Surface Charges Contribution to Protein Stability of *Thermococcus celer L30e*

CHAN, Chi Ho

A Thesis Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

in

Molecular Biotechnology

The Chinese University of Hong Kong

September 2010

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Thesis/Assessment Committee

Professor P.C. Shaw (Committee Chairman)

Professor K.B. Wong (Thesis Supervisor)

Professor S.M. Ngai (Committee Member)

Prof. D. Li (External Examiner)

Statement

Double mutant cycle of K46/E62 using charge-to-Ala mutants was conducted by Miss T.H. Yu. All experimental works reported here were conducted by the author, unless stated the otherwise.

CHAN, Chi Ho

Acknowledgements

I would like to express my gratitude to my supervisor, Prof. K.B. Wong, for his guidance and opportunities given throughout my 4-year doctoral study. In addition to the various experimental and computational techniques, more importantly, I have also leant the attitude to be a good researcher.

I would also like to thanks Prof. P.C. Shaw and Prof. S.M. Ngai for their useful comments and suggestions to my study. Tribute is also paid to all colleagues in Prof. K.B. Wong's laboratory. Special thanks should be given to Miss T.H. Yu who has put her effort on the double-mutant-cycle study.

Finally, I would like to express my greatest thank to my girl-friend Dr. H.L. Chan for her unlimited support and advices in all aspects during my Ph.D. study. Whenever I get frustrated, she is always the first one who stands by me and gives me courage to overcome any difficulties.

Abstract

Electrostatic interaction has long been proposed to be an important factor for stabilizing protein. Charge-charge interaction may especially be important to the thermostability of protein, as having more surface electrostatic interactions is one of the common structural features found in thermophilic proteins when compared to their mesophilic homologues. In order to quantitatively investigate the electrostatic contribution to protein stability, two complementary approaches, namely the double mutant cycle approach and pK_a shift approach, were carried out.

In the double mutant cycle approach, the coupling free energies of two salt bridges (E6/R92 and K46/E62) and one a long range ion pair (E90/R92) were estimated by using circular dichroism, to find out the thermodynamic parameters of the protein model *Thermococcus celer L30e* and its charge-to-neutral mutants. It was found that the coupling free energy was temperature independent and was about 3 kJ mol^{-1} per salt bridge. By using a novel analysis of double mutant cycle of ΔC_p, it was also found that the interaction of salt bridge plays an important role in the reduction of ΔC_p. The temperature independency of coupling free energy and the effect of reducing ΔC_p could explain the general observation very well that thermophilic proteins have highly up-shifted protein stability curves is due to its elevated electrostatic

interactions when compared with their mesophilic homologs.

Wild-type *T. celer L30e* has also shown to have no observable residual structure in the guanidine HCl-induced denatured state, indicating that denatured state of *T. celer L30e* should not have large effect on the overall protein stability.

To overcome the unwanted crystallization problem of wild-type *T. celer L30e* in the low ionic strength neutral pH NMR conditions, which were essential for the pK_a shift approach, a quintuple Arg-to-Lys variant was designed to dramatically improve the crystalline solubility, while the surface charges, as well as the structural, thermodynamic, and electrostatic properties, were conserved. It has also shown that electrostatic interaction played a critical role in crystallization at low ionic strength conditions, and arginine residue was especially important in crystal packing because of its high ability of forming salt bridges and hydrogen bonds.

In the pK_a shift approach, the native state pK_a values of acidic residues were obtained by fitting the side chain carboxyl ^{13}C chemical shifts to microscopic model or global fitting of titrational event (GloFTE), whereas the denatured state pK_a values were obtained by conventional pH titration of terminal protected 5-residue glycine-based model peptide. It was found that the surface charge-charge interactions, either attractive or repulsive, were strong and complicated because of the high surface charge density of *T. celer L30e*. However, the fact that most of the acidic residues

have significantly downshifted native state pK_a values indicated the surface charge distribution of *T. celer L30e* is optimized for stabilizing the protein. In addition, we have shown that temperature has negligible effect on pK_a values in both native state and denatured state, therefore temperature can only marginally amplify the stabilizing effect in linear manner.

摘要

靜電作用是一個穩定蛋白質的重要因素。靜電作用很可能對蛋白質的熱穩定性特別重要，因為當與嗜中溫同源蛋白質比較時，其中一個嗜熱蛋白質的共同結構特點是更多的表面靜電作用。為了定量地調查靜電作用對蛋白質穩定性的貢獻，我們利用了兩種互補的方法，即雙突變循環和酸解離平衡常數轉移方法。

在雙突變循環方法中，我們利用了圓二色光譜來估計兩組鹽橋之間（E6/R92和K46/E62），以及一對遠程離子對（E90/R92）之間的偶合自由能，來找出蛋白質模型，*Thermococcus celer L30e*和它的帶電變中立的突變體的熱力學參數。偶合自由能被發現是與溫度變化無關的，而數值大約是每鹽橋3 kJ mol^{-1}。另外，通過使用嶄新的蛋白熱容量差雙突變循環分析，發現了鹽橋的偶合自由能在減少蛋白熱容量中扮演一個重要的角色。獨立於溫度變化的偶合自由能以及減少蛋白熱容量的特性很好地解釋了嗜熱蛋白質有高度地提高了的蛋白質穩定曲線這一普遍觀察所得的現象，歸結於當與嗜中溫同源蛋白質比較時，嗜熱蛋白質有提升了的靜電作用。

野生型*T. celer L30e*在胍導致的變性狀態沒有可測的殘餘結構，表明*T. celer L30e*的變性狀態對全蛋白質穩定不應該有巨大影響。

為了克服野生型*T. celer L30e*在酸解離平衡常數轉移方法中必要的低離子強度及中性酸鹼值的核磁共振實驗條件中的結晶問題，我們設計了一種五倍Arg變

Lys變種。這變種在大大改進了蛋白質的結晶可溶性的同時，保存了野生型的表面電荷，以及其結構的、熱力學的、及靜電的特性。我們同時發現，靜電作用在低離子強度條件的結晶扮演一個重要角色，而由於精氨酸殘基在形成鹽橋和氫鍵的高能力，它對晶體排列特別重要。

在酸解離平衡常數轉移方法中，酸性殘基的天然狀態酸解離平衡常數透過適應側鏈羧基^{13}C化學位移的數據到微觀模型或總體擬合滴定事件方法（GloFTE）而獲得，而變性狀態的酸解離平衡常數則由基於終端被保護的五連甘胺酸殘基模型肽的常規酸鹼值滴定法得到。我們發現，表面的靜電作用，不論是不同電荷之間的吸引，或者是相似電荷之間的排斥，都由於*T. celer L30e*有很高的表面電荷密度而變得強而複雜。然而，大多酸性殘基的天然狀態酸解離平衡常數顯著調低表明了*T. celer L30e*的表面電荷分佈是為了穩定蛋白質而優化了的。另外，溫度對天然狀態以及變性狀態的酸解離平衡常數的作用都是微不足道的，因此溫度可能只少量地以線性方式擴大穩定效應。

Content

List of Figures

List of Tables

Chapter 1: Concept, overview, and methods on electrostatic contribution to protein stability

1.1 Definition of protein stability

In general, proteins are folded polymers (polypeptides) that built from 20 common amino acids. Different combinations of these 20 common building blocks enable proteins to exhibit different structural, functional, and thermodynamic properties. One of the important thermodynamics properties is "protein stability". Speaking in layman terminology, protein stability reflects how tough the folded protein can resist its breakage. More precisely, protein stability usually refers to the Gibbs free energy of unfolding (ΔG_u), which in turn refers to the difference between the free energy of the denatured (G_D) and native state (G_D):

$\Delta Gu = G_D - G_N$

Before going to further discussion on protein stability, we have to clarify two terms that have mentioned – native state and denatured state.

All proteins are synthesized from ribosome as linear polypeptides *in vivo*. Although having high conformational entropy, this linearly arranged conformation without forming any "structures" should be energetically unfavourable due to the factors such as exposure of hydrophobic groups, unpairing of hydrogen donors and

acceptors, lack of favourable electrostatic interactions, etc. Since the energy level of the linear polypeptide is high, this conformation is not expected to be stable. Instead, there must be countless of unstable conformations that have similar energy levels, and the polypeptides should populate in these conformations according to the Boltzmann distribution. Since the energy difference between these conformations is very small, the polypeptides should dynamically and evenly populate in these unstable conformations. In another word, the polypeptides are inter-converting between these unstable conformations with no major populated conformation. These unstable conformations are termed "random coil structure", or more commonly, "unfolded state" or "denatured state".

The unstable denatured state conformations are in general unable to retain in physiological condition *in vivo*. They will spontaneously convert into other conformations that have lower energy levels. This process is "protein folding". Since the conformational entropy decreases during protein folding, the number of possible conformations at lower energy level will also decrease. Under this manner, the protein folding process can be thermodynamically described as a free energy funnel (Figure 1.1). As the folding proceeds, the number of possible conformation decreases until the final stable confirmation is reached. However, whether the observable protein structure in the real world is the only globally

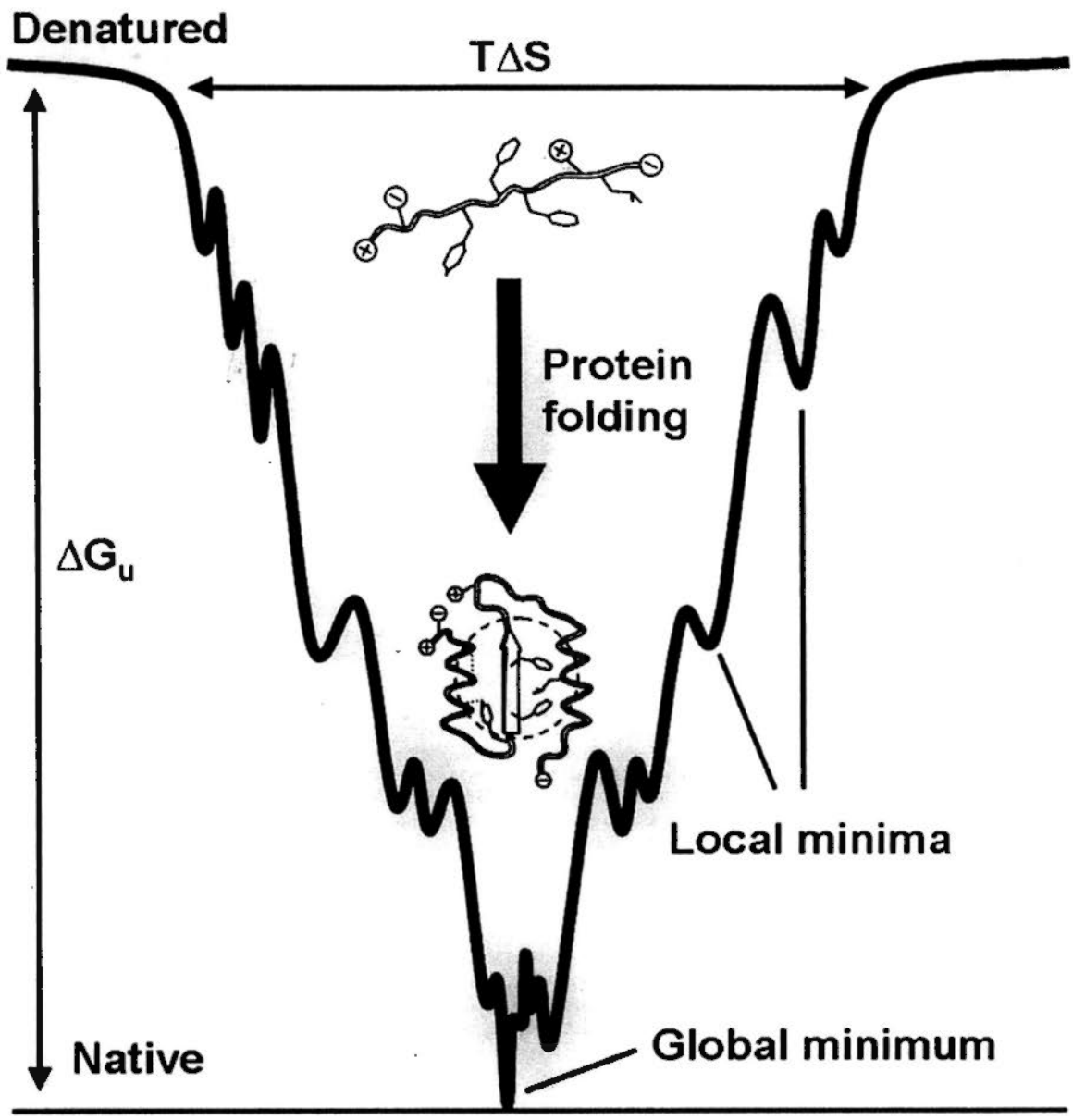


Figure 1.1. Free energy funnel for protein folding. Unstable polypeptides become more native-like upon protein folding, where its conformational entropy (TΔS) decreases while the free energy of unfolding (ΔG_u) increases in the folding process. Local minima along the sides of free energy funnel represent folding intermediates, whereas the global minimum represents the native protein fold according to thermodynamics hypothesis.

energetically minimum state has been being debated. Levinthal C. proposed a famous paradox in his early research[1]. He postulated that there are too many possible conformations in the denatured state and a protein needs to take a time longer than the age of the universe to achieve its correct structure if all the conformations are searched randomly. However, the truth is that proteins usually folded spontaneously on ms or even μs time scale. Therefore, the folding process must be kinetically controlled, and the final conformation may not be the energetically global minimum. On the contrary, based on the denaturation-renaturation experiments, Anfinsen C. had suggested the thermodynamic hypothesis of protein folding that unstable polypeptide reaches its final stable conformation by searching the energetically global minimum[2]. Clear solution for this argument is difficult to achieve as experimental supports for both sides have also been reported[3-8]. In a more recent computational simulation study, it was suggested the final stable conformation is most often the state of minimum free energy due to the evolution of protein[9]. Nevertheless, the final stable conformation is called the "native state" of the protein.

The folding pathway has not been clearly understood till now. It is still widely being investigated using structurally simple models like ankyrin repeat proteins[10-18]. However, whether it is kinetically controlled, or follows other models like the hierarchical model which proposes the folding initiates locally and folded elements

associate step-wisely to obtain final native structure[19,20], the protein folding pathway is not important in the thermodynamic point of view. This is because energy level is a state function that the kinetics between different states (folding pathway) should have no effect on the equilibrium (energy level) of states. Therefore, protein folding pathways would not be further discussed here.

Protein stability is the difference between energy level of a single stable native conformation, and the averaged energy level of an ensemble of unstable denatured conformations (Figure 1.2). Gibb's free energy can be calculated from thermodynamic parameters:

$$\Delta G_u = -RT \ln\left(\frac{k_u}{k_f}\right)$$

where R is the gas constant, T is the absolute temperature, k_u and k_f are the unfolding and folding rate constants. One should notice that protein stability is not a constant, but a function of its physical and chemical environments. For example, it is well known that Gibb's free energy changes with temperature and pH. In other words, the global energy minimum in the funnel and its corresponding native conformation may change according to the change of the physical and chemical conditions. That is why denaturation of protein can be induced by temperature or denaturants like urea and guanidine-HCl.

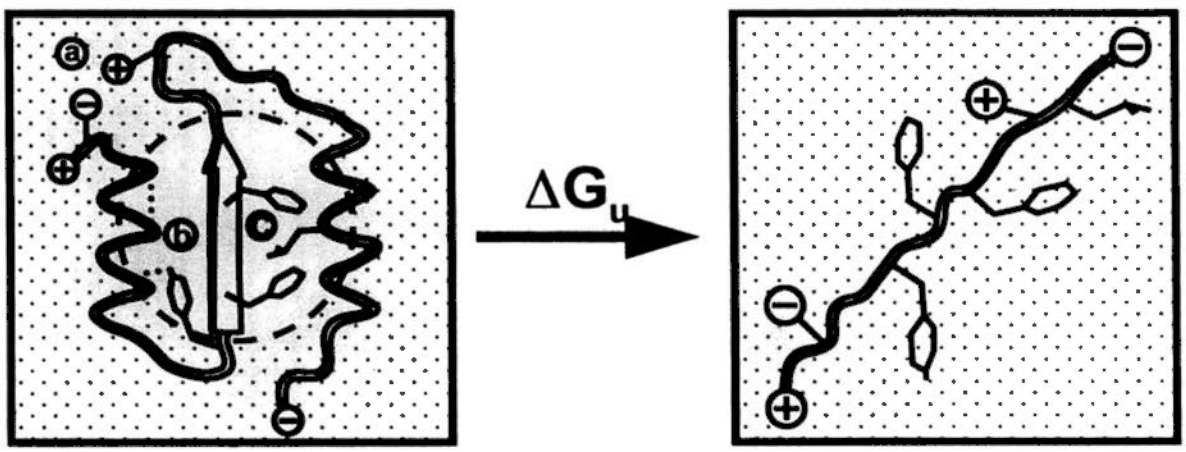
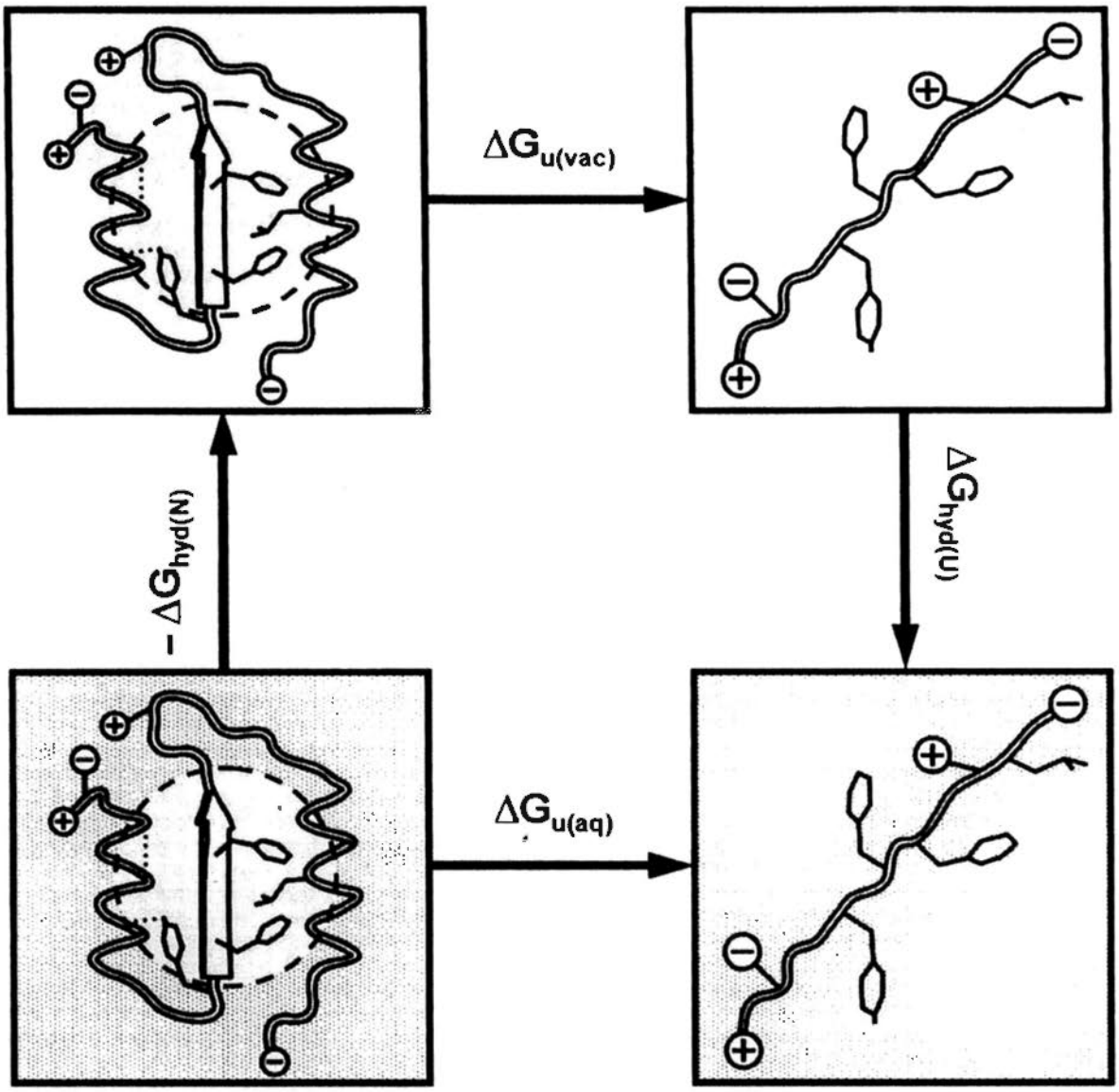


Figure 1.2. Two state protein unfolding. Free energy of unfolding (ΔG_u) is defined as the free energy difference of the single stable native conformation and the ensemble of unstable denatured conformations. Various kinds of interaction such as (a) charge-charge interactions, (b) hydrogen bonds, and (c) hydrophobic interactions were formed in the native protein. The hydrophobic core of the native state protein is indicated by a dashed circle.

