	146	−424.39	51	−157.51	414	−979.63	−2.37	2.92
Glycosyltransferase B	2EXO (312)	30.0	244	−407.65	185	−350.96	288	−816.17	73	−445.06	790	−2019.84	−2.56	1.68
(β-glycanase)	1XYZ (320)	60.0	237	−417.99	273	−588.64	242	−721.41	68	−432.65	820	−2160.70	−2.63	1.66
Citrate synthase	1CSH (435)	37	306	−489.09	284	−560.16	266	−753.01	78	−297.99	934	−2100.26	−2.25	1.01
	1AJ8 (371)	100	278	−421.34	219	−427.48	243	−668.08	90	−297.24	830	−1814.15	−2.19	1.23
Rubredoxin	8RXN (52)	35.5	25	−35.97	38	−71.66	28	−77.72	8	−9.65	99	−195.00	−2.0	9.16
	1CAA (53)	100	26	−43.61	34	−63.92	30	−87.30	15	−39.36	105	−234.19	−2.23	9.31
Elongation factor	1EFU (364)	37	304	−554.13	211	−462.21	296	−772.83	77	−312.55	888	−2101.73	−2.37	1.38
	1EFT (403)	71	343	−601.12	208	−433.57	353	−985.75	127	−576.38	1031	−2596.82	−2.52	1.31
Glutamate dehydrogenase	1HRD (449)	33.5	386	−603.41	300	−625.14	307	−940.11	86	−433.40	1079	−2602.05	−2.41	1.10
	1GTM (417)	87.5	377	−624.84	257	−536.91	303	−883.93	113	−640.80	1050	−2686.48	−2.56	1.24
Neutral protease	1NPC (315)	30	168	−241.12	305	−561.32	200	−499.41	48	−222.50	721	−1524.35	−2.11	1.51
	1LNF (313)	52.5	183	−318.66	315	−646.09	202	−592.98	51	−257.90	751	−1815.63	−2.42	1.59

Notes: T_{env}, average environmental temperature; the data are selected from Refs. 3, 15, and 42; L, length of protein sequence; N and E, contact number and energy; N_T and E_T, the total contact number and energy of a protein; E_{av}, averaged contact energy = E_T/N_T; C.D., contact density = the total contact number/the total number of residue pairs.