1.2 Contribution of electrostatic interaction in native state to protein stability

Although how protein rapidly and accurately folds into its particular native state conformation is still not yet clearly understood, the fact that the structure of a protein is determined by the amino-acid sequence and the structure of protein determines its function, has come to consent for many years. Some recent studies suggested the internal dynamics can influence protein activity[21-23]. However, these suggestions do not affect the big picture of protein function-structure relationship.

Having the most critical importance, the protein stability of the native functional conformation is surprisingly only the relatively small difference of about 20 – 60 kJ mol^{-1} [24] as a result of balancing between very large opposing forces[25,26]. The terms that affect protein stability have been summarized in figure 1.3. The major opposing forces of protein folding come from conformational entropy (ΔS_{conf})[27] and the hydration effect upon protein unfolding ($\Delta\Delta G_{hyd}$)[26,28]. Their summation can account for destabilizing energy in the order of several hundreds kJ mol^{-1}. On the other hand, the stabilizing forces of comparable size have been suggested to be the enthalpy of various kinds of interactions in native protein (ΔH_{int}), which comes from factors like hydrophobic effect[5,29], van der Waals forces (native state packing)[30-33], and peptide



$$\Delta G_{u(aq)} = -\Delta G_{hyd(N)} + \Delta G_{u(vac)} + \Delta G_{hyd(U)}$$

$$\Delta G_{u(aq)} = \Delta\Delta G_{hyd} + \Delta H_{int} - T\Delta S_{conf}$$

Figure 1.3. Thermodynamic cycle of unfolding and dehydration. Thermodynamic cycle can separate the free energy change due to hydration ($\Delta\Delta G_{hyd}$) from the free energy change due to the intra-molecular interactions (ΔH_{int}) and conformation entropy ($T\Delta S_{conf}$). Be noted that protein stability ($\Delta G_{u(aq)}$) is a delicate balance between two opposing forces.

hydrogen bonds[5,34,35].

In addition to the factors mentioned above, electrostatic interaction is another intra-molecular force commonly observed in proteins. Electrostatic interaction refers to the non-directional columbic force between two partial or full point charges. The electrostatic potential between the 2 charges is inversely proportional to the separating distance and the dielectric constant between them. The interaction energy between two charges in vacuum can be calculated by Coulomb's Law:

$$U_{ij} = k\frac{q_i q_j}{\varepsilon_c r_{ij}}$$

where U_{ij} is the electric potential (kJ mol^{-1}) induced by the ion pair, q_i and q_j ($0 < q_i,q_j \leq 1$) are the charge (eV) of the ionizable groups, r_{ij} is the separating distance (Å) between the two charges, ε_c is the averaged dielectric constant which is equal to 1 in vacuum, and k is the conversion factor which is equal to 1385.5 Å kJ mol^{-1}.

In protein, electrostatic interactions mostly originated from the interactions of the ionizable groups in the side chain of Asp, Glu, Arg, Lys, and His, as well as more rarely the N-terminus and C-terminus. The interacting oppositely-charged ionizable groups are called "ion pair". Ion pair is in general classified as "salt bridge" if the r_{ij} is shorter than 4.0 Å, whereas "long range" interaction is usually describing interaction of ion pair with r_{ij} is larger than 8.0 Å. Since the charging of the ionizable groups is dependent on their pK_a values as well as the pH of the solvent, the effect of

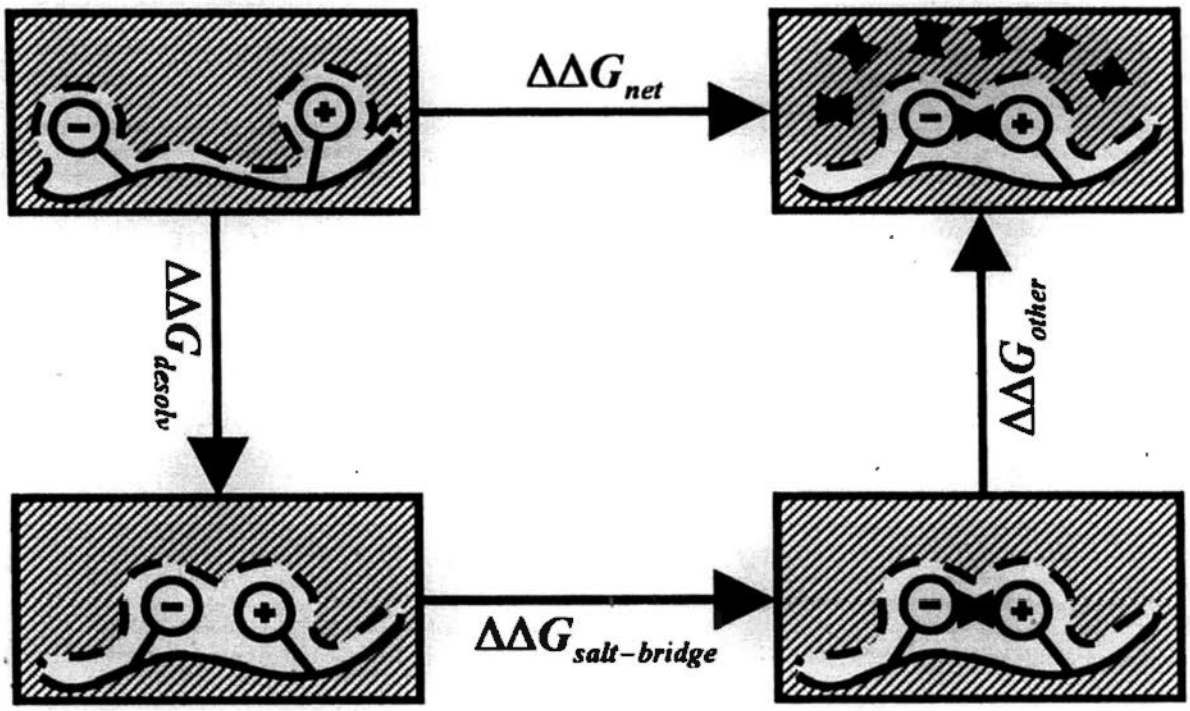
electrostatic interactions on protein stability is therefore pH-dependent.

In vacuum, electrostatic interaction can be very strong, even if the interaction is "long range". Assuming the ionizable groups in proteins are point charges and are fully protonated or deprotonated (only carry +1 or -1 eV per group). Neglecting the short distance Lennard-Jones potential and van der Waals interaction of which the effects should be negligible in long distance, electrostatic interaction of an ion pair is 173.2 kJ mol^{-1} with r_{ij} equal to 8.0 Å. The interaction energy is still significant (6.9 kJ mol^{-1}) even the r_{ij} is increased to 200 Å (approximately equal to the hydrodynamic diameter of large monomeric globular proteins).

This apparently very strong electrostatic interaction in native protein, however, is controversial to be a factor that contributes to protein stability. Instead, electrostatic interaction has been widely suggested to be important in protein folding[36] and specific interactions such as protein and ligand binding)[37]. The controversy of electrostatic contribution to protein stability is due to its experimental inconsistence to stabilize proteins. Very often, removal of surface charges from protein results in minimal changes in protein stability[37-43], or even have destabilizing effect[44-46]. Bosshard H.R. and co-workers have found that the pK_a values of all 13 ion-pair forming Glu in designed coiled coil AB zipper were similar to the pK_a values of Glu in unfolded peptide, suggesting electrostatic interaction has very

limited effects on protein stability[47]. However, experimental evidence also showed electrostatic interaction can also contribute to protein stability[19,20,48-53]. As an example, our group has demonstrated removal of most of the charges on ribosomal protein *T. celer L30e* will lead to destabilizing effect[54]. One trivial reason for the controversy is that electrostatic can be repulsive among like charges. Since the electrostatic interaction can act in a long range manner[55] (up to 8 Å), having repulsive interaction between like charges is not rare, especially for thermophilic proteins which are well known to have elevated number of charged residues[56]. Moreover, the stabilizing effect of electrostatic interaction has been challenged by the large desolvation penalty[37,38,57] and entropic cost[43,44,46] of salt bridge formation.

These seemingly contradicting data and explanations simply alert us one important fact – electrostatic interaction is a complicated summation of various effects. To describe the electrostatic contribution for computational simulation, Hendsch and Tidor had used a thermodynamic cycle (Figure 1.4) to dissect net electrostatic contribution of a charge pair ($\Delta\Delta G_{net}$) into three terms: the contribution due to the interaction between charge pair ($\Delta\Delta G_{salt\text{-}bridge}$), the contribution due to interaction between the charges to the rest of the polar group in the protein



$$\Delta\Delta G_{net} = \Delta\Delta G_{salt-bridge} + \Delta\Delta G_{desolv} + \Delta\Delta G_{other}$$

Figure 1.4. Thermodynamic cycle of salt bridge contribution to protein stability. The net electrostatic contribution of forming a salt bridge ($\Delta\Delta G_{net}$) is the sum of desolvation effect ($\Delta\Delta G_{desolv}$), interaction of the salt bridge ($\Delta\Delta G_{salt\text{-}bridge}$), and the interaction of the two charges residues to the rest of the protein ($\Delta\Delta G_{other}$).

($\Delta\Delta G_{other}$), and the contribution due to desolvation of ion pair formation ($\Delta\Delta G_{desolv}$)[57]. More complicated models have also been built to describe electrostatic interactions for computational simulation, where some are based on physical effective energy functions[58-60], while the others are based on statistical effective energy functions[61-63]. For details of energy functions please refer to reviews of Sippl M.I.[64], Lazaridis T. & Karplus M.[65], and the references therein. Using various models in theoretical calculations, the role of electrostatic interaction remains unclear. While some calculation suggested electrostatic interaction has only little effect on protein stability[61], some suggested electrostatic interactions can have a large effect on protein stability[66].

As more and more experimental data (mutagenesis data) of electrostatic interaction studies are accumulated, the best way to have a comprehensive overview about the stabilizing effect of electrostatic interactions is to have a statistic analysis on the reported data. Afterall, computation simulations have a list of assumptions, which are usually difficult to validate. To achieve this, we have analyzed the electrostatic contribution ($\Delta\Delta G_u$) for mutagenesis deposited in the ProTherm database[67,68]. A database of 164 charge-to-alanine mutations of 25 protines has been established, where the 25 corresponding protein crystal structures were available in high resolution. As expected, these $\Delta\Delta G_u$ values were found to have a very broad

Gaussian-like distribution, ranging from stabilizing by 16 kJ mol^{-1} to destabilizing by 14 kJ mol^{-1} (Figure 1.5). This broad distribution indicated the effect of charge removal can be very context dependent. However, most of these mutations were slightly stabilizing with the average $\Delta\Delta G_u$ value of 2.4 ± 4.2 kJ mol^{-1}. Correlating these data to protein structures, it was found that the average $\Delta\Delta G_u$ of a mutagenesis subset data (31 out of 164) of isolated charges (no charges within 8 Å) was 1.9 ± 3.0 kJ mol^{-1}, which indicates the long range interaction of a charge is in average slightly stabilizing. Similarly, the average $\Delta\Delta G_u$ of mutagenesis data where at least one salt bridge (< 4.0 Å) was broken in the removal of the charge (48 out of 164) was 3.0 ± 3.9 kJ mol^{-1}. Although the variation of these data is too large to make a concrete conclusion here, the apparently higher mean $\Delta\Delta G_u$ value of salt bridge breaking charges than that of isolated charges cues us that salt bridges also play a role in electrostatic contribution. Data from double mutant cycles also suggested salt bridge stabilizes protein by 3.2 ± 3.9 kJ in average (see Chapter 3), which is similar to the statistics of single site-directed mutagenesis data. Summing up, electrostatic interaction is in average slightly stabilizing (about 3 kJ mol^{-1} per salt bridge), but the effect of removing charges can be very much context dependent. Be noted that $\Delta\Delta G_u$ is not only reflecting the electrostatic interaction of all ion pairs involved, but a sum of many effect. In analogy to the protein stability (Figure 1.3), the net electrostatic

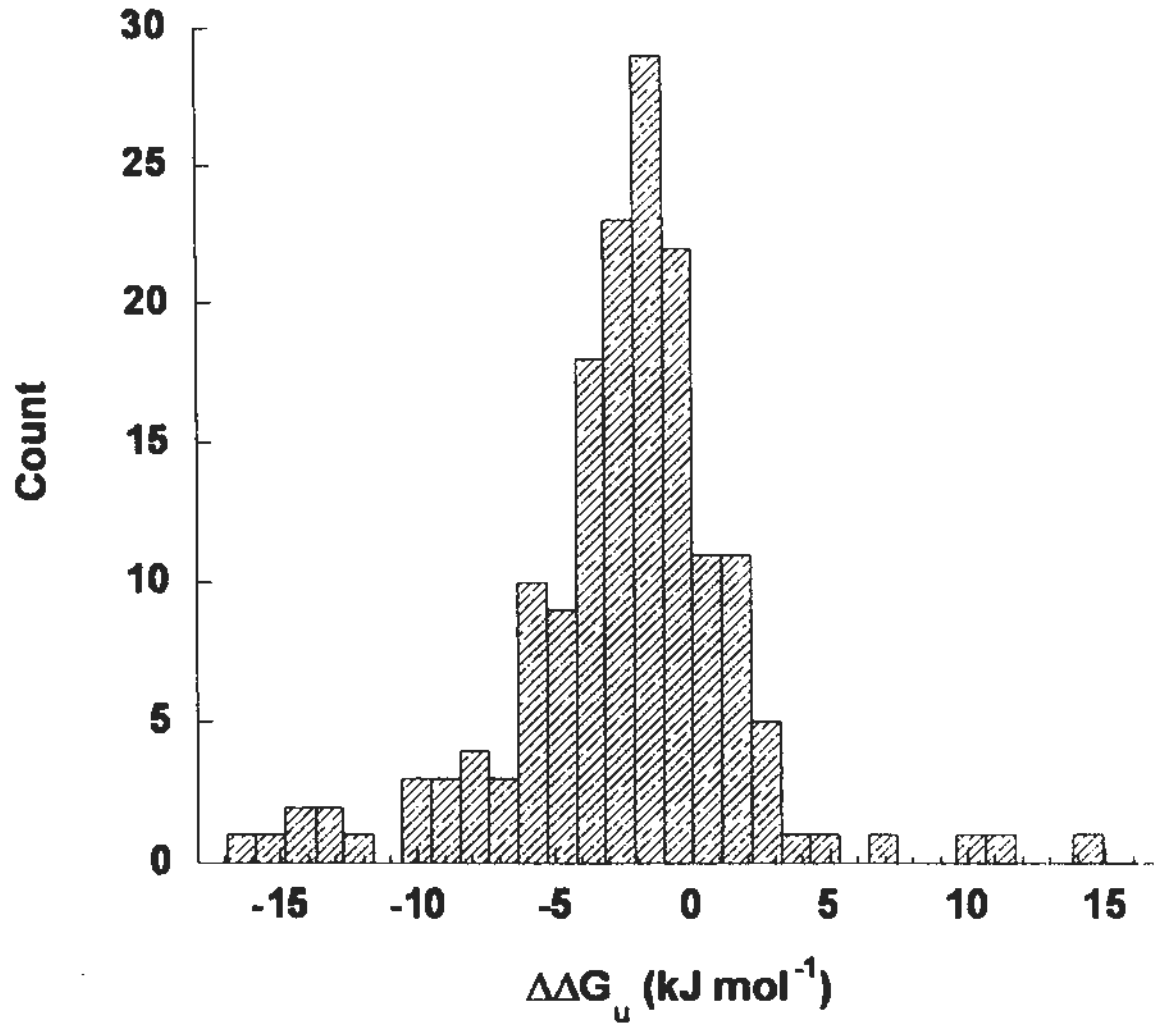


Figure 1.5. Histogram of charge-to-Ala $\Delta\Delta G_u$. The distribution of change in ΔG_u due to single charge-to-Ala mutation ($\Delta\Delta G_u$) is represented in histogram. Non-redundant data of 164 mutation of 25 proteins were used in the 31-bin histogram. A very broad Gaussian-like distribution was observed.

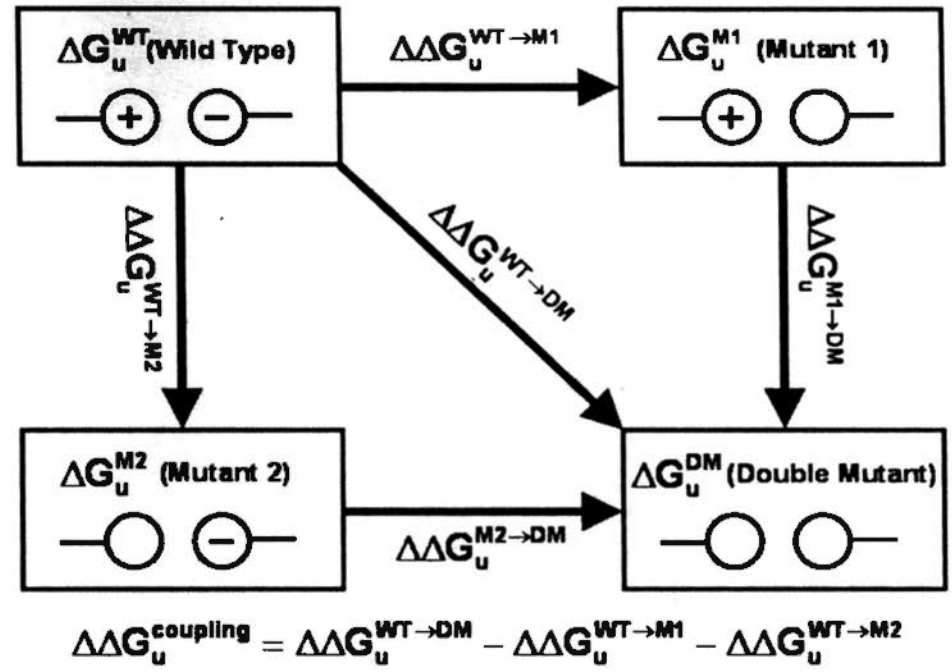
contribution (ΔG_{net}) is also a complicated value taking the change of hydration free energy upon breaking of the ion pair ($\Delta\Delta G_{hyd}$) and side chain conformational entropy (ΔS_{conf}) into account, in addition to the electrostatic enthalpy of ion pair in vacuum (ΔH_{int}). Imagine an ideal case that a protein which only has one acidic and one basic residue. The electrostatic contribution of this ion pair (ΔG_{net}) found by single charge-to-neutral mutation is not expected to be the same if the value is obtained by mutating different charges. It is because mutation will inevitably change the $\Delta\Delta G_{hyd}$ and ΔS_{conf} in addition to removal of ΔH_{int}, but the value of $\Delta\Delta G_{hyd}$ and ΔS_{conf} are not the same for mutating different residues, not to mention the difference in electrostatic interaction of charge to the rest of the protein which will be present in real protein.

1.3 Methods for experimental studies of electrostatic interaction

To determine the electrostatic contribution of an ion pair ideally, the interaction of this ion pair have to be removed without affecting structural and thermodynamic properties, as well as the electrostatic properties of the rest of the protein, in both native and denatured state. Although there is currently no simple experimental method to fulfill all these idealistic criteria, there are two complementary methods namely "double mutant cycle approach" and "pK_a shift approach", which disrupt ion pair interactions by mutation and pH respectively, can in principle experimentally

estimate the interaction energies between charged residues.

In double mutant cycle approach, the ion pair interaction is invasively removed by mutation. As discussed above, the loss of free energy from mutation cannot be equated with the loss of ion pair interaction energy. However, quantitative information about the electrostatic contribution of ion pair can be estimated by combining mutations in a double mutant cycle[69-72], which is designed to cancel out all effects of mutations except the interaction energy between the ion pair in study[72]. The interaction energy of ion pair found by double mutant cycle is called "coupling free energy"[72]. The principle of double mutant cycle is illustrated in Figure 1.6. The lost in unfolding free energy of mutating one charged residue from wild-type protein ($\Delta\Delta G_u^{WT\rightarrow M1}$ and $\Delta\Delta G_u^{WT\rightarrow M2}$) consist 2 components, the coupling free energy ($\Delta\Delta G_u^{coupling}$) and the interaction energy of the mutated residue to the rest of the protein. Mutating another charged residue will lead to further decrease in free energy ($\Delta\Delta G_u^{M1\rightarrow DM}$ and $\Delta\Delta G_u^{M2\rightarrow DM}$), which is the interaction energy between the mutated charge and the rest of the protein only (Figure 1.6). If two charged residues are not interacting with each other (i.e. $\Delta\Delta G_u^{coupling} = 0$), the unfolding free energy change by mutating both charges (mutation sequence has no effect) will be equal to the sum of the unfolding free energy changes of the two single mutants:



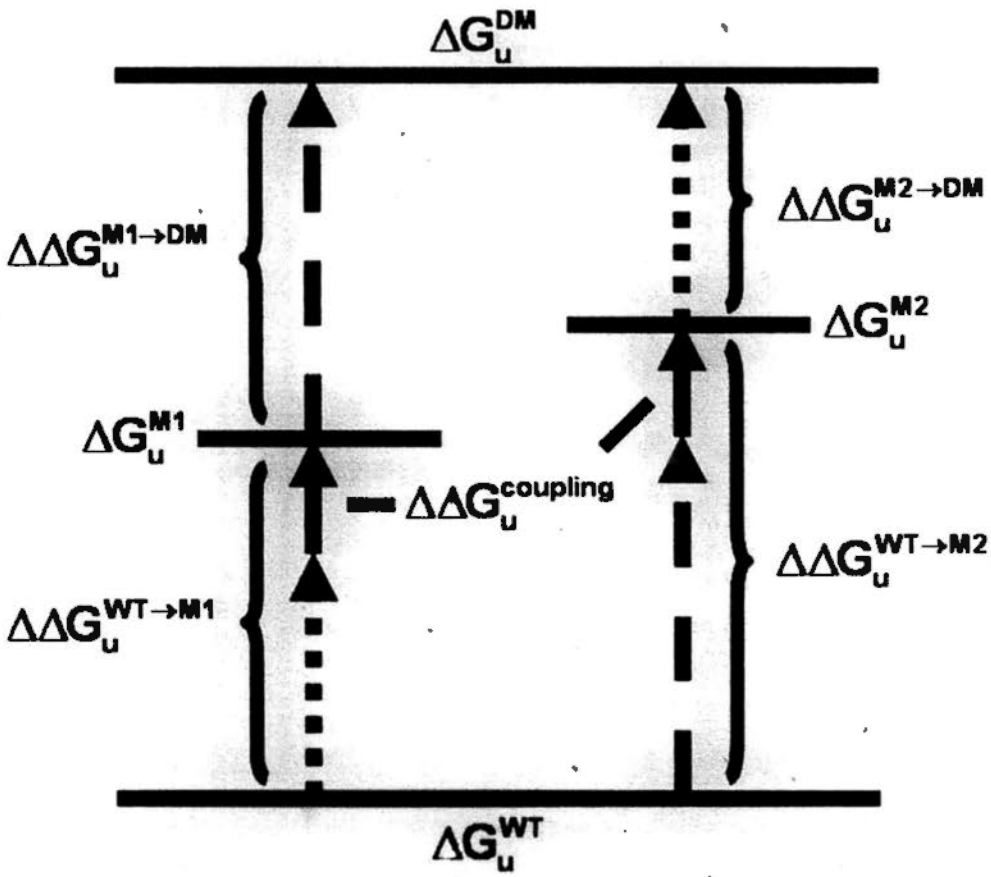


Figure 1.6. The scheme of double mutant cycle of free energy of unfolding. According to the double mutant cycle shown above, the effect of a charge on ΔG_u due to its interaction to the rest of the protein is cancelled in the cycle, so that the coupling free energy of the ion pair can be calculated, assuming the effects of charges on ΔG_u are additive.

$$\begin{aligned}\Delta\Delta G_u^{WT\to DM} &= \Delta\Delta G_u^{WT\to M1} + \Delta\Delta G_u^{WT\to M2} \\ &= \Delta\Delta G_u^{WT\to M1} + \Delta\Delta G_u^{M1\to DM} \\ &= \Delta\Delta G_u^{WT\to M2} + \Delta\Delta G_u^{M2\to DM}\end{aligned}$$

However, if the two charges are interacting with each other, the effect of mutating one residue will couple to the mutation of another residue, so that the additivity of unfolding free energy change of single mutation breaks down:

$$\Delta\Delta G_u^{WT\to DM} \neq \Delta\Delta G_u^{WT\to M1} + \Delta\Delta G_u^{WT\to M2}$$

The failure in additivity is because the unfolding free energy change of single mutation is not simply equal to the interaction energy between the mutated residue and the rest of the protein:

$$\Delta\Delta G_u^{WT\to M1} \neq \Delta\Delta G_u^{M2\to DM} \quad \text{and} \quad \Delta\Delta G_u^{WT\to M2} \neq \Delta\Delta G_u^{M1\to DM}$$

The coupling free energy is in fact the difference between the unfolding free energy change of single mutation and that due to interaction energy between the mutated residue and the rest of the protein:

$$\begin{aligned}\Delta\Delta G_u^{coupling} &= \Delta\Delta G_u^{WT\to M1} - \Delta\Delta G_u^{M2\to DM} \\ &= \Delta\Delta G_u^{WT\to M2} - \Delta\Delta G_u^{M1\to DM} \\ &= \Delta\Delta G_u^{WT\to DM} - \Delta\Delta G_u^{WT\to M1} - \Delta\Delta G_u^{WT\to M2}\end{aligned}$$

which can be simplified to:

$$\Delta\Delta G_u^{coupling} = \Delta G_u^{WT} + \Delta G_u^{DM} - \Delta G_u^{M1} - \Delta G_u^{M2}$$

Large positive $\Delta\Delta G_u^{coupling}$ indicates the interaction of the two residues is strongly stabilizing.

Double mutant cycle approach is a powerful tool to estimate ion pair interaction energy. However, one should be reminded the yielded coupling free energy is not necessarily purely electrostatic. Other assumptions and limitations such as no structural changes induced by mutation and no new interactions formed by the substituting residues[50,73,74] should also be noticed and validated. In practice, it is virtually impossible to validate every assumption due to technical difficulties (especially validating denatured state properties). Probably the best way to validate the coupling free energy found is solely due to the interaction energy between two residues in study is to repeat independent double mutant cycles using different kinds of substitutions for the same ion pair[75]. If the coupling free energies found by different independent double mutant cycles are the same, then the value can be confidently concluded as the interaction energy of the ion pair.

In pK_a shift approach, the ion pair interaction was non-invasively removed by pH, or more precisely by protonation of acidic residues and deprotonation of basic residues. As the name indicates, pK_a shift approach estimates the electrostatic contribution of a charge from the difference in the pK_a values in native and denatured state. Be noted that one can neither separate any non-charge contribution, such as contribution due to structural rearrangement induced by protonation/deprotonation, from the electrostatic contribution, nor determine the net energetic contributions of

both residues of a salt bridge to protein stability at a given pH.

The principle of pK_a shift approach can be understood by considering the thermodynamic cycle of unfolding and protonation/deprotonation (Figure 1.7). Let's take acidic residue (Asp or Glu) to illustrate the idea. The electrostatic contribution to free energy of unfolding ($\Delta\Delta G_u^{ele}$) can be calculated by:

$$\Delta\Delta G_u^{coupling} = \Delta G_u^{p} - \Delta G_u^{dp}$$

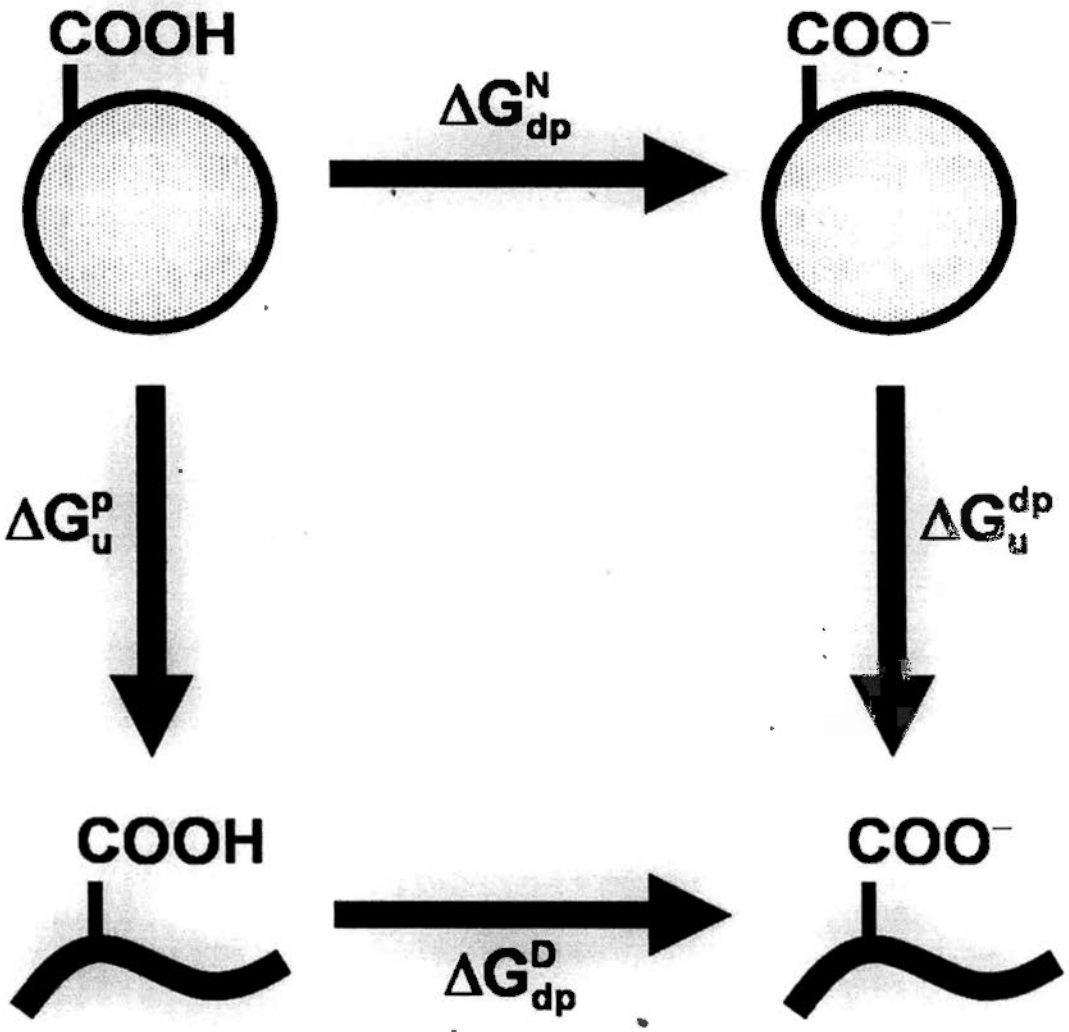
where ΔG_u^{p} and ΔG_u^{dp} are the free energy of unfolding of which the charge in study is in protonated and deprotonated form respectively. Be noted that the $\Delta\Delta G_u^{ele}$ value of basic residue should be multiplied by -1, as the protonated form of basic residue is charged. However, in practice, ΔG_u^{dp} and ΔG_u^{p} cannot be measured as the pure protonated and deprotonated population can by no means be separated. Yet, according to the thermodynamic cycle, we can see that:

$$\Delta G_{dp}^{N} + \Delta G_u^{dp} = \Delta G_u^{p} + \Delta G_{dp}^{D}$$

where ΔG_{dp}^{N} and ΔG_{dp}^{D} are the free energy of deprotonation in the native and denatured state respectively. By simple transformation, we have:

$$\Delta\Delta G_u^{ele} = \Delta G_u^{p} - \Delta G_u^{dp} = \Delta G_{dp}^{N} - \Delta G_{dp}^{D} = \Delta\Delta G_{dp}^{fold}$$

Therefore, the electrostatic contribution to free energy of unfolding is equal to the folding contribution to free energy of deprotonation ($\Delta\Delta G_{dp}^{fold}$), which can in turn be calculated by:



$$\Delta\Delta G_u^{ele} = \Delta\Delta G_{dp}^{fold} = -\ln(10)RT\left(pK_a^N - pK_a^D\right)$$

Figure 1.7. Thermodynamic cycle of deprotonation and unfolding. The thermodynamic cycle showed the scheme of pK_a shift approach. According to the thermodynamic cycle, the electrostatic contribution to unfolding free energy of an acidic residue ($\Delta\Delta G_u^{ele}$) is equal to folding contribution to free energy of deprotonation ($\Delta\Delta G_{dp}^{fold}$), which in turn can be calculated if the pK_a values of that acidic residue in native state (pK_a^N) and in denatured state (pK_a^D) are known.

$$\Delta\Delta G_{dp}^{fold} = -\ln(10)RT\left(pK_a^N - pK_a^D\right)$$

Therefore, by finding the pK_a values of the charge in native and denatured state, the electrostatic contribution to protein stability can be found. Thus, the accuracy of the experimentally found pK_a values is of the major consideration when estimating the electrostatic contribution of a charge to protein stability. To found out the pK_a values, various approaches have been proposed, such as finding pH dependency of enzyme activity or inhibitor binding, or comparing any pH-dependent properties (e.g. protein stability) before and after the removal of the charged residue in interest[76-80]. However, the most straight forward and accurate method is by NMR[81-83] because of its atomic level resolution. However, using NMR to trace the protonation and deprotonation of residues still have other data interpretation problems (see Chapter 6).

1.4 Denatured state can also contribute to protein stability

Be reminded that protein stability is the energy difference between native state and denatured state. Therefore, the denatured state is in principle has the same importance on protein stability as native state does. In fact, improvement of protein stability has long been tried by manipulating the denatured state, such as decreasing the entropy of the denatured state by introducing disulfide crosslinks[84-86] or by

substituting high entropy residues (Gly and Ala) with a more rigid residue (Pro)[87]. Although not all trials had encouraging results, the possibility of improving protein stability by manipulating denatured state is clearly demonstrated. Also, many studies showed clearly that residual structure can be present in denatured state of proteins, and the presence of residual structures may destabilize the global protein stability by stabilizing the denatured state[88-91]. Be clear that residual structure in denatured state is not a stable structure, but an ensemble of preferred conformation of the denatured polypeptide.

1.5 Objectives of this study

Base on the structural comparison of thermophilic proteins and their mesophilic homologues, it has long been argued that electrostatic interaction plays an important role in improving protein thermostability (protein stability at high temperatures)[56,92-98]. In a recent study in our group, some surface charges have been shown to be important for reducing ΔC_p of thermophilic proteins[99], which is a common strategy for improving thermophilic proteins to enhance their thermostability[100]. However, quantitative experimental study on pair-wise and global electrostatic contribution on thermostability is still lacking. In this study, one of the objectives is to quantitatively study how pair-wise salt bridge interaction reduces ΔC_p of a protein and contributes to its thermostability by double mutant cycle approach.

Thermophilic proteins are well known to have much more charges on its surfaces[56]. Since charge-charge interaction can be repulsive, whether the highly elevated number of charges on protein surface can improve protein stability is doubted. In this study, the global effect of surface charges on thermophilic protein would also be investigated by pK_a shift approach.

As discussed above, NMR is the best method to estimate pK_a values of charges in a protein. However, protein sample of very high concentration in low ionic strength condition was needed for the pK_a shift approach. Unfortunately, our protein model *Thermococcus celer L30e*, was found to crystallize in some NMR conditions. Therefore, a general de-crystallization method without affecting protein structural, thermodynamic, and electrostatic properties would be presented.

Lastly, denatured state can also contribute to global protein stability. Thus, various NMR-related strategies were applied to wild-type *T. celer L30e* to find out if there is any residual structure which can destabilize the protein.