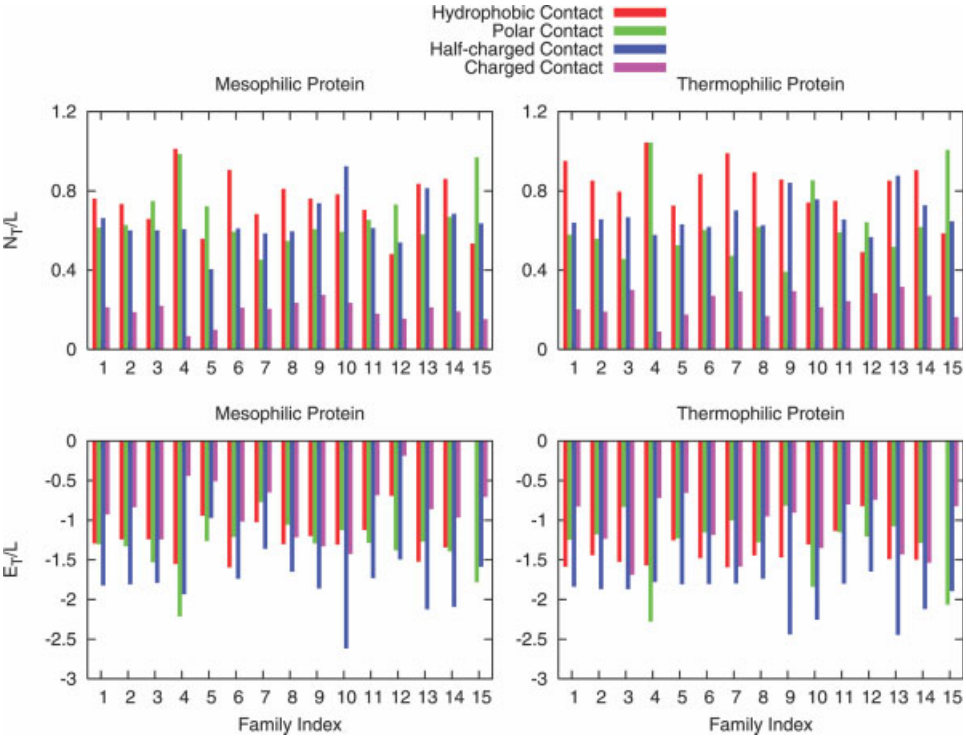

FIGURE 5 Normalized contact number (N/L) and normalized contact energy (E/L) for four kinds of contacts in fifteen mesophilic proteins and fifteen thermophilic proteins (the data are from Table I). L is the amino acid number of a protein.

the four proteins. We find that the overall patterns have changed greatly (compare to Figure 1), especially for charged–charged residue pairs. This is because the electrostatic energies between them are greatly moderated by the polar effect due to the solvent. So the averaged interactions of most pairs are very small (between +0.6 and −0.6 kcal/mol). But this is only the averaged values. For any actual residue pairs, the potential energy varies greatly. Some of them are around zero, no matter how close they are. Some of them are much lower (−10 kcal/mol) even they are comparably far away from each other. This may explain why previous statistical contact potentials are not so accurate. Their energies only depend on the centroid distances between residues, which is not very consistent with real situation.

The inter-residue potential energy is the base on which we define contacts. We assume a contact is formed when the potential energy between two residues is lower then −0.5 kcal/mol (it approximately equal to define a contact when the distance between residues is 7.5 Å).

Contacts in Mesophilic and Thermophilic Proteins

Now we apply the contact analysis to the mesophilic and thermophilic proteins. The database we use is from those used by Vogt and Argos,[3] Kumar et al.[15] and Gromiha et al.[42] There are fifteen groups of proteins. Each group represents a protein family and has two proteins with great different preferred environmental temperature (Table I). The proteins in the same family have the similar sequences and structures.

We calculate the contacts for the proteins and divide them into four categories according to the properties of their amino acid components, which are hydrophobic contacts (nonpolar–nonpolar residue contacts), polar contacts (polar–nonpolar and polar–polar residue contacts), half-charged contacts (charged-nonpolar and charged-polar residue contacts) and charged contacts (charged–charged residue contacts). We are interested in the contribution of different types of contacts to the thermostability of proteins. If one type of contact has larger number or lower energy in thermophilic proteins than in their mesophilic counterparts, it should contribute to the thermostability of protein.

To understand the data in Table I, we visualize them in different ways in Figures 5–7. From them, we observed some interesting results

1. In Figure 5 we present the normalized contact numbers (N/L) and the normalized contact energies (E/L) of the four kinds of contacts in the fifteen mesophilic proteins