Chapter 2: Materials and methods

In this chapter, general materials and methods used in this study will be reported. The techniques used are classified in seven categories: "General techniques" for the general techniques that were routinely used. These general techniques are well developed and are usually for supporting purpose; "Cloning, expression, and purification" for all DNA and proteins sample preparation techniques; "Circular dichroism experiments" for methods and techniques to obtain thermodynamic parameters of proteins; "Crystallization and crystal structure refinement" for procedures to obtain refined crystal structures; "Nuclear Magnetic Resonance experiments and pH titration" for the experimental details to obtain good NMR data and the strategies to obtain pK_a values of charged residues; and lastly, "Computational and statistical techniques" for reporting various softwares, in-house scripts, and the procedures of all statistics used; "Reagents and Buffers" for the detail recipe of all reagents and buffer used. Unless otherwise specified, all the experimental details used throughout this study will strictly follow the conditions reported in this chapter.

2.1 General techniques

2.1.1 Agarose gel electrophoresis

To yield a 1 % agarose gel, 0.5 g of argarose was mixed with 50 ml 1X TAE

buffer. The suspension mixture was then heated in microwave oven for about 2 minutes. After cooling the mixture to hand-hot by running water (clear solution should be observed), about 2 µl of ethidium bromide (EB) was added and mixed gently. The gel was then poured into gel tank with comb and allowed to set for about 15 minutes, and the solidified gel was then placed into TAE buffer. DNA samples (mixed with 6X DNA loading dye) and marker could than be loaded into the wells. Constant voltage of 120 V was applied for the electrophoresis for 15 – 30 minutes until the front was electrophoresed to the middle of the gel.

2.1.2 DNA extraction from agarose gel (Gene clean)

DNA extraction was performed with using commercial "gene clean" kits (Invitrogen), and reagents and procedures were provided by manufacturer. DNA in EB added agarose gel was detected by UV transilluminator, and the DNA containing slice was excised. The slice was then dissolved in about 0.5 ml gel solubilization buffer provided at 70 °C for about 10 minutes, with occasional shaking. The completely dissolved gel solution was then transferred to the spin cartridge privided with a 2-ml wash tube, followed by centrifugation at *13,000* rpm for 1 minute. After discarding the flow-through, 0.7 ml of ethanol added wash buffer provided in the kit was added to spin cartridge and incubated at room temperature for 2 minutes. Then, centrifuge the tube at *13,000* rpm for 1 minute. Flow-through was discarded and the

tube was centrifuged again for 3 minutes to remove any residual wash buffer. Then placed the spin cartridge into a new 1.5 ml centrifuge tube, and added 50 µl of pre-warmed nanopure water. After incubation at room temperature for about 2 minutes, the whole tube was centrifuged at *13,000* rpm for 1 minute to recover the DNA in the 1.5 ml centrifuge tube.

2.1.3 Plasmid DNA extraction from *E. coli* broth culture (mini-prep)

DNA extraction was performed with using commercial "mini-prep" kits (Invitrogen), and reagents and procedures were provided by manufacturer. Transformed E coli cells were collected by centrifuging 5 ml of overnight broth culture at 13,000 g for about 1 minute, and the pellet was resuspendend in 0.25 ml cell resuspension buffer provided. 0.25 ml of cell lysis solution provided was then added and mixed by inverting several times. Clear solution should be observed after mixing. After incubation for about 5 minutes, 0.35 ml of neutralization buffer provided was added and mixed immediately by inverting several times. The white precipitant formed was removed by centrifugation at 13,000 g for 10 minutes. The supernatant was transferred in a spin cartridge provided with 2-ml wash tube. After centrifugation at 13,000 g for about 1 minute and discarding the flow-through, 0.7 ml of ethanol added wash buffer provided was added to the cartridge. The whole tube

was centrifuged at *13,000* g for 1 minute followed by discarding the flow-through. Repeated the washing step once, and the spin cartridge was then placed into a new 1.5 ml centrifuge tube, and added 50 µl of pre-warmed nanopure water. After incubation at room temperature for about 2 minutes, the whole tube was centrifuged at *13,000* rpm for 1 minute to recover the DNA in the 1.5 ml centrifuge tube.

2.1.4 Sodium dodecylsulphate-polyacrylamide gel electrophoresis (SDS-PAGE)

All SDS-PAGE were run using Mini-PROTEAN III electrophoresis cell (BioRad). All components needed for SDS-PAGE were assembled according to manufacturer's instructions. The resolving and stacking gel solution mixture for casting sodium dodecylsulphate-polyacrylamide gels was mixed from pre-mixed high concentration stock reagents. The preparation of resolving and stacking gel solution for different acrylamide concentration are summarized in Table 2.1. About 3.6 ml of resolving gel solution was transferred to the kit and wait for about 10 minutes for polymerization. About 200 µl of isopropanol was layered on the top of the resolving gel solution to isolate it from oxygen and to remove any bubbles. After removing isopropanol and washing of the resolving gel phase, 5 % stacking gel solution was transferred to the kits with comb inserted. After polymerization time of about 5 to 10 minutes, the double layered gel was ready for electrophoresis. After

Table 2.1. Recipe of resolving and stacking gel for SDS-PAGE

For resolving gel:			
Components	10.0 % Volume (ml)	12.5 % Volume (ml)	15.0 % Volume (ml)
4X resolving gel buffer (pH 8.8)	1	1	1
40 % acrylamide	1	1.25	1.5
10 % ammonium persulfate (APS)	0.05	0.05	0.05
TEMED	0.01	0.01	0.01
H_2O	1.94	1.69	1.44
Total Volume	4	4	4

For stacking gel:	
Components	5.0 % Volume (ml)
4X stacking gel buffer (pH 6.8)	1
40 % acrylamide	0.5
10 % ammonium persulfate (APS)	0.03
TEMED	0.01
H_2O	2.46
Total Volume	4

loading suitable amount (< 20 μl) of protein marker (LMW) and protein samples, which were pre-mixed with 2X SDS loading dye and pre-heated by air bath at 95 °C for 5 minutes, constant voltage of 180 V, or constant current of 60 mA per gel was applied for the electrophoresis for about 30 – 45 minutes, until the dye front reached the bottom of the gel.

2.1.5 Staining of polyacrylamide gel

Polyacrylamide could be temporarily stained by zinc-imidazole reverse stain in quick-and-dirty manner by immersed in staining solution of 0.2 M imidazole and 0.1 % SDS for about 5 minutes followed by brief rinsing and the subsequence immersion of the gel in 0.2 M $ZnSO_4$ solution for about 30 second. Transparent bands with white background should be observed.

For persistent staining of protein bands, conventional Coomassie blue staining or sensitive colloidal Coomassie G-250 staining[101] could be used. The conventional staining could be achieved simply by immersing the gel in about 30 ml of conventional staining solution for about 1 hour with shaking, followed by immersing the stained gel in destaining solution of 25 % (v/v) ethanol, 8 % (v/v) acetic acid. The conventional staining solution was replaced by new staining solution for the sensitive colloidal Coomassie G-250 staining. No destaining step was needed in colloidal Coomassie staining. Both stainings yield blue protein bands with

transparent background.

2.1.6 Manipulation of protein concentration and buffer condition

Diluted protein samples were concentrated by adding less than 15 ml of protein sample in Vivaspin R15 (Vivascience), followed by centrifugation at 4500 g for 20 minutes per round. Pipette up-and-down was applied after each round of centrifugation to prevent protein precipitation due to high local concentration.

Buffer of protein samples were changed by immersing dialysis bag with protein solution filled (< 40 ml) in 4 L of stirring desired buffer for dialysis at 4 °C for at least 5 hours. Alternatively, dilution-concentration cycle of diluting protein sample with desired buffer and then concentrating back the sample by Vivaspin R15 could be used.

For determining the crystalline solubility by dialysis approach, 0.1 ml of 10 mg ml^{-1} protein samples in storage buffer (0.2 M Na_2SO_4, 20 mM sodium phosphate, pH 5.4) were dialyzed in 1 L dialysis buffer (10 mM citrate/phosphate at pH 2.5, 4.0, or 6.5) at room temperature overnight. For low protein concentration (0.5 and 1 mg ml^{-1}), 1.0 ml of samples and 2 L dialysis buffer were used instead. Precipitates (microcrystals) were collected by centrifugation and observed under Zeiss microscopes. Microcrystals solubility in high ionic strength was tested by

re-dissolving them in 0.1 ml of dialysis buffer with 1 M NaCl. Protein concentrations of supernatant were determined by measuring $O.D._{280nm}$.

For determining crystalline solubility by concentration approach, 0.1 ml of 10 mg ml^{-1} protein samples in storage buffer (0.2 M Na_2SO_4, 20 mM NaH_2PO_4, pH 5.4) were diluted in 20 ml low ionic strength buffers (10 mM citrate/phosphate at pH 6.5), and then concentrated by Vivaspin R15 until white precipitates were observed.

2.1.7 Determination of DNA and protein concentration

Double-stranded DNA and protein samples were diluted 10 folds by suitable buffer before adding to 1 cm light path cuvette for measuring $O.D._{260nm}$ and $O.D._{280nm}$. by spectrophotometer (Eppendorf) for concentration determination.

DNA concentration in μg ml^{-1} was calculated by:

$$[dsDNA] = 10 \times O.D._{260nm} \times 50$$

whereas protein concentration in mg ml^{-1} was calculated by:

$$[protein] = 10 \times (O.D._{280nm} / \varepsilon) \times M.W.$$

where ε is molar extinction coefficient in cm^{-1} M^{-1}, and M.W. is the molecular weight in g mol^{-1}.

2.1.8 Preparation of competent *E. coli* strains (DH5α, C41, and BL21(DE3)pLysS)

Single colony of E. coil strain obtained by streaking just-thaw-frozen stock on

LB agar plate with appropriate antibiotics (none for DH5α and C41; 50 μg ml^{-1} chloramphenicol for BL21(DE3)pLysS) followed by incubation at 37 °C for 16 hours, was picked to 5 ml LB with appropriate antibiotics and incubated at 37 °C with shaking for about 3 hours until $O.D._{600nm}$ reached 0.3 – 0.4. The culture was then scaled up to 100 ml LB by 1 % inoculation, and incubated in the same condition for about 2 hours until $O.D._{600nm}$ reached 0.4 – 0.5. The culture was then chilled on ice for 5 minutes and centrifuged at 6000 g for 10 minutes at 4 °C. The bacterial pellet was resuspended in 40 ml RF1 solution and chilled on ice for 5 minutes, followed by another centrifugation at 6000 g for 10 minutes at 4 °C. The bacterial pellet was then resuspended in 4 ml RF2 solution and chilled on ice for 15 minutes. After chilling, the competent cells were aliquoted in 100 ml per tube manner, immediately followed by quick frozen in liquid nitrogen. The frozen cells were stored at 70 °C until use.

2.1.9 Calibration of pH meter

pH meter (Beckman Coulter ϕ510) was calibrated according to the manufacturer's instruction. 3-point calibration was used. Before calibration, 10 ml of pH buffer standard solution of pH 4.01, 7.00, and 10.01 were equilibrated at 25 °C water bath for 10 minutes, and set the "temperature" of the pH meter at 25 °C, and then follow the instruction strictly. For pH measurement at 60 °C, the calibration was repeated and equilibrated with three temperature insensitive pH standard solutions at

60 °C, and the "temperature" of the pH meter was set at 60 °C instead.

2.2 Cloning, expression, and purification

2.2.1 Site-directed mutagenesis strategy

All single and double site-directed mutagenesis were achieved by a two-stage PCR procedure modified from the QuikChange® (Stratagene) mutagenesis protocol [102]. Before site directed mutagenesis, mutagenic mega primers were generated by primer extension using mutageneic primer (standard QuikChange primer, either coding or complementary sequence), universal gene flanking primer (either 5' coding sequence or 3' complementary sequence, dependent on the mutageneic primer used), and wild-type *T. celer L30e* cloned in expression vector pET8c (Novagen) as the template in polymerase chain reaction (PCR). The condition for primer extension was summarized in Table 2.2. The subsequence steps were the same as the standard QuikChange protocol: 40 rounds of whole plasmid linear amplification using thermo-cycler (ABI PCR System 2400 or ABI PCR System 2700) using megaprimers and wild-type *T. celer L30e* cloned pET8c as the template (Table 2.3), followed by a DpnI (NEB) digestion (Table 2.4) at 37 °C for 3 – 5 hours. The

Table 2.2. Reagents and conditions for generating mega primers

Reagents for primer extension	
Components	Volume (µl)
0.1 M Flanking forward primer (or mutagenic forward primer)	0.5
0.1 M Mutagenic reverse primer (or flanking reverse primer)	0.5
Template	0.5
10 mM dNTPs	0.5
5X HF buffer	5.0
Phusion DNA polymerase	0.3
Autoclaved dH_2O	17.7
Total Volume	25

Conditions for primer extension (using thermo-cycler)			
Cycle	Denaturation	Annealing	Extension
1^{st}	98 °C, 5 min	60 °C, 30 s	72 °C, 15 s
2^{nd} to 39^{th}	98 °C, 1 min	60 °C, 30 s	72 °C, 15 s
40^{th}	98 °C, 1 min	60 °C, 30 s	72 °C, 15 s

Table 2.3. Reagents and conditions for mega-primer-Quikchange

Reagents for mega-primer-QuikChange	
Components	Volume (µl)
Primers mixture (DNA product of primer extension)	1.0
Template	0.5
10 mM dNTPs	0.5
5X CG buffer	5.0
Phusion DNA polymerase	0.3
Autoclaved dH_2O	17.7
Total Volume	25

Conditions for mega-primer-QuikChange			
Cycle	Denaturation	Annealing	Extension
1st	98 °C, 5 min	65 °C, 30 s	72 °C, 1 min
2nd to 39th	98 °C, 1 min	65 °C, 30 s	72 °C, 1 min
40th	98 °C, 1 min	65 °C, 30 s	72 °C, 1 min

Table 2.4. Reagents for DpnI (NEB) digestion

Reagents for DpnI (NEB) digestion	
Components	Volume (µl)
PCR product	30
10X Buffer 4 (NEB)	4.0
10X BSA (NEB)	4.0
DpnI (NEB)	1.0
Autoclaved dH_2O	1.0
Total Volume	40

digested DNA was then transformed into DH5α (see Section 2.2.2) to obtain single colony, which was subsequently picked into LB broth to yield 5 ml broth culture. The plasmid DNA in the broth culture was extracted (see Section 2.1.3) and confirmed by DNA sequencing (BGI).

The clone of quintuple Arg-to-Lys variant of *T. celer L30e* was brought from Mr. Gene GmbH, and subcloned into the expression vector, pET3d (Novagen). DNA sequencing was performed to check the sequences.

2.2.2 Plasmid transformation to competent *E. coli* strain

The whole transformation was performed with aseptic techniques. An aliquot of 100 μl competent cells (see Section 2.1.8) was thawed on ice for about 5 minutes and 0.5 – 1 μl of plasmid (see Section 2.1.3) was added into the competent cells, and left for cold shock on ice for about 30 minutes. After first cold shock, the competent cells were treated with 2 minutes heat shock at 42 °C, followed by second cold shock on ice for about 15 minutes. Then 0.4 ml of LB broth was added and the competent cells were incubated at 37 °C with shaking for about 45 minutes for recovery. After recovery, competent cells were spin down and 0.4 ml of supernatant was removed. Resuspended the cells and then spread on LB plate with appropriate antibodies (100 μg ml^{-1} ampicillin for DH5α and C41; additional 50 μg ml^{-1} chloramphenicol for BL21(DE3)pLysS). For expressing NMR labeled proteins, the cells were spread on

M9 plate with appropriate antibodies instead. The plate was then incubated at 37 °C for 16 hours.

2.2.3 Expression of recombinant proteins

Plasmid-transformed single colony obtained after transformation (see Section 2.2.2) were picked to 10 ml LB broth with appropriate antibiotics (see transformation Section) and incubated at 37 °C with shaking for 16 hours as the starter culture. After incubation, 10 ml of the starter culture was inoculated to 1 L of LB broth (0.5 L in 2 L flask) with 10 ml 40% filtered glucose, 1 ml of 1 M $MgSO_4$, and appropriate antibiotics added. The scaled up culture was incubated at 37 °C with shaking for 2 – 3 hours until $O.D._{600nm}$ reached 0.6. 1 ml of 0.4 M IPTG, 1 ml of 1 M $MgSO_4$, and appropriate antibiotics were then added in 1 L culture. The culture was incubated at 37 °C for 16 hours and the cells were then collected by centrifugation at 6000 g at 4 °C for 5 minutes.

For expression of NMR $^{13}C^{15}N$-labeled protein, M9 broth was used instead of LB broth in both starter and scaled up cultures. Also, 4 g of ^{13}C glucose, 2 g of ^{15}N NH_4Cl and 4 ml of solution Q have to be additionally added in the initial scaled up culture.

2.2.4 Protein extraction from *E. coli* by sonication

Cell pellet of 1 L culture was resuspended in about 30 ml of buffer (buffer A of

following chromatography, see Section 2.2.5) with 0.75 mM PMSF (Sigma). Resuspended cells were broken by sonication on ice (Sonic & Materials, Vibra Cell) with pulse-program of pausing for 4 seconds with 1 second pause, with total sonication time of 1 minute. This sonication pulse-program was repeated for at least 10 times until clear solution was observed. Sonicated cell mixture was then centrifuged at 16000 g at 4 °C for 40 minutes and the supernatant was filtered (Schleicher & Schuell) and stored. A small portion of pellet and supernatant were subjected to SDS-PAGE to analyze if the desire protein was in soluble fraction or insoluble inclusion body.

2.2.5 Purification by chromatography

For purification of *T. celer L30e* using SP and Heparine column, pre-packed 5-ml fast-flow column (GE Health Care) was first connected to AKTA prime machine (GE Health Care) which pressure limited was set to be 0.2 MPa, and washed by about 30 ml of buffer B (1.0 M NaCl, 20 mM sodium acetate, pH 5.4) until $O.D._{280nm}$ became steady. The washing was followed by equilibration of column with about 50 ml of buffer A (20 mM sodium acetate, pH 5.4) until conductivity became steady. Then the 0.2 μm filtered solution with desire soluble protein in buffer A was loaded to the column at flow rate of 5 ml min^{-1}, leaving about 30 μl of input for SDS-PAGE analysis. Flow-through was also collected. After loading input, the

column was equilibrated with buffer A again until $O.D._{280nm}$ became steady. Bound protein was eluted using a 300 ml linear NaCl gradient ranging from 0 to 1 M NaCl. For SP column, *T. celer L30e* was approximately eluted at 0.2 M NaCl, whereas *T. celer L30e* was approximately eluted at 0.4 M NaCl for Heparine column. The elution was collected as fractions of 5 ml per tube. After elution, the column was washed with buffer B and then equilibrated with 20 % ethanol until conductivity became steady for storage at 4 °C. The collected input, flow-through, and elution fractions were analyzed by SDS-PAGE to determine relative protein concentration and purify after purification.

For purification of roughly 300 ml gel filtration column Superdex G-75 HiLoad 26/60 (GE Health Care), the pressure limit was set to 0.3 MPa, and the connected column was pre-equilibrium with about 300 ml buffer C of 0.2 M Na_2SO_4, 20 mM sodium acetate, pH 5.4, until the conductivity became steady. Maximum 5 ml of protein sample was loaded to column in flow rate of 3 ml min^{-1}. Purified *T. celer L30e* was approximately eluted at 200 ml. Elution was collected as fractions of 3 ml per tube. After elution, the column was equilibrated with 20 % ethanol until conductivity became steady for storage. Again, input and all elution fractions were analyzed by SDS-PAGE.

The purification of all *T. celer L30e* mutants were the same as that of the

wild-type, but the elution condition might have small changes.

2.2.6 Purification by refolding

Yeast L30e was expressed in inclusion bodies. The pellet collected after sonication was resuspended in 20 ml of detergent buffer (0.2 M NaCl, 1 % deoxycholic acid, 1 % NP-40). The mixture was sonicated with the same pulse-program for sonicating cell suspension (see Section 2.2.4) for once. The sonicated mixture was then centrifuged at 10000 g at 4 °C for 5 minutes. The resuspension of the pellet, sonication, and centrifugation cycle was repeated for 3 times but using 20 ml Triton/EDTA buffer (1 % Triton X-100, 1 mM EDTA) instead of detergent buffer for resuspension. After these washing cycles, the pellet collected was dissolved by 4 ml denaturing buffer (~ 4 M guanidine HCl, 10 mM sodium phosphate, pH 7.4). The mixture was then sonicated with the same pulse-program 1 time and centrifuged at 5000 g at 4 °C for 5 minutes. The supernatant was collected and the protein concentration was roughly estimated by spectrophotometer (see Section 2.1.7). The supernatant was then slowly added drop by drop to a large bulk of renaturing buffer (~ 200 ml of 0.2 M NaCl, 20 mM sodium acetate, pH 5.4) with continuous stirring, making the final protein concentration of less than 0.2 mg ml^{-1}. The precipitate in the bulk renaturing buffer was spinned down, and the supernatant (renatured Yeast L30e) was collected.

2.3 Circular dichroism experiments

2.3.1 General usage of spectropolarimeter

0.4 ml of thoroughly degassed protein sample was transferred into securely stoppered quartz cuvette with 1 mm path-length, and the cuvette was inserted in the spectropolarimeter (JASCO J-810) equipped with a Peltier-type temperature control unit. Protein sample was allowed to equilibrate at the experimental temperatures for 1 minute inside the cuvette before circular dichroism (CD) measurement. For estimating the secondary structure of the protein sample by far-UV CD, the molar ellipticity in deg cm^2 mol^{-1} of a full spectrum with wavelength ranging from 260 nm to 190 nm was recorded. For tracing the thermal- and denaturant-induced denaturation, spectra with wavelength ranging from 217 nm to 227 nm were recorded at different temperature and denaturant concentration.

2.3.2 Obtaining thermodynamic parameters from thermal denaturation

20 μM protein samples were dialyzed in experimental buffer condition for 16 hours before CD measurement. After equilibrating protein sample at 298 K for 5 minutes, all protein samples were heated from 298 K to 383 K at a heating rate of 1 K min^{-1}. The thermal denaturations were then monitored by molar ellipticity at 222 nm, and the thermal denaturation data were fitted to a two-state model[103]:

$$y_{obs} = \frac{\{(y_n + m_n T) + (y_u + m_u T) e^{-\Delta H_m / R(1/T - 1/T_m)}\}}{\{+ e^{-\Delta H_m / R(1/T - 1/T_m)}\}}$$

where y_{obs} is the observed molar ellipiticity at 222 nm; y_n and m_n are the y-intercept and slope of the pre-transition baseline; y_u and m_u are the y-intercept and slope of the post-transition baseline; R is the gas constant; T is the temperature in Kelvin. Melting temperature values (T_m) and the van't Hoof enthalpy (ΔH_m) were directed obtained from the fitting.

Thermal denaturations of yeast L30e, *T. celer L30e* E6A, K46M, E6AR92A, E6AR92M, and K46ME62Q were found to be irreversible in this condition. Since these mutants have a similar range of molar ellipiticity at 222 nm with those reversible mutants in the thermal denaturation, all of them should have their T_m values obtained at almost the same molar ellipticity ($-1.2 \pm 0.2 \times 10^5$ deg cm^2 mol^{-1}). Therefore, the apparent T_m values of the thermally irreversible mutants were estimated by finding the corresponding temperature at which the molar ellipticity was equal to $-1.2 \pm 0.2 \times 10^5$ deg cm^2 mol^{-1}.

2.3.3 Obtaining thermodynamic parameters from denaturant-induced denaturation

20 μM protein samples were equilibrated with suitable range of denaturant (0 M – 10.2 M for urea; 0 M – 7.2 M for guanidine HCl) with 0.2 M interval in the experimental buffer condition for 30 minutes at room temperature before CD

measurement. Concentration of denaturant was determined from refractive index measurements using refractometer (Leica AR200). The denaturatant-induced denaturations were then monitored by molar ellipticity at 222 nm, and the chemical denaturation data were fitted to a two-state model[103]:

$$y_{obs} = \frac{\{y_n + m_n[D]\} + \{y_u + m_u[D]\}e^{-\Delta G_{(D)}/RT}}{\{+ e^{-\Delta G_{(D)}/RT}\}}$$

where y_{obs} is the observed molar ellipiticity at 222 nm; y_n and m_n are the y-intercept and slope of the pre-transition baseline; y_u and m_u are the y-intercept and slope of the post-transition baseline; R is the gas constant; T is the temperature in Kelvin; [D] is the concentration of urea; $\Delta G_{(D)}$ is the free energy change of unfolding at [D]. The free energy change of unfolding without denaturant, $\Delta G_{(H2O)}$, was obtained by linear extrapolation model[103]:

$$\Delta G_{(D)} = \Delta G_u - m[D]$$

To obtain the temperature dependency of protein stability, ΔG_u at 298 K, 308 K, 318 K, 328 K, 338 K, and 348 K were found for *T. celer L30e* and its mutants, and at 283 K, 288 K, 293 K, 298 K, 303 K, and 308 K were found for yeast L30e.

To determine the thermodynamics parameters ΔH_m, ΔC_p, and T_m, the free energy change of unfolding (ΔG_u) at temperature ranging from 298 K to 348 K for *T. celer L30e* and 283 K to 308 K fo yeast L30e obtained from urea-induced denaturation and melting temperature (T_m) obtained from thermal denaturation were

fitted to the Gibbs-Helmholtz equation:

$$\Delta G_u = \Delta H_m\left(1-\frac{T}{T_m}\right)-\Delta C_p\left[(T_m - T)+T\ln\left(\frac{T}{T_m}\right)\right]$$

For yeast L30e, *T. celer L30e* E6A, K46M, E6AR92A, E6AR92M, and K46ME62Q, only urea-induced denaturation data were used for fitting as their thermal denaturations were irreversible. The thermal dynamic parameters obtained with and without using their thermal denaturation data for fitting agree very well with each other.

2.4 Crystallization and crystal structure refinement

2.4.1 Sparse-matrix screening and fine gridding for high quality crystals

Protein sample eluted from gel filtration column (see Section 2.2.5) was directly concentrated to about 10 mg ml^{-1}. The sparse-matrix screening was performed with using commercial screening kits Crystal Screen 1 & 2, SaltRx, PEG/ion screen (Hampton Research), as well as JCSG+ (Qiagen). Crystallization by sitting drop vapour diffusion was carried out by using Phoenix™ RE robotic system (Rigaku) to aliquot screening chemical reagent and mix with concentrated protein sample in 3 different protein-to-buffer ratios of 1:2, 1:1, and 2:1 with final volume of 0.3 μl in the 3 crystallization wells of CrystalQuick 96-well round-bottom plates (Greiner).

Crystal plates were equilibrated at 16 °C until crystals were observed in crystal viewing by light microscopes (Zeiss) or Rhombix Crystal Imager (Thermo).

The initial hits were then finely grided by changing the chemical parameters, usually pH and precipitant concentration, in the step-wise manner. In the grid search, crystallization was achieved by either sitting drop or hanging drop vapour diffusion, taking place in CrystalQuick 24-well round-bottom sitting drop or hanging drop plates (Greiner). 0.5 or 1 ml crystallization buffer was pipetted into the reservoir, and the crystallization drop was mixed in protein-to-buffer ratio of 1:1 with final volume of 4 – 10 µl in the initial grid. Crystal plates were equilibrated at 16 °C until crystals were observed in crystal viewing by light microscopes (Zeiss) or Rhombix Crystal Imager (Thermo). The crystallizability and nucleation could be altered by using additive screen kit (Hampton Research) or changing the crystallization conditions such as temperature, protein-to-buffer ratio, as well as gridding range and intervals in the subsequence grid searches.

For crystallization of wild-type *T. celer L30e* and all its single and quintuple Arg-to-Lys variants in low ionic strength conditions, their crystals were obtained either by dialyzing 10 mg ml^{-1} protein samples concentrated from elution of gel filtration column (high ionic strength) in nano-pure water, or using sitting drop vapour diffusion with the same setting as grid search reported above, but using 10

mg ml^{-1} protein sample in low ionic strength in low pH (10 mM citrate/phosphate, pH 4.0).

2.4.2 Data collection and processing

Single crystal obtained from grid search was carefully fished out using appropriate cryoloop (Hampton Research) under light microscope (Zeiss), or collect the crystals of *T. celer L30e* and its Arg-to-Lys variants obtained from dialysis by centrifugation. Then transferred the crystal to appropriate cryoprotectant (15 – 25 % glycerol or 20 – 30 % PEG 400 in crystallization buffer) and allowed for equilibration for 1 minute. The crystal was then quickly mounted to the 110 K pre-cooled goniometer (Rigaku) of X-ray diffraction setup using MicroMax™-007 HF generator (Rigaku) or FR-E+ SuperBright™ generator (Rigaku) at wavelength 1.54 Å. Datasets were acquired and collected at 110 K and detected by R-AXIS IV++ image plate detector (Rigaku) controlled by CRYSTALCLEAR, and the diffraction data were processed, merged, scaled, and reduced with programs (MOSFLM, SCALA, TRUNCATE) from the CCP4 suite[104]. The space group was determined by POINTNESS in CCP4 suite. The crystal structures were solved by molecular replacement using PHENIX (phenix.automr)[105] or CCP4 suite (MOLREP) with the previously solved wild-type *T. celer L30e* crystal structure (1H7M) as the search model. The structures were refined with REFMAC5[106] or PHENIX (phenix.refine)[105].

The refineing structures were analyzed and modified by COOT[107,108]. The refined structures were validated using WHATCHECK[109]

2.5 Nuclear Magnetic Resonance experiments and pH titration

2.5.1 NMR sample preparation

After expressing and purifying the labeled protein as described (see Section 2.2.5 and 2.2.6), the buffer condition of the protein samples were changed to the experimental buffer condition by several round dilution-concentration cycles (see Section 2.1.6). D_2O (for deuterium lock), DSS (for ^{1}H and ^{13}C internal reference), and sodium azide (for preventing bacterial growth) were added to the final concentration of 10 %, 0.5 mM, and 0.05 % respectively when the volume of the protein sample was concentrated to around 1 ml. The protein sample was further concentrated to about 0.5 ml for using 5 mm NMR tube (Norell) or about 0.2 ml for 3 mm NMR tube (Norell). The final concentration of should be around 1.0 mM and 2.8 mM for using 5 mm and 3 mm NMR tube. The protein sample was then incubated at 25 °C and 60 °C for 5 minutes. The pH values of the protein sample at specified temperatures were checked by pre-calibrated pH meter (see Section 2.1.9). Thoroughly degassed the protein sample and then transferred it to pre-washed NMR tube (washing by nano-pure water followed by acetone). Briefly spin down the

sample by hand and gassed the NMR tube with N_2 gas by several vacuum-gassing cycles. Securely stoppered the NMR tube and then sealed with parafilm.

2.5.2 Data collection and processing of NMR experiments

NMR experiments were performed by Avance™ II 500 MHz NMR spectrometer (Bruker BioSpin) controlled by TopSpin™ 2.0 (Bruker BioSpin). Adjust the experimental temperature (298 K or 333 K) and then insert the NMR tube in the NMR probe. After that, deuterium lock (for stable magnetic field), turning and matching of the probe (for better sensitivity), and shimming (for bettr magnetic field homogeneity) were performed. The center offset frequency of ^{1}H dimension was found by acquiring 1D ^{1}H-water presaturation spectrum (ZGPR), followed by adjusting the center offset frequency interactively to achieve the best water suppression. The ^{1}H 90° pulse width was determined by Bruker AU program (PULSECAL), or by traditional method of finding 360° pulse width and then divided by 4. Then, read in parameters of the pulse program of desired NMR experiment, and got the probe-specific parameters such as 90° pulse of ^{13}C and ^{15}N. Some preliminary 2D experiments were acquired and processed by TopSpin 2.0 for determination of center offset frequency, spectral width, and the digital resolution of various dimensions, as well as the signal-to-noise ratio of the experiment. For TOCSY type and NOESY type experiment, the mixing time was adjusted to optimize the signal of

the experiments. The pulse program was modified if necessarily. After setting up all acquisition parameters, optimized the receiver gain and then start the acquisition.

All multidimensional NMR data were processed with the NMRPipe[110] with optionally applying zero-filling, window-functions, p0 and p1 phasing, baseline-correction, and linear-prediction dependent on the spectrum. All processed data were analyzed using CARA, including performing sequential assignment and integration of resonance peaks.

2.5.3 pH titration

pH titration of titratable groups in the whole protein were achieved by repeating suitable NMR experiments (modified 2D H(CA)CO for side chain carboxylic groups of Asp and Glu) using different NMR samples with pH values ranging from 1.0 to 6.5 at 298 K or 333 K. The suitable chemical shifts (chemical shifts of CO in 2D H(CA)CO for Asp and Glu) were then plotted against the pH values, and fitted by different means.

For finding the microscopic pK_a values, the chemical shifts were fitted to modified Henderson-Hasselbalch equations describing from 1 to 3 transitions which were assumed to be independent[111,112]:

$$\delta(\mathrm{pH}) = \delta_b + \sum_{i=1}^{j} \frac{C_i(\delta_a - \delta_b)10^{n(pK_a - pH)}}{1 + 10^{n(pK_a - pH)}}$$

for $\sum_{i=1}^{j} C_i = 1$ and $j = 1,2,\text{or}3$

where δ(pH) was the experimental chemical shift at the given pH, δ_a and δ_b denoted the chemical shifts at the protonated and deprotonated form respectively. pK_{ai} was the microscopic pK_a value of titratable group i. C_i was the percentage contribution coefficient of titratable group i for the total absolute chemical shift ($\delta_a - \delta_b$). The Hill coefficient was allowed to set to free parameter for fitting for single transition fitting (j = 1), or for the major transition (Ci > 0.7) of complex transition fitting (j = 2 or 3). The choice of fitting model (different in degree of freedom) was determined by F-test statistics using a confidence level of 95%, and by Akaike information criterion with probability higher than 95% (see Section 2.6.6).

For fitting the chemical shift data by global fitting of titrational events (GloFTE), pKaTool (pKa system)[113,114] was used for fitting. 2 or more titratable groups were used for pair-wise fitting or charge-cluster fitting respectively. The pH step was set to 0.1 and the Monte Carlo step number was set to 300. The titration curves were calculated based on the explicit evaluation of the Boltzmann sum and the fitted pK_a values were determined by finding the pH where the titratable groups were half-protonated[113].

The model peptides Ac-GGD/EGG-NH_2 were commercially purchased (GL Biochem). The pH values in the titration of model peptides were performed by directly measuring the pH values of the accurately about 10 ml 10 mM peptide

solution of dissolving accurately about 40.2 mg and 41.6 mg lyophilized peptide powder of Ac-GGDGG-NH_2 and Ac-GGEGG-NH_2 respectively in 10 ml nano-pure water using 10 ml volumetric flask at 298 K and 333 K using pre-calibrated pH meter. Before titration with purchased 1.0 M NaOH (Sigma), the pH of the peptide solution was titrated from about 2.2 back to accurately about 2.0 by known volume of purchased 4.0 M HCl. The NaOH titration was performed with step volume of 10 µl per step, until the pH value higher than 8.0, which was far away from the expected pK_a values for both peptides. The concentrations of NaOH (mM) in the peptide solution after each step were calculated and plotted against the measured pH values. To obtain the pK_a values, the data was fitted by transformed Henderson-Hasselbalch equation with only one transition and set the Hill coefficient to be one:

$$C^{expt}_{[OH^-]} = 1000 \times \left(10^{-pH} - \frac{10^{-pK_a} \times \frac{C^{expt}_{[peptide]}}{1000}}{10^{-pH} + 10^{-pK_a}} \right) - C^{offset}_{[OH^-]}$$

where $C^{expt}_{[OH^-]}$ was the experimentally observed concentration of OH^- in mM, $C^{expt}_{[peptide]}$ was the averaged peptide concentration of the whole titration in mM, which should not be deviated more than 2 % from the initial peptide concentration. $C^{offset}_{[OH^-]}$ was the averaged OH^- concentration offset due to the initialization of pH value to 2.0 by HCl. Since the concentration and amount of HCl added were known, the experimental offset could be calculated and should not deviated more than 2 % from

he fitted values.

2.6 Computational and statistical techniques

2.6.1 Supporting in-house scripts

Throughout this study, numerous in-house scripts were written in PERL, CSH, and SH for supporting purpose such as finding the number of salt bridges based on crystal structures, processing NMR and crystal diffraction data using various programs, finding the interacting residues in crystal contact and assigning the type of the interactions, calculating the secondary shifts of each residue based on the calculated sequence corrected random coli shift, calculating the NOE ratios, calculating ideal polyproline II coordinates, simulation of pH titration curves of different buffers, and formatting ACSII text files (such as .pdb).

2.6.2 Protein model building

Models of native *T. celer L30e* and its mutants with flexible C-terminal residues were generated by MODELLER[115,116] using corresponding crystal structures as the template. 300 rounds of slow variable target function method (VTFM) optimization (automodel.library_schedule = autosched.slow), slow molecular dynamics (MD) optimization (automodel.md_level = refine.slow) were applied for refinement. For the polyproline II based denaturation state model of the quintuple Arg-to-Lys variant,

a 101-residue ideal polyproline helix II, of which the coordinates of every atom were calculated by in-house perl script, were used as the template in MODELLER. The same refinement cycle as that for native proteins was used, but the backbone coordinate were fixed and the refinement cycle was repeated fro 100 times. The final polyproline based denatured state model was the averaged model for the 100 refined models. All the models built were viewed and checked by PYMOL.

2.6.3 Calculation of solvent accessible surface area (ASA)

The solvent accessible surface area was calculated by the NACCESS program using a probe radius of 1.4 Å. Surfaces of N and O atoms were considered polar, whereas the surfaces of other atoms were nonpolar. C^{α} was excluded from calculation of side chain accessibility. The contact (buried) area of salt bridge ($\Delta ASA_{salt-bridge}$) was calculated by:

$$\Delta ASA_{salt-bridge} = \frac{\left(ASA_i^{M1} - ASA_i^{WT}\right) + \left(ASA_j^{M2} - ASA_j^{WT}\right)}{2}$$

where ASA denote the calculated accessible surface area, i and j denote the two residues involved in the salt bridge, WT, M1, and M2 denote the structures of wild-type, glycine substitution of residue j (M1), and glycine substitution of residue i (M2) respectively. The relative accessibility of each residue was calculated by comparing the calculated accessibility of the residue to that of the corresponding residue type in an extended Ala-X-Ala tripeptide (form amino acids).