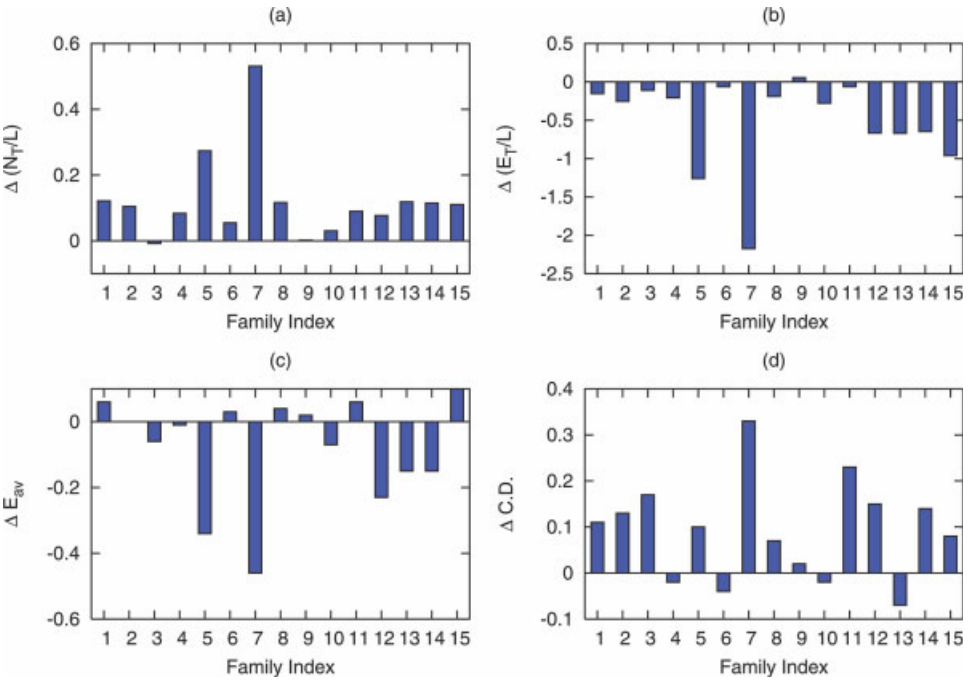

FIGURE 6 The differences of the normalized total contact number $\Delta(N_T/L)$ (a) normalized total contact energy $\Delta(E_T/L)$, (b) averaged contact energy ΔE_{av}, (c) and contact density ΔC.D., and (d) between thermophilic and mesophilic proteins in each of the fifteen protein groups (the data are from Table I).

and fifteen thermophilic proteins. L is the sequence length of a protein.

It is generally accepted that charged contacts and hydrophobic contacts are the most important factors to the stability of proteins.[5,6,13–15] Charged contacts provide the strongest forces while the hydrophobic contacts make stable hydrophobic cores. However, Figure 5 shows that the half-charged contacts have the lowest normalized potential energies in most cases (12 of 15 in mesophilic proteins, 13 of 15 in thermophilic proteins). This can be explained in two ways: on one hand, the number of charged contacts is usually much less than those of other three types of contacts (the top two figures in Figure 5), even if generally one single charged contact has lowest energy. On the other hand, although the hydrophobic contacts have the largest contact number in most cases (10 of 15 in mesophilic proteins, 11 of 15 in thermophilic proteins), the potential energy of one hydrophobic contact is generally higher than those of other kinds of contacts. In contrast, the half-charged contact not only has large contact number but also has lower single contact energy than that of hydrophobic contact or polar contact. In 24 of 30 proteins, the number of the half-charged contacts is the largest or second largest in all of the four kinds of contacts.

These results suggest that the half-charged interactions may also be the important interaction to the thermostability of proteins. Charged residues may exert their profound influence not only by forming contacts with other charged residues but also with polar or nonpolar residues. The favorable interaction between a charged residue and a nonpolar residue may be resulted from two aspects. The first is that even the charged residues have some nonpolar fragments in the side-chains. This would favor the aggregation between charged-nonpolar residue pairs. The second reason may still be the electrostatic interaction. In the definition of Charmm27 force field, all the atoms in the nonpolar residues are weak electric. So in fact even the nonpolar residues would have electrostatic interaction with charged residues.

2. Figure 6 shows the differences of the normalized contact number ($\Delta(N_T/L)$), normalized contact energy ($\Delta(E_T/L)$), averaged contact energy (ΔE_{av}), and contact density (ΔC.D.) between the thermophilic and mesophilic proteins in each of the fifteen protein families, respectively.