2.6.4 On-line pK_a prediction servers

The pK_a values of the naïve state quintuple Arg-to-Lys variant were predicted by on-line pK_a prediction servers of PROPKA[58,59], H++[117,118], and KARLSBERG+[119,120]. The pK_a values of denatured state quintuple Arg-to-Lys variant were predicted by H++ online server.

2.6.5 Curve fitting

All the curves were fitted in the least-square fitting manner using the Levenburg-Marquardt algorithm for all by non-linear regressions. All data for fitting were not weighted and the allowable error was set to 0.1 %. The initial guessed value for each parameters were tried to set at different values to prevent local minimum solution for fitting.

2.6.6 Comparison of curve fitting models

For comparing two fitting models with different degree of freedom, Akaike's Information Criteria (AIC), which use information theory to determine the relative likelihood, and one-tailed F-test, use the extra sum of square test, were performed by in-house PERL script to distinguish which fitting model should be used. For AIC, AIC values were calculated by:

$$AIC = 2V + N\ln\left(\frac{SS}{N}\right)$$

where V was the number of parameter, N was the number of data point, and SS was

the sum of square. The more possible model has the lower AIC value.

For one-tailed F-test, F-values (F) were calculated by:

$$F = \frac{\frac{SS_n - SS_d}{df_n - df_d}}{\frac{SS_d}{df_d}}$$

where subscript n and d denoted numerator and denominator, SS was the sum of square, and df is the degrees of freedom. Be noted that numerator degrees of freedom was the degrees of freedom for the fitting model that had the larger SS. The P values (P(c)) were obtained from the corresponding F values by the F-distribution probability density function:

$$P(c) = \frac{\frac{1}{B\left(\frac{df_n}{2}, \frac{df_d}{2}\right)} \times \left(\frac{df_n c}{df_n c + df_d}\right)^{\frac{df_n}{2}} \times \left(1 - \frac{df_n c}{df_n c + df_d}\right)^{\frac{df_d}{2}}}{c}$$

where c was the confident level, B was the Beta function:

$$B(x, y) = \int_0^1 t^{x-1}(1-t)^{y-1} dt$$

for Re(x) and Re(y) > 0

In this study, confident level of 95 % was used in F-test. Fitting model that has smaller SS was chosen if the calculated P value was smaller than 0.05.

2.6.7 Detection of outliners

Huber's method was used for detection of outliners in the data sample of N data point and was carried out by in-house PERL script. Data was progressively

transformed by winsorization. The initial values of the mean ($\hat{\mu}_0$) and standard deviation ($\hat{\sigma}_0$) were simply the median (m_0) and the standard deviation (σ_0) of the data. If a data point x_i had a value falls outside the range of $\hat{\mu}_0 \pm 1.5\hat{\sigma}_0$,

$$\tilde{x}_i = \hat{\mu}_0 \pm 1.5\hat{\sigma}_0$$

where $\tilde{x}_i$ was the new x_i used in the next round of winsorization. otherwise,

$$\tilde{x}_i = x_i$$

For the winsorized mean ($\hat{\mu}_1$) and the standard deviation ($\hat{\sigma}_1$) was calculated by:

$$\hat{\mu}_1 = \frac{\sum_{i=1}^{N} \tilde{x}_i}{N}, \text{ and } \hat{\sigma}_1 = 1.134 \times \sigma(\tilde{x}_i)$$

The factor 1.134 was derived from the normal distribution, given a value 1.5 for the multiplier used in the winsorization process. Winsorization were performed for at least 100 rounds until the Huber's mean ($\hat{\mu}_{Hub}$) and Huber's standard deviation ($\hat{\sigma}_{Hub}$) were converged.

2.6.8 Detection of monotonic trends

Wald-Wolfowitz run tests were carried out by in-house PERL script to detect if the coupling free energies of various ion pairs had any trends with respect to temperature ($\Delta\Delta G_u(T)$). If $\Delta\Delta G_u(T)$ was a stationary random process (coupling free energy has trend with respect with temperature), the number of runs, which was defined as a sequence of identical observations that is followed or preceded by a different observation or no observation at all, was a random variable with mean value

(μ_0) and sum of square (SS_0) of:

$$\mu_0 = \frac{2n_{pos}n_{neg}}{n_{pos}+n_{neg}} + 1 \quad \text{and} \quad SS_0 = \frac{2n_{pos}n_{neg} \times (2n_{pos}n_{neg} - n_{pos} - n_{neg})}{(n_{pos}+n_{neg})^2 (n_{pos}+n_{neg}-1)}$$

where n_{pos} and n_{neg} was the number of points higher and lower than μ_0 respectively. Significant difference between observed number of runs (r) and μ_0 indicated nonstationarity because of the possible presence of a trend in $\Delta\Delta G_u(T)$. The Z-score of the difference was calculated by:

$$Z = \left| \frac{r - \mu_0}{\sqrt{SS_0}} \right|$$

and the P value was calculated by:

$$P = \frac{1}{2(1 - (1 + Z(0.050 + Z(0.021 + Z(0.003)))))^6}$$

With using 95% confident level, $\Delta\Delta G_u(T)$ was regarded as stationary if the P value was larger than 0.05. Be noted the coupling free energy was assumed to have no seasonality with respect to temperature, therefore the 6-point dataset of each ion pair was regarded as stationary if r is larger than 3.

2.7 Reagents and Buffers

In this section, recipe and preparation method of reagents and buffers are listed.

2.7.1 Buffers for preparation of competent cells

RF1

30 mM potassium acetate, 100 mM $RbCl_2$, 10 mM $CaCl_2$, 50 mM $MnCl_2$, and

15 % glycerol. The pH was adjusted to 5.8 by acetic acid without back titration. The solution was then sterilized by 0.2 μm filtration.

RF2

10 mM MPOS, 75 mM $CaCl_2$, 10 mM $RbCl_2$, and 15 % glycerol. The pH was adjusted to 6.5 by KOH without back titration. The solution was then sterilized by 0.2 μm filtration.

2.7.2 Buffers for agarose gel electrophoresis

50X Tris-acetate-EDTA (TAE) stock solution

242 g L^{-1} Tris base, 57.1 ml L^{-1} glacial acetic acid, and 37.2 g L^{-1} $Na_2EDTA\,2H_2O$, and stored the solution at room temperature.

6X agarose gel loading buffer

0.25 % (w/v) bromophenol blue, 40 % (w/v) sucrose, and 60 mM EDTA (pH 8.0), and stored the solution at 4 °C.

2.7.3 Buffers for SDS-PAGE

10X SDS-electrophoresis buffer stock solution

0.125 M Tris base, 1.92 M glycine, and 0.5 % (w/v) SDS, and stored the solution at 4 °C.

2X SDS-PAGE loading buffer

200 mM DTT, 4 % SDS, 0.2 % bromophenol blue, 20 % glycerol, and 100 mM

Tris-HCl, pH 6.8, and stored the solution at -20 °C.

4X SDS resolving gel buffer

181.7 g L^{-1} Tris base, and 8 g L^{-1} SDS. Adjusted the pH to 8.8 with HCl, and stored the solution at 4 °C.

4X SDS stacking gel buffer

60.6 g L^{-1} Tris base, and 8 g L^{-1} SDS. Adjusted the pH to 6.8 with HCl, and stored the solution at 4 °C.

Conventional Coomassie staining solution

0.15 % Coomassie Brilliant Blue R-250, 30 % (v/v) ethanol, and 10 % (v/v) acetic acid, and store the solution at room temperature.

Conventional destaining solution

25 % ethanol and 10 % glacial acetic acid, and stored the solution at room temperature.

Sensitive colloidal Coomassie staining solution

5 % aluminium sulfate-(14-18)-hydrate, 10 % ethanol, 2% orthophosphoric acid and 0.02 % (w/v) Coomassie Brilliant Blue G-250, and store the solution at room temperature.

Media for bacterial culture

Luria-Bertani (LB) broth

400 g kg^{-1} NaCl, 400 g kg^{-1} peptone, and 200 g kg^{-1} yeast extract, and store the powder mixture at room temperature. For liquid LB broth, dissolved 25 g L^{-1} of the powder mixture and optionally adjust the pH to 7.0. The solution was sterile by autoclave at 121 °C for 20 minutes, and stores the solution at room temperature.

M9ZB broth

1 g L^{-1} NH_4Cl, 3 g L^{-1} KH_2PO_4, 6 g L^{-1} Na_2HPO_4, 5 g L^{-1} NaCl, and 10 g L^{-1} bacto-trypton and optionally adjust the pH to 7.0. The solution was sterile by autoclave at 121 °C for 20 minutes, and stores the solution at room temperature.

M9 broth

6 g L^{-1} Na_2HPO_4 $7H_2O$, 3 g L^{-1} KH_2PO_4 (anhydrous), 0.5 g L^{-1} NaCl and adjust the pH to 7.0. The solution was sterile by autoclave at 121 °C for 20 minutes, and stores the solution at room temperature.

Before using the M9, dissolved vitamin mixture (Table 2.5), 2 – 4 g L^{-1} ^{13}C glucose, 1 g L^{-1} ^{15}N NH_4Cl in about 50 – 100 ml of autoclaved M9 solution and sterile by 0.2 μm filtered. Mixed all solutions and added 4 ml L^{-1} 0.2 μm filtered Solution Q (Table 2.6), 1 ml L^{-1} 0.2 μm filtered 1.0 M $MgSO_4$, and appropriate antibiotics.

Table 2.5. Vitamin mix for expression in M9 medium

Reagents for vitamin mix for 1 L M9 medium	
Components	Weight (mg)
Thamine	10
d-bioin	2
Chloline chloride	2
Folic acid	2
Niacinamide	2
D-pantohenic acid	2
Pyridoxal	2
Riboflavin	2
Total Weight	24

Table 2.6. Solution Q for expression in M9 medium

Reagents for mixing 100 ml Solution Q	
Components	Weight (mg)
$FeCl_2 4H_2O$	500.0
$CaCl_2 2H_2O$	18.4
H_3BO_3	6.4
$CoCl_2 6H_2O$	1.8
$CuCl_2 2H_2O$	0.4
$ZnCl_2$	34.0
$Na_2MoO_4 2H_2O$	60.5
$MnCl_2 4H_2O$	4.0
5 N HCl	0.8 ml
Total Weight	625.5

Agar medium

Mix 15 g L^{-1} bacto-agar in the desired medium and autoclaved the mixture at 121 °C for 20 minutes. After cooled the solution down to about 50 °C (hand hot), the solution was aliquot in about 20 ml per plate manner. The plates were store at room temperature.

Chapter 3: Electrostatic interactions improve protein thermostability by up-shifting protein stability curve

3.1 Introduction

Protein stability curve describing the temperature dependency of ΔG_u can be obtained by using Gibbs-Helmholtz equation, where thermodynamic parameters T_m, ΔH_m, and ΔC_p are needed for the description. Comparison of protein stability curves and the thermodynamic parameters of thermophilic and mesophilic proteins revealed the elevated thermostability of thermophilic proteins is obtained in general by up-shifting and broadening of protein stability curves[100], which can be indicated by the increase of maximal ΔG_u and reduced ΔC_p respectively. However, the origin of the up-shift of protein stability curve is not yet clearly known.

Electrostatic interaction is one of the possible candidates. The contribution of sat bridge on protein thermal stability has been first suggested by Perutz and Raidt[94,95] on the basis of sequence comparisons between mesophilic and thermophilic proteins. This suggestion has been reinforced by more recent researches which showed thermophilic proteins have generally more surface charges and ion-pairs than their mesophilic homologues[56,92,93,96-98]. Despite of the counter argument that the

large desolvation penalty[37,38,57] and entropic cost[43,44,46] of salt bridge formation cannot be fully compensated by the enhanced coulombic interactions due to salt bridge, supporting experimental evidences that salt bridge contributes to thermal stability of protein has been accumulated[48,50-52]. The stabilizing effect of salt bridge therefore improves thermostability of proteins.

The change in solvent accessible surface area upon unfolding (ΔASA) and residual structure in protein denatured state have been suggested to be the origin of the reduced ΔC_p in thermophilic protein[121-126]. However, they are failed to account for the large reduction of ΔC_p in thermophilic proteins[99,127-129]. Our group has previously first experimentally demonstrated electrostatic interaction contributes to ΔC_p[99], indicating the difference of protein stability curves among thermophilic and mesophilic homologues is originated from electrostatic interactions.

As mentioned above, protein stability curve reflects the temperature dependency of protein stability. Therefore, to deepen the knowledge on how electrostatic interaction contribute to thermostability, the effect of salt bridges on protein stability at temperature ranging from 25 °C to 75 °C, as well as the effect on another thermodynamic parameter ΔC_p, were examined. A well studied thermophilic protein *Thermococcus celer L30e* provided us with an excellent model for this investigation. Two salt bridges E6/R92, and K46/E62, as well as a long range ion pair E90/R92

(separation distance 10.8 Å), located in two separated large charge clusters of *T. celer L30e* (Figure 3.1) were chosen for investigation. While the coupling free energy of salt bridges was analyzed by double mutant cycles of free energy, the effect of salt bridge on ΔC_p was investigated by a novel analysis originated from double mutant cycle. By comparing the protein stability of *T. celer L30e* and its mesophilic yeast homolog, we have demonstrated salt bridge is the major factor that contributes to the large up-shift in the protein stability curve of thermophilic proteins. In addition, we have also shown the salt bridge interaction is temperature independent, indicating temperature would not affect the salt-bridge interactions.

3.2 Results

3.2.1 Removal of salt bridge forming charged residues in charge clusters reduces protein stability of *T. celer L30e*

To demonstrate the importance of salt bridge to protein stability, four charged residues of two salt bridges located in two charge clusters (E6, R92, K46, E62), and an isolated charge (E90) were mutated to generate 14 single and double charge-to-neutral mutants (Figure 3.1). T_m values of the mutants were determined by thermal denaturation at low ionic strength condition (Figure 3.2). Removal of charged residues involved in salt bridge (E6, R92, K46, E62) was found to induce large decrease in T_m values (ΔT_m), which ranged from -4.7 °C to -13.7 °C (Table 3.1).

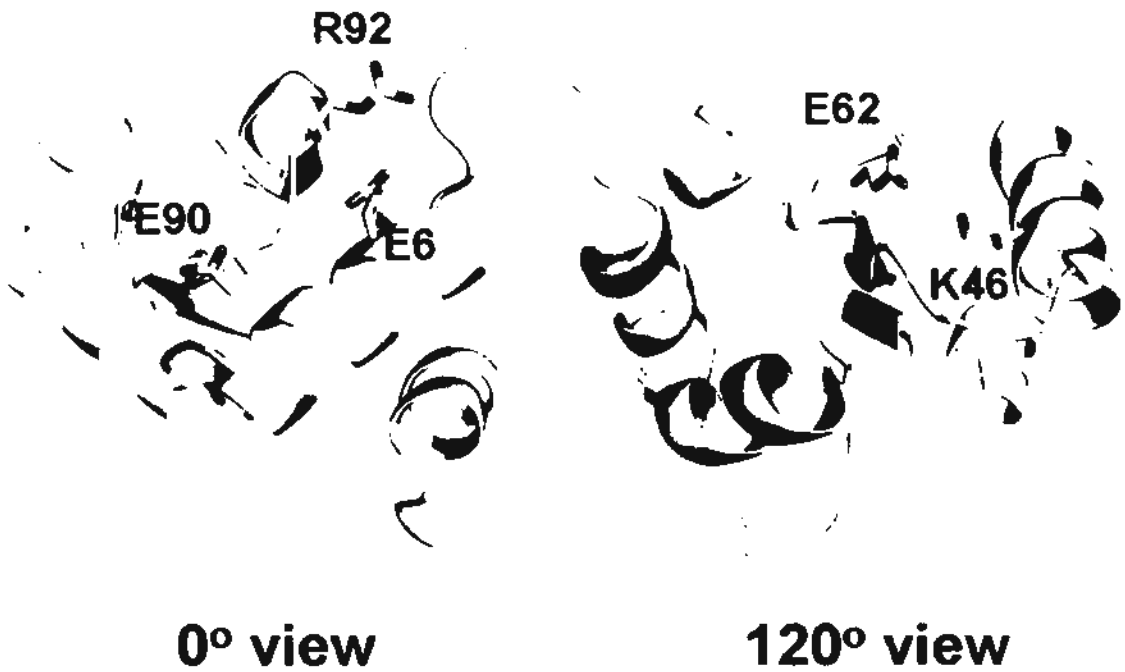


Figure 3.1. Location of three ion pairs. Salt bridge E6/R92 and the long range ion pair E90/R92 are located in charge cluster 1 (0° view). Salt bridge K46/E62 is located in another charge cluster 2 (120° view).

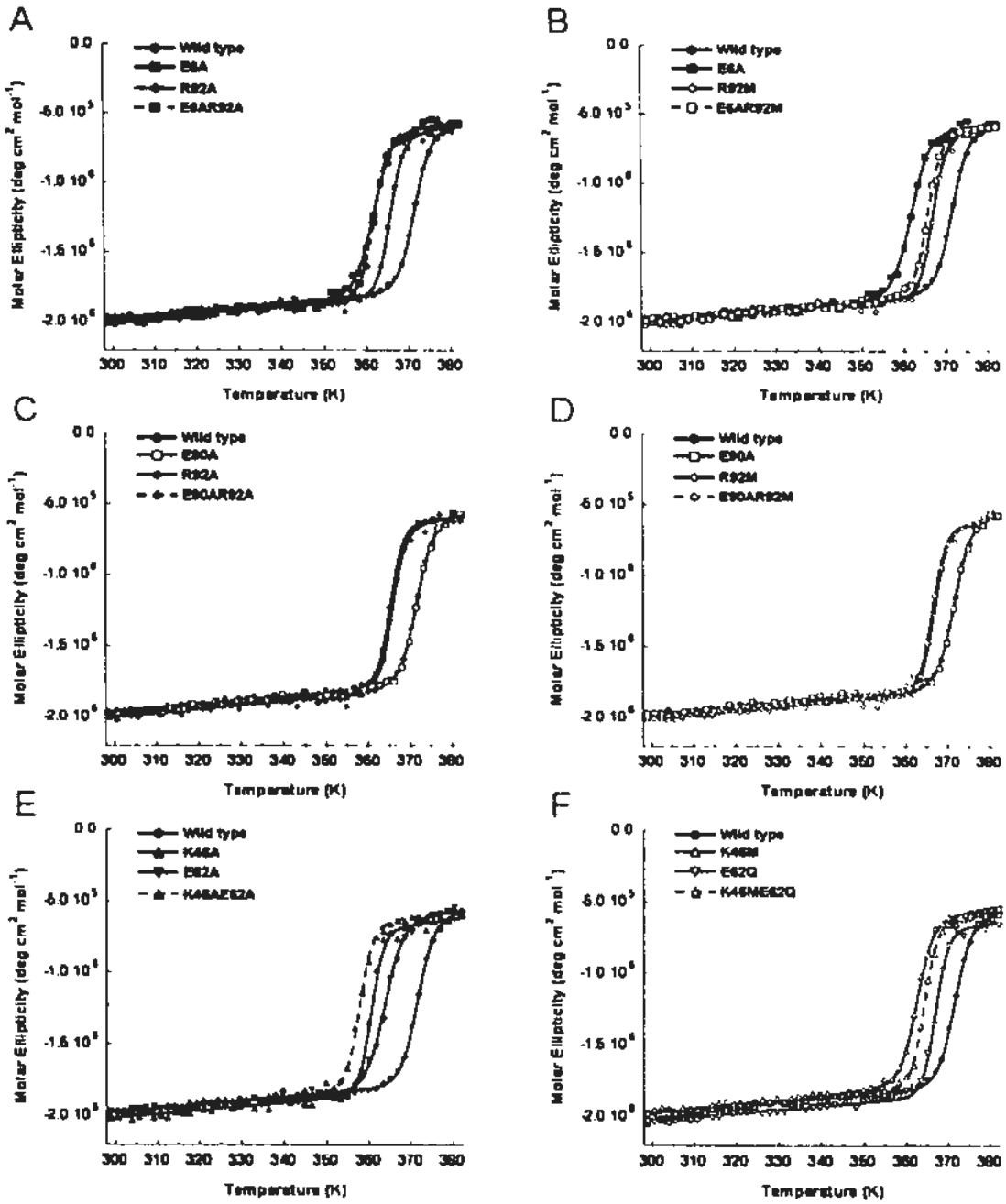


Figure 3.2. Thermal denaturation of *T. celer L30e* and its mutants. Thermal denaturation curves of protein samples involved in charge-to-Ala double mutant cycle of (A) Glu-6 and Arg-92, (C) Glu-90 and Arg-92, and (E) Lys-46 and Glu-62, as well as those of protein samples involved in charge-to-neutral double mutant cycle of (B) Glu-6 and Arg-92, (D) Glu-90 and Arg-92, and (F) Lys-46 and Glu-62, were plotted. The symbols and lines used were consistent in all panels: wild-type *T. celer L30e* (filled circles), E6A (filled squares), E90A (open squares), R92A (filled diamonds), R92M (open diamonds), K46A (filled triangles), K46M (open triangles), E62A (filled inverted triangles), and E62Q (open inverted triangles) were fitted by continuous lines, while the double mutants E6AR92A (filled squares), E6AR92M (open squares), E90AR92A (filled diamonds), E90AR92M (open diamonds), K46AE62A (filled triangles), and K46ME62Q (open triangles) were fitted by dashed line. All data were collected in low ionic strength (10 mM sodium acetate, pH 5.4).

Table 3.1. Thermodynamic parameters of yeast L30e, *T. celer L30e* and its mutants

Protein sample	Thermal denaturation[a]		Gibbs-Helmholtz analysis[b]		
	T_m (K)	ΔT_m (K)	ΔH_m (kJ mol^{-1})	ΔC_p (kJ mol^{-1} K^{-1})	$\Delta\Delta C_p^{total}$ (kJ mol^{-1} K^{-1})
Mesophilic homologues:					
Yeast L30e	328.1 ± 0.1*	-43.5 ± 0.1	316 ± 19*	7.9 ± 0.5	2.6 ± 0.3
Thermophilic homologues:					
T. celer L30e	371.6 ± 0.1	---	382 ± 9	5.3 ± 0.3	---
E6A	361.9 ± 0.1*	-9.7 ± 0.1	360.± 9*	6.1 ± 0.2	0.9 ± 0.1
K46A	360.7 ± 0.1	-10.9 ± 0.1	391 ± 13	6.8 ± 0.1	1.5 ± 0.1
K46M	362.5 ± 0.3*	-9.1 ± 0.2	412 ± 12	6.8 ± 0.3	1.5 ± 0.2
E62A	363.8 ± 0.1	-7.8 ± 0.1	354 ± 12	5.7 ± 0.2	0.5 ± 0.1
E62Q	366.9 ± 0.1	-4.7 ± 0.1	378 ± 7	5.8 ± 0.2	0.5 ± 0.1
E90A	371.5 ± 0.1	-0.1 ± 0.1	374 ± 6	5.4 ± 0.1	0.2 ± 0.1
R92A	365.9 ± 0.1	-5.7 ± 0.1	424 ± 14	6.8 ± 0.2	1.5 ± 0.1
R92M	366.9 ± 0.2	-4.6 ± 0.1	423 ± 13	6.5 ± 0.2	1.2 ± 0.1
E6AR92A	362.0 ± 0.1*	-9.6 ± 0.1	395 ± 8	6.7 ± 0.2	1.4 ± 0.1
E6AR92M	363.5 ± 0.2*	-8.1 ± 0.2	387 ± 7	6.5 ± 0.3	1.2 ± 0.1
E90AR92A	365.4 ± 0.1	-6.1 ± 0.1	408 ± 12	6.6 ± 0.1	1.3 ± 0.1
E90AR92M	366.5 ± 0.1	-5.1 ± 0.1	412 ± 10	6.5 ± 0.3	1.3 ± 0.2
K46AE62A	357.9 ± 0.1	-13.7 ± 0.1	354 ± 12	6.2 ± 0.1	0.9 ± 0.1
K46ME62Q	364.3 ± 0.3*	-7.3 ± 0.2	398 ± 8	6.3 ± 0.2	1.0 ± 0.1

[a] The T_m values in low ionic strength condition (10 mM sodium acetate, pH 5.4) are shown. Thermal denaturations of yeast L30e, T. celer L30e E6A, K46M, E6AR92A, E6AR92M, and K46ME62Q were irreversible in this condition (indicated by asterisks). Their apparent T_m and ΔH_m values were estimated as described in Materials and Methods.

[b] ΔC_p values were obtained by fitting ΔG_u obtained from urea-induced denaturation to the Gibbs-Helmholtz equation.

On the other hand, charge-to-Ala mutation of isolated residue E90 showed negligible effect on T_m value (ΔT_m -0.1 °C).

The free energy of unfolding (ΔG_u) of *T. celer L30e* and its single and double charge-to-neutral variants at temperature ranging from 25 °C to 75 °C were determined by urea-induced denaturation at low ionic strength condition (Figure 3.3) and summarized in Table 3.2. The chemical denaturation of all samples at all temperatures studied followed a reversible two-state transition. All protein samples with the same concentration had a similar value of molar ellipticity at 222 nm in the urea-induced denatured state of the CD spectrum at all temperatures studied, suggesting those charge-to-neutral mutations did not create or remove residual structures if any in the urea-induced denatured state. Although removal of single charge reflected a combination of effects on ΔG_u, we could still observe that the removal of all salt bridge involving residues in general resulted in the decrease in ΔG_u by about 6 kJ mol^{-1}.

3.2.2 Charge-to-neutral mutations of salt bridge involving residues increase in ΔC_p value

The ΔG_u values of *T. celer L30e* and its charge-to-neutralmutants at temperatures ranging from 25°C to 75°C were fitted to the Gibbs-Helmholtz equation to determine ΔC_p and ΔH_m (Figure 3.4), and the resultant thermodynamic parameters

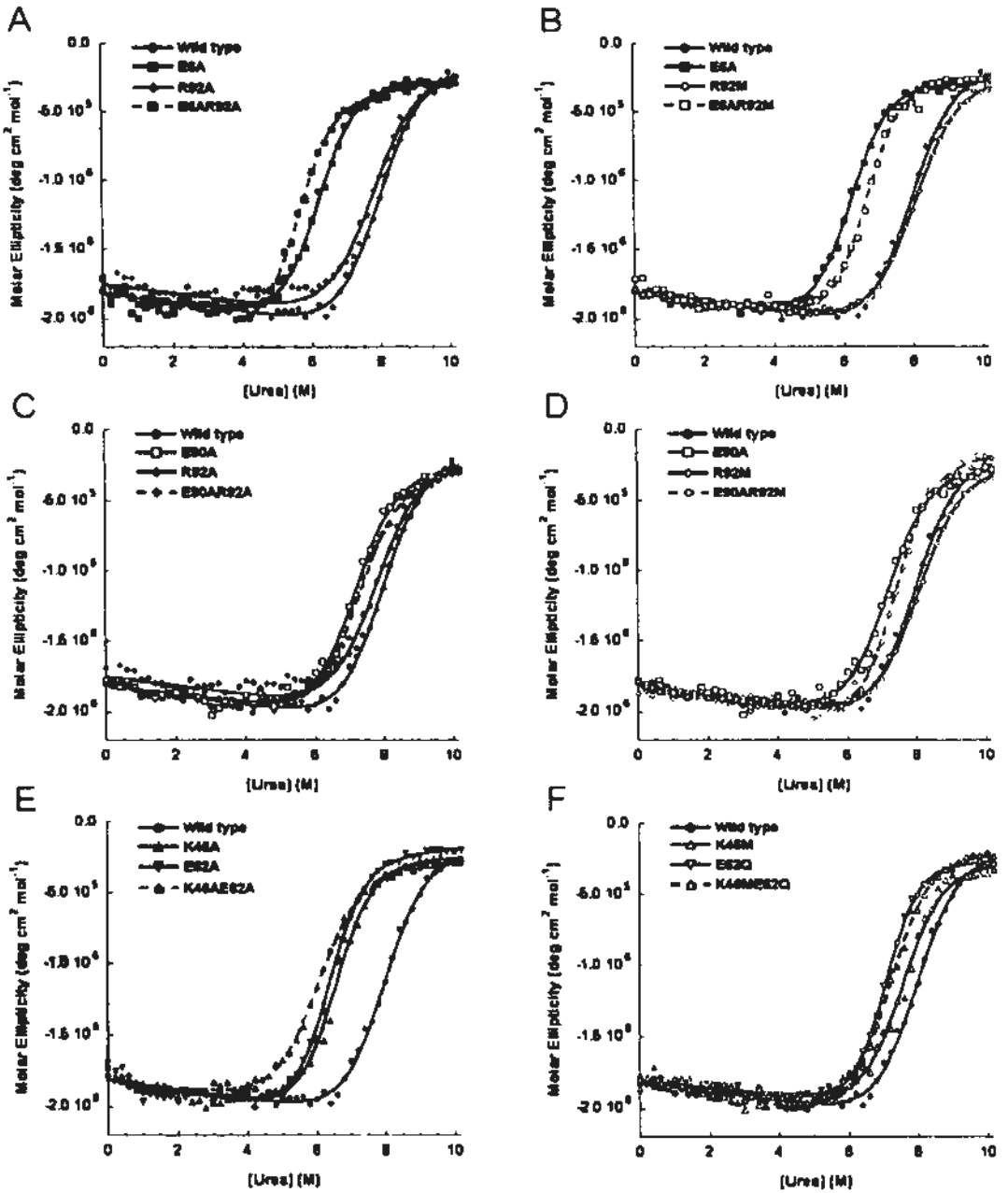


Figure 3.3. Uread-induced denaturation of *T. celer L30e* and its mutants. Urea-induced denaturation curves of protein samples involved in charge-to-Ala double mutant cycle of (A) Glu-6 and Arg-92, (C) Glu-90 and Arg-92, and (E) Lys-46 and Glu-62, as well as those of protein samples involved in charge-to-neutral double mutant cycle of (B) Glu-6 and Arg-92, (D) Glu-90 and Arg-92, and (F) Lys-46 and Glu-62, were plotted. The symbols and lines used were consistent in all panels: wild-type *T. celer L30e* (filled circles), E6A (filled squares), E90A (open squares), R92A (filled diamonds), R92M (open diamonds), K46A (filled triangles), K46M (open triangles), E62A (filled inverted triangles), and E62Q (open inverted triangles) were fitted by continuous lines, while the double mutants E6AR92A (filled squares), E6AR92M (open squares), E90AR92A (filled diamonds), E90AR92M (open diamonds), K46AE62A (filled triangles), and K46ME62Q (open triangles) were fitted by dashed line. All data were collected in low ionic strength (10 mM sodium acetate, pH 5.4).

Table 3.2. ΔG_u of *T. celer L30e* and its charge-to-neutral mutants

Protein	$\Delta G_{u(298K)}$ kJ mol^{-1}	$\Delta G_{u(308K)}$ kJ mol^{-1}	$\Delta G_{u(318K)}$ kJ mol^{-1}	$\Delta G_{u(328K)}$ kJ mol^{-1}	$\Delta G_{u(338K)}$ kJ mol^{-1}	$\Delta G_{u(348K)}$ kJ mol^{-1}
T. L30e	34.9 ± 0.5	35.0 ± 0.5	32.4 ± 0.5	30.8 ± 0.4	26.6 ± 0.4	20.9 ± 0.5
E6A	27.5 ± 0.3	27.7 ± 0.3	25.5 ± 0.3	23.3 ± 0.3	19.0 ± 0.3	14.4 ± 0.3
K46A	29.8 ± 0.3	29.7 ± 0.4	26.7 ± 0.4	24.2 ± 0.3	19.0 ± 0.3	13.3 ± 0.4
K46M	32.9 ± 0.4	32.3 ± 0.4	29.5 ± 0.4	27.2 ± 0.4	22.4 ± 0.3	17.4 ± 0.2
E62A	28.5 ± 0.3	28.7 ± 0.3	25.7 ± 0.3	23.3 ± 0.3	18.0 ± 0.2	12.1 ± 0.3
E62Q	31.5 ± 0.3	32.1 ± 0.4	29.0 ± 0.4	27.7 ± 0.3	23.1 ± 0.3	18.0 ± 0.3
E90A	32.7 ± 0.4	32.7 ± 0.4	29.8 ± 0.5	29.0 ± 0.4	24.2 ± 0.4	19.8 ± 1.5
R92A	33.9 ± 0.5	33.6 ± 0.5	31.3 ± 0.6	30.8 ± 0.5	25.0 ± 0.4	19.1 ± 0.3
R92M	35.2 ± 0.5	35.4 ± 0.4	32.7 ± 0.4	31.3 ± 0.5	26.0 ± 0.4	19.6 ± 0.3
E6AR92A	28.4 ± 0.3	28.5 ± 0.4	26.2 ± 0.4	25.1 ± 0.3	19.9 ± 0.3	15.3 ± 0.3
E6AR92M	29.7 ± 0.3	30.0 ± 0.4	27.7 ± 0.4	26.1 ± 0.3	21.4 ± 0.3	16.0 ± 0.3
E90AR92A	32.5 ± 0.4	32.0 ± 0.5	29.7 ± 0.5	29.4 ± 0.5	22.9 ± 0.4	18.3 ± 0.3
E90AR92M	33.2 ± 0.3	33.3 ± 0.4	30.3 ± 0.4	29.7 ± 0.4	24.7 ± 0.3	19.8 ± 0.3
K46AE62A	27.0 ± 0.3	27.0 ± 0.3	24.0 ± 0.3	21.2 ± 0.3	14.5 ± 0.3	8.2 ± 0.5
K46ME62Q	32.6 ± 0.4	32.6 ± 0.4	30.1 ± 0.4	27.8 ± 0.4	22.4 ± 0.3	17.3 ± 0.2

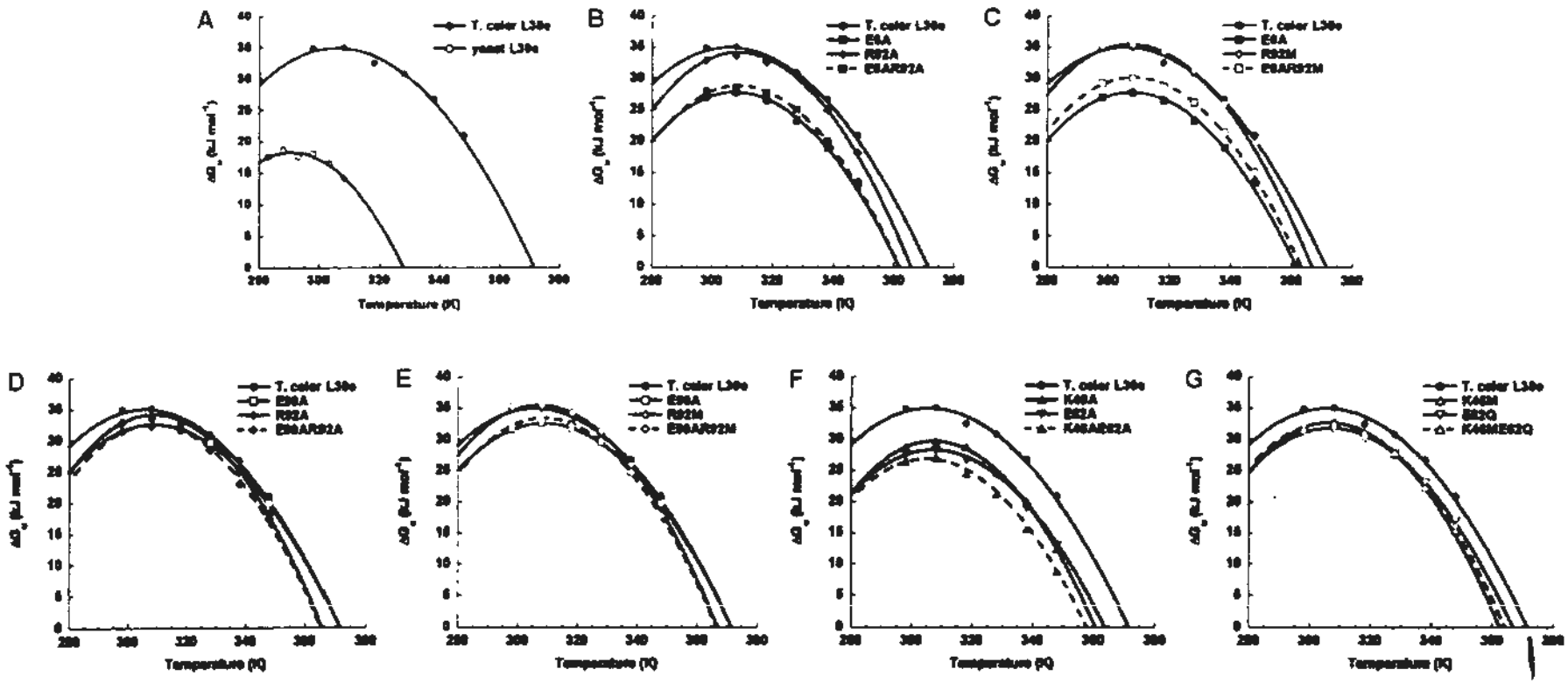


Figure 3.4. Protein stability curves of yeast L30e, T. celer L30e and its mutants. Free energy change of unfolding (ΔG_u) at 283 K, 288 K, 293 K, 298 K, and 333 K for yeast L30e, and 298 K, 308 K, 318 K, 328 K, 338 K, and 348 K were derived from urea-induced denaturation in low ionic condition. Tm values were determined by thermal denaturation (other figure) in the same condition. Lines represent the best fits of DGu obtained at various temperatures (indicated by symbols) to the Gibbs-Helmholtz equation, where continuous lines represent wild-type proteins and single mutants, and dashed lines represent double mutants. All symbols and lines represent the same as those in thermal and urea-induced denaturation curves (Figure 2). The fitted parameters, DCp and DHm, are summarized in Table 2. (A) The protein stability curve of wild-type T. celer L30e is compared to those of yeast L30e. Protein stability curves for the six double mutant cycles using (B) E6AR92A, (C) E6AR92M, (D) E90AR92A, (E) E90AR92M, (F) K46AE62A, and (G) K46ME62Q as the reference states are also shown.

were summarized in Table 3.1. The ΔC_p value of wild type *T. celer L30e* was 5.3 ± 0.3 kJ mol^{-1} K^{-1}, which was consistent with the previous result[99]. For the single charge-to-neutral mutants (E6A, R92A/M, K46A/M, E62A/Q) and double charge-to-neutral mutants (E6AR92A/M, K46A/ME62A/Q) which broke one of the two salt bridges, their ΔC_p values were approximately increased by about 1.1 ± 0.1 kJ mol^{-1} K^{-1} to 6.4 ± 0.1 kJ mol^{-1} K^{-1}. Double mutants of long range ion pair (E90AR92A/M) also increased its ΔC_p values by 1.2 ± 0.1 J mol^{-1} K^{-1} to the value of 6.5 ± 0.1 kJ mol^{-1} K^{-1} because the charge-to-neutral mutation of Arg-92 broke its salt bridge with Glu-6. For the mutation of isolated charge Glu-90 which did not form any salt bridge (E90A), it only marginally increased it ΔC_p value by 0.2 ± 0.1 kJ mol^{-1} K^{-1}.