We can see from Figures 6a and 6b that most of the thermophilic proteins have larger normalized contact number and lower normalized contact energies than their

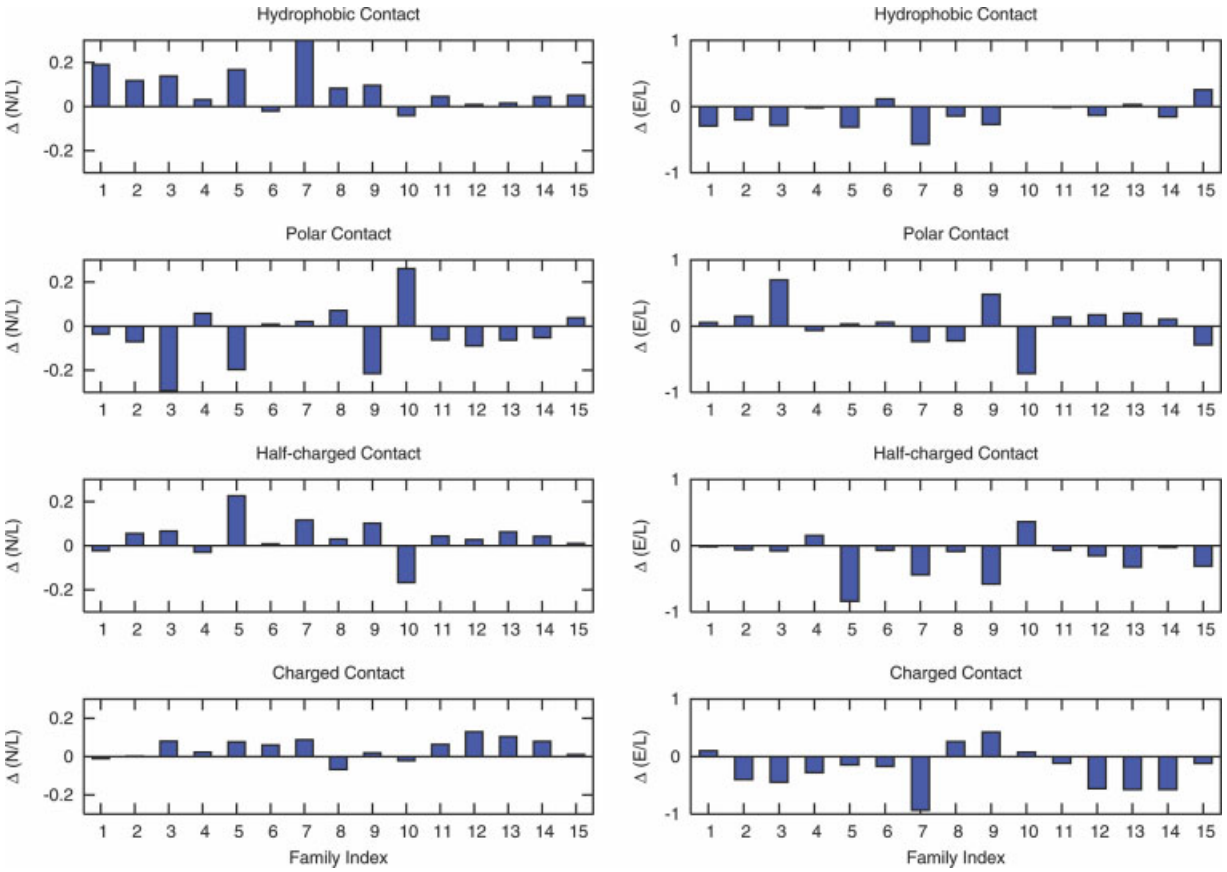

FIGURE 7 The differences of the normalized contact numbers $\Delta(N/L)$ and energies $\Delta(E/L)$ of the four kinds of contacts between the thermophilic and mesophilic proteins in the fifteen protein families (the data are from Table I).

mesophilic counterparts (13 out of 15 proteins). This is reasonable since it make the thermophilic proteins more stable. This indicates that the normalized contact number and energies are strongly correlated with the thermostability of proteins and are good indices to be used to predict the relative stability of two proteins in practice. This also means that the contacts defined by all-atom potential are more suitable to the description of protein stability. It is different from other famous criterions, such as hydrogen bonds, nonpolar surface area, and hydrophobic core and ion pairs. All of them depend on the geometry of the protein directly, while ours involves the all-atom force field. Although it is more complicated, it can effectively characterize the residue–residues interactions.

3. Figure 7 shows that only one kind of contact correlates strongly with temperature: the hydrophobic contact. In 13 of 15 groups thermophilic proteins have more hydrophobic contacts than their mesophilic counterparts. Thus the hydrophobic interaction may be one of the most important factors to determine the thermostability of proteins. This is in agreement with other studies.[5,6,23] Furthermore, we found that in most cases thermophilic proteins also have more half-charged and charged contacts and lower half-charged and charged contact energies than mesophilic proteins. This suggests that these two kinds of contacts may be also important to the thermostability of proteins. This is confirmed to the conclusions of Tanner et al.[43] They studied the determinants of thermostability for GAPDH in the extreme thermophile *Thermus aquaticus*, and found that the number of hydrogen bonds between charged side chains and neutral partners can greatly enhanced the thermostability of proteins. An exception is the thermophilic protein (1xyz) in the 10th family. It enhances its thermostability by greatly increasing the number of polar contacts but not hydrophobic and charged-related contacts. However, we found that the polar contacts have no significant correlation to the thermal stability of proteins.

In conclusion, we proposed a new criterion to analyze the thermostability of proteins: the contacts based on all-atom force field. We find that almost all the thermophilic proteins

have larger normalized contact number and lower normalized contact energies than their mesophilic counterparts. This indicates a strong correlation of these two variables and thermostability. It can be used in practice to predict the relative stability of two proteins.

To increase the thermostability of proteins, they need to form more compact contact networks. There are two way to do this. One is increasing contact number or contact density and another is substituting weaker contacts with stronger ones. It is generally accepted that the former is to increase the number of hydrophobic contacts and the latter is to replace weaker contacts with charged ones. Our results confirm this but further show that the half-charged contacts can play both roles: increasing contact number and enhancing contact strength or reduce contact energy. The charged residues exert their profound influence not only by forming contacts with other charged residues but also with polar or nonpolar residues, thus further increasing the strength of contact network and thermostability of proteins. Generally, whether increasing hydrophobic contacts or not, increasing the number of charged-related contacts (charged or/and half-charged contacts) is always an efficient way to increase the thermostability of proteins. We expect that our results will be helpful to understand the mechanism of enhancing thermostability of proteins and furthermore, predict the thermostability of proteins.

REFERENCES

1. Jaenicke, R.; Bohm, G. Curr Opin Struct Biol 1998, 8, 738–748.
2. Ladenstein, R.; Antranikian, G. Adv Biochem Eng Biotechnol 1998, 61, 37–85.
3. Vogt, G.; Argos, P. Fold Des 1997, 2, 840–846.
4. Vogt, G.; Woell, S.; Argos, P. J Mol Biol 1997, 269, 631–643.
5. Haney, P.; Konisky, J.; Koretke, K. K.; Luthey-Schulten, Z.; Wolynes, P. G. Proteins: Struct Funct Genet 1997, 28, 117–130.
6. Salminen, T.; Teplyakov, A.; Kankare, J.; Cooperman, B. S.; Lahti, R.; Goldman, A. Protein Sci 1996, 5, 1014–1025.
7. Xiao, L.; Honig, B. J Mol Biol 1999, 289, 1435–1444.
8. Szilagyi, A.; Zavodszky, P. Structure 2000, 8, 493–504.
9. Petsko, G. A. Methods Enzymol 2001, 334, 469–478.
10. Zhou, H. X. Biochemistry 2002, 41, 6533–6538.
11. Zhou, H. X.; Dong, F. Biophys J 2003, 84, 2216–2222.
12. Dominy, B. N.; Perl, D.; Schmid, F. X.; Brooks, C. L., III. J Mol Biol 2002, 319, 541–554.
13. Elcock, A. H. J Mol Biol 1998, 284, 489–502.
14. Querol, E.; Perez-Pons, J. A.; Mozo-Villarias, A. Protein Eng 1996, 9, 265–271.
15. Kumar, S.; Tsai, C. J.; Nussinov, R. Protein Eng 2000, 13, 179–191.
16. Das, R.; Gerstein, M. Funct Integr Genomics 2000, 1, 76–88.
17. England, J. L.; Shakhnovich, B. E.; Shakhnovich, E. I. Proc Natl Acad Sci USA 2003, 100, 8727–8731.
18. Mozo-Villiarías, A.; Querol, E. Curr Bioinformatics 2006, 1, 25–32.
19. Colacino, F.; Crichton, R. R. Biotechnol Genet Eng Rev 1997, 14, 211–277.
20. Russell, R. J. M.; Taylor, G. L. Curr Opin Biotechnol 1995, 6, 370–374.
21. Scandurra, R.; Consalvi, V.; Chiaraluce, R.; Politi, L.; Engel, P. C. Biochimie 1998, 80, 933–941.
22. Vieille, C.; Burdette, D. S.; Zeikus, J. G. Biotechnol Annu Rev 1996, 2, 1–83.
23. Gromiha, M. M. Biophys Chem 2001, 91, 71–77.
24. Berezovsky, I. N.; Shakhnovich, E. I. Proc Natl Acad Sci USA 2005, 102, 12742–12747.
25. Rose, G.; Geselowitz, A.; Lesser, G.; Lee, R.; Zehfus, M. Science 1985, 229, 834–838.
26. Cootes, A. P.; Curmi, P. M. G.; Cunningham, R.; Donnelly, C.; Torda, A. E. Proteins: Struct Funct Genet 1998, 32, 175–189.
27. Vendruscolo, M.; Najmanovich, R.; Domany, E. Phys Rev Lett 1999, 82, 656–659.
28. Gromiha, M. M.; Selvaraj, S. Biophys Mol Biol 2004, l86, 235–277.
29. Miyazawa, S.; Jernigan, R. L. Macromolecules 1985, 18, 534–552.
30. Miyazawa, S.; Jernigan, R. L. J Mol Biol 1996, 256, 623–644.
31. Sharp, K. A.; Friedman, A. R.; Honig, B. Biochemistry 1991, 30, 9686–9697.
32. Bahar, I.; Jernigan, R. L. J Mol Biol 1997, 266, 195–214.
33. Gromiha, M. M.; Selvaraj, S. Biophys Chem 1999, 77, 49–68.
34. Zhang, C.; Kim, S. H. Proc Natl Acad Sci USA 2000, 97, 2550–2555.
35. Chelli, R.; Gervasio, F. L.; Procacci, P.; Schettino, V. Proteins: Struct Funct Genet 2004, 55, 139–151.
36. Khatun, J.; Khare, S. D.; Dokholyan, N. V. J Mol Biol 2004, 336, 1223–1238.
37. Still, V. C.; Tempezvk, A.; Hawley, R. C.; Hendrickson, T. J Am Chem Soc 1990, 112, 6127–6129.
38. Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. J Phys Chem A 1997, 101, 3005–3014.
39. MacKerell, A. D., Jr.; Bashford, D.; Bellott, R. L.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. J Phys Chem B 1998, 102(18), 3586–3616.
40. Bernstein, F.; Koetzel, T.; Williams, G.; Meyer, E.; Brice, M.; Rodgers, J.; Kennard, O.; Shimanouchi, T.; Tasumi, M. J Mol Biol 1977, 112, 535–542.
41. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. Nucleic Acids Res 2000, 28, 235–242.
42. Gromiha, M. M.; Oobatake, M.; Sarai, A. Biophys Chem 1999, 82, 51–67.
43. Tanner, J.; Hecht, R.; Krause, K. Biochemistry 1996, 35, 2597–2609.

Reviewing Editor: David Case