3.2.3 *T. celer L30e* has significantly up-shifted protein stability curve when compared to that of yeast L30e

The ΔG_u of *T. celer L30e* and its mesophilic yeast homolog was determined by urea-induced denaturation in low ionic strength condition at temperatures ranging from 10 °C to 75 °C. The chemical denaturations followed a reversible two-state transition. ΔG_u values at various temperatures were fitted to Gibbs-Helmholtz equation to obtain ΔC_p, ΔH_m, and the T_m values of the proteins (Figure 3.4). The T_m of *T. celer L30e* (98.6 °C) was 43.5 °C higher than that of the yeast homolog (55.1

°C), whereas the ΔC_p of *T. celer L30e* (5.3 ± 0.3 kJ mol^{-1} K^{-1}) was 2.7 kJ mol^{-1} K^{-1} smaller than that of the yeast homolog (7.9 ± 0.7 kJ mol^{-1} K^{-1}). Comparison of the protein stability curves of T. celer and yeast L30e revealed that the former has a much higher T_m because of the significant up-shifting (by about 16.8 kJ mol^{-1} at 25 °C) and broadening (ΔC_p value reduced by 2.6 ± 0.6 kJ mol^{-1} K^{-1}) of the protein stability curve.

3.2.4 Coupling free energies of ion pairs are stabilizing at all experimental temperatures

To determine the temperature dependency of the coupling free energy of ion pairs ($\Delta\Delta G_u^{coupling}$), $\Delta\Delta G_u^{coupling}$ values of two salt bridges (E6/R92 and K46/E62) and one long range ion pair with separation distance of 10.8 Å (E90/R92) were obtained by double mutant cycles at temperatures ranging from 25 °C to 75 °C. To improve the reliability of the $\Delta\Delta G_u^{coupling}$ values found, two different double mutant cycles were applied to each of the three ion pairs (Figure 3.5). No significant difference can be observed between the mean values of coupling free energies found by different double mutant cycles according to ANOVA. The coupling free energies of ion pairs at different temperatures were summarized in Table 3.3.

$\Delta\Delta G_u^{coupling}$ values of salt bridges were found not to vary with temperatures. The mean $\Delta\Delta G_u^{coupling}$ values of E6/R92 and K46/E62 at the six experimental

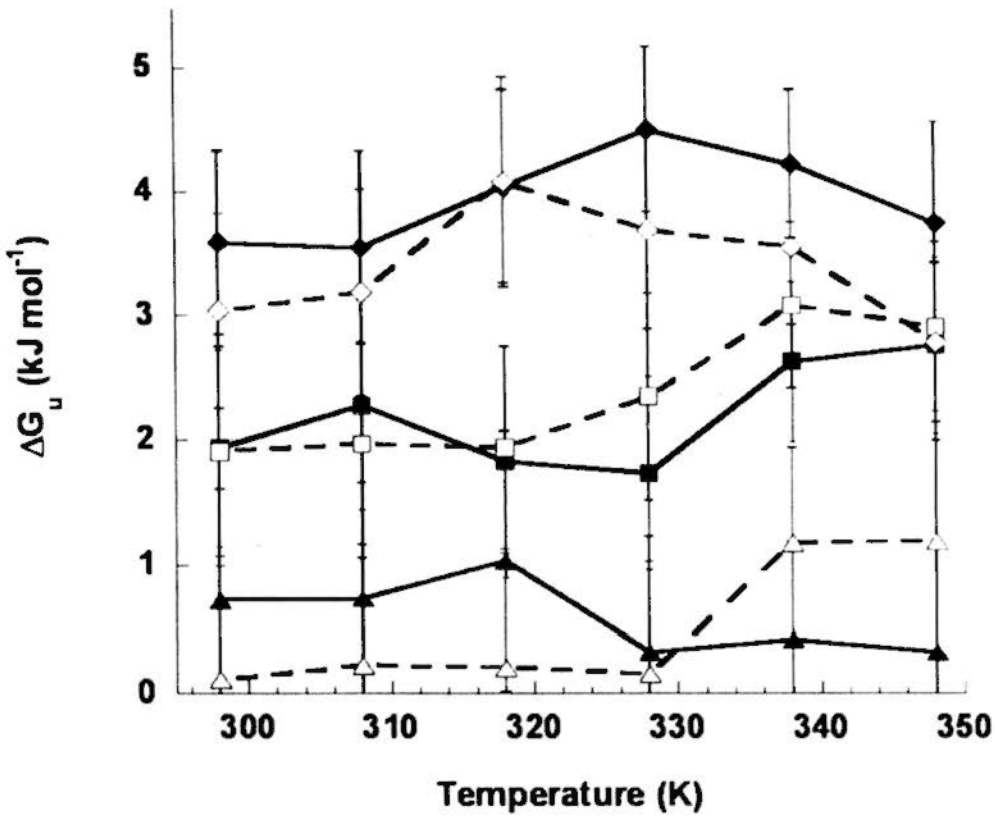


Figure 3.5. Coupling free energy of ion pairs. Coupling free energy of ion pairs were found by double mutant cycles using reference state of E6A/R92A (filled squares), E6A/R92M (open squares), E90A/R92A (filled triangles), E90A/R92M (open triangles), K46AE62A (filled diamonds), and K46ME62Q (open diamonds) at temperatures ranging from 25 °C to 75 °C.

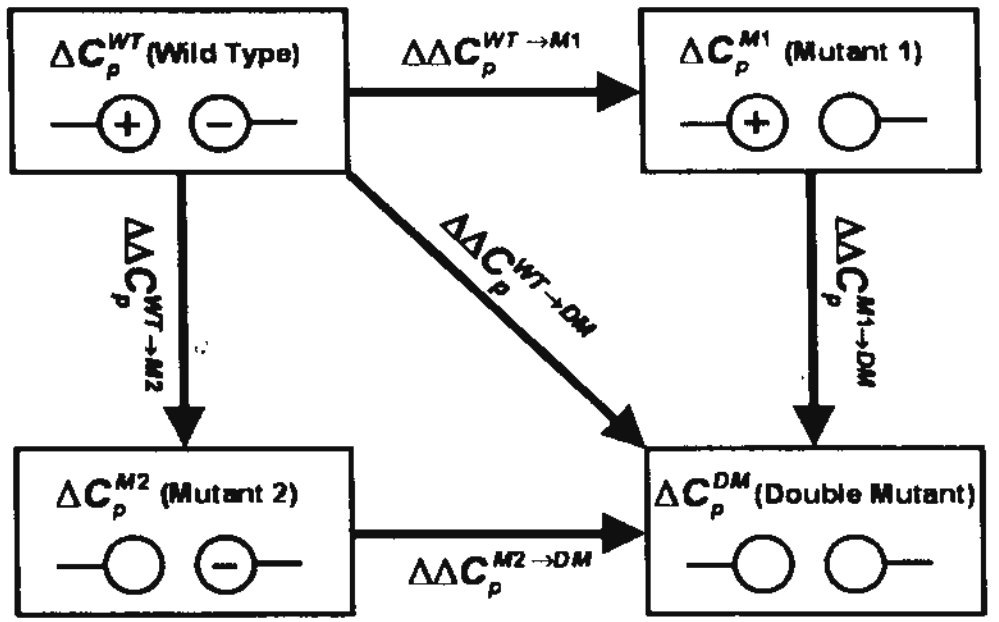
Table 3.3. Contribution of ion pairs to ΔG_u and ΔC_p of *T. celer L30e*

Temperature (K)	E6A / R92A	E6A / R92M	E90A / R92A	E90A / R92M	K46A E62A	K46M / E62Q
Coupling free energy ($\Delta\Delta G_u^{coupling}$) (kJ mol^{-1})						
298	1.9 ± 0.8	1.9 ± 0.8	0.7 ± 0.9	0.1 ± 0.9	3.6 ± 0.7	3.0 ± 0.8
308	2.3 ± 0.8	2.0 ± 0.8	0.7 ± 0.9	0.2 ± 0.8	3.6 ± 0.8	3.2 ± 0.8
318	1.8 ± 0.9	1.9 ± 0.8	1.0 ± 1.0	0.2 ± 0.9	4.0 ± 0.8	4.1 ± 0.9
328	1.7 ± 0.8	2.4 ± 0.8	0.3 ± 0.9	0.2 ± 0.9	4.5 ± 0.7	3.7 ± 0.8
338	2.6 ± 0.6	3.1 ± 0.7	0.4 ± 0.8	1.2 ± 0.8	4.2 ± 0.6	3.6 ± 0.6
348	2.8 ± 0.7	2.9 ± 0.7	0.3 ± 1.7	1.2 ± 1.7	3.7 ± 0.8	2.8 ± 0.6
Coupling free energy ($\Delta\Delta C_p^{coupling}$) (kJ mol^{-1} K^{-1})						
	-1.0 ± 0.2	-0.9 ± 0.3	-0.4 ± 0.2	-0.1 ± 0.2	-1.0 ± 0.2	-1.0 ± 0.2

temperatures were 2.3 ± 0.4 kJ mol^{-1} and 3.7 ± 0.3 kJ mol^{-1} respectively. To test the randomness of the $\Delta\Delta G_u^{coupling}$ values, Wald-Wolfowitz runs test with 95 % confidence were performed and the standard deviations (s.d.) of $\Delta\Delta G_u^{coupling}$ values were found. The s.d. values of both salt bridges were about 0.4 kJ mol^{-1}. The result of the runs tests together with the small s.d. values suggested that the $\Delta\Delta G_u^{coupling}$ should not be a function of temperature. On the other hand, the $\Delta\Delta G_u^{coupling}$ values of the long range ion pair E90/R92 were found only to be marginally stabilizing at all temperatures, having the mean value of 0.6 ± 0.5 kJ mol^{-1}. The runs test result as well as the small s.d. values (0.3 kJ mol^{-1} for E90A/R92A cycle; 0.5 kJ mol^{-1} for E90A/R92M cycle) suggested this long range pair-wise interaction also did not vary with temperatures.

3.2.5 Salt bridge accounts for most of the increased ΔC_p value in charge-to-neutral mutants

Like ΔG_u, ΔC_p is also a state function of mutation. For example, the ΔC_p value of a doubly substituted variant, e.g. E6AR92A, will always be the same no matter we substitute E6 before substituting R92, or vice versa. Assuming the effects of charge-to-neutral mutations on ΔC_p values are additive, the effect of the salt bridge on ΔC_p value (i.e. the coupling ΔC_p, $\Delta\Delta C_p^{coupling}$) can be found by double mutant cycle of ΔC_p (Figure 3.6). The logic is the same as that of the traditional double



$$\Delta\Delta C_p^{coupling} = \Delta\Delta C_p^{WT \to DM} - \Delta\Delta C_p^{WT \to M1} - \Delta\Delta C_p^{WT \to M2}$$

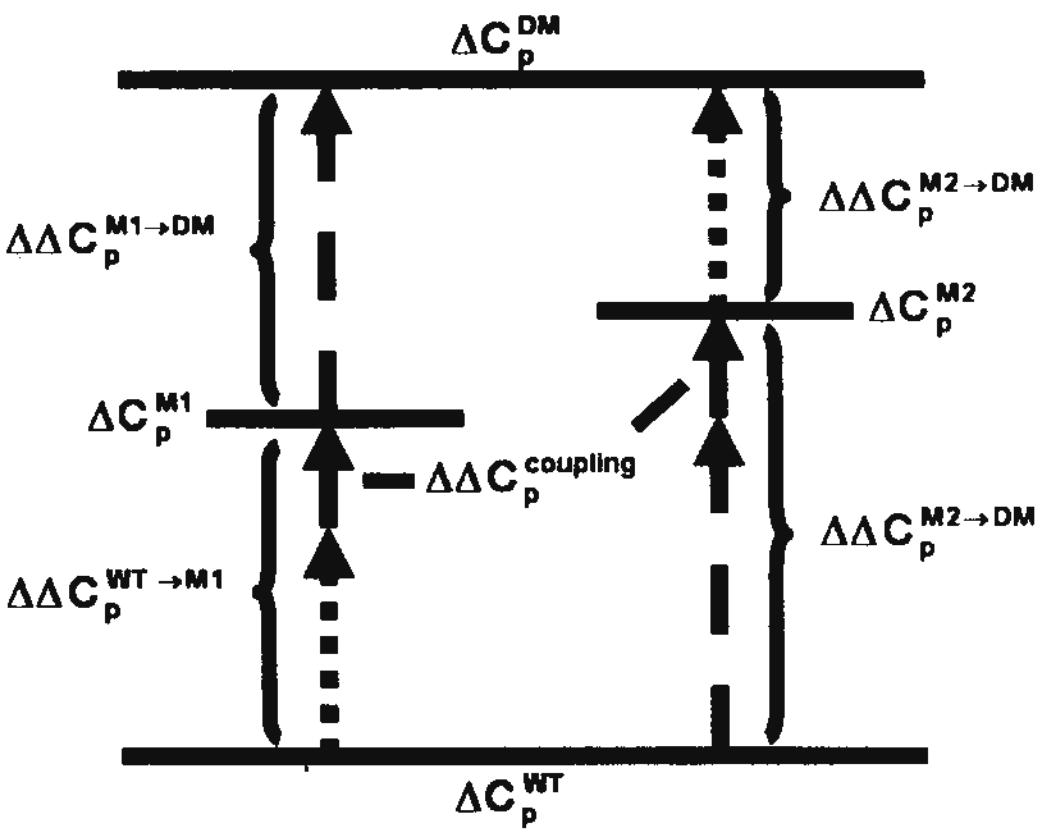


Figure 3.6. The scheme of double mutant cycle of ΔC_p for ion pair. According to the ΔC_p cycle shown above, the effect of a charge on ΔC_p due to its interaction to the rest of the protein is cancelled in the cycle, so that the coupling ΔC_p of the ion pair can be calculated, assuming the effects of charges on ΔC_p are additive.

mutant cycle of free energy. The total effect of removing a charge on ΔC_p ($\Delta\Delta C_p^{WT\rightarrow M1}$, $\Delta\Delta C_p^{WT\rightarrow M2}$) is the sum of the $\Delta\Delta C_p^{coupling}$ and the effect due to the electrostatic interactions between the charge and the rest of the protein ($\Delta\Delta C_p^{M1\rightarrow DM}$, $\Delta\Delta C_p^{M2\rightarrow DM}$). If the salt bridge does not have any effect on ΔC_p value (the salt bridged residues are decoupled with respect to ΔC_p), the total effect of removing both charges on ΔC_p ($\Delta\Delta C_p^{WT\rightarrow DM}$) mutant should be the sum of $\Delta\Delta C_p^{M1\rightarrow DM}$ and $\Delta\Delta C_p^{M2\rightarrow DM}$. Any difference between these two values indicates the salt bridge has effect on ΔC_p value, and the difference is the $\Delta\Delta C_p^{coupling}$. In other words, if $\Delta\Delta C_p^{WT\rightarrow DM}$ has a comparable values to $\Delta\Delta C_p^{WT\rightarrow M1}$ and $\Delta\Delta C_p^{WT\rightarrow M2}$, than the salt bridge interaction should dominant the effect on the change of ΔC_p. Mathematically, by substituting the ΔC_p values of wild-type protein, two single and one double mutant into the corresponding ΔG_u values in the equation of normal double mutant cycle of free energy, $\Delta\Delta C_p^{coupling}$ can be calculated.

The $\Delta\Delta C_p^{coupling}$ of the salt bridges E6/R92 and K46/E62 were found to be -0.9 ± 0.1 kJ mol^{-1} K^{-1} and -1.0 ± 0.1 kJ mol^{-1} K^{-1} respectively. In analogy to the double mutant cycle of free energy in which positive coupling free energy ($\Delta\Delta G_u^{coupling}$) indicates the pair-wise interaction increases the protein stability, the negative coupling ΔC_p ($\Delta\Delta C_p^{coupling}$) indicates the pair-wise interaction decreases the ΔC_p of the protein. Noting that the $\Delta\Delta C_p^{coupling}$ found was comparable to the $\Delta\Delta C_p^{WT\rightarrow M1}$

and $\Delta\Delta C_p^{WT \rightarrow M2}$ (Table 3.1 & 3.2), suggesting that the salt bridge interaction dominated the effect of increasing ΔC_p values when the charge is removed. On the other hand, the $\Delta\Delta C_p^{coupling}$ of the long range ion pair E90/R92 was found to be -0.2 ± 0.1 kJ mol^{-1} K^{-1}, showing this long range interaction only had a marginal effect on increasing ΔC_p value. Be noted that the $\Delta\Delta C_p^{WT \rightarrow M1}$ of the isolated charge Glu-90 also showed little effect on increasing ΔC_p value (Table 3.1), which was consistent with the finding from the salt bridges that the salt bridge interactions dominant the effect of the change in ΔC_p.

3.2.6 No structural change was observed in crystal structures of charge-to-Ala mutants

In the previous study[99], the crystal structures of two singly charge-to-Ala mutants (E90A, R92A) have been solved. In this study, the structures of another single charge-to-Ala mutants (E6A) and 3 double charge-to-Ala mutants for the three ion pairs studied (E6A/R92A, E90A/R92A, K46A/E62A) were determined by X-ray crystallography at resolution ranging from 1.8 Å to 2.4 Å. The mutant structures were solved by molecular replacement with using wild-type *T. celer L30e* (PDB code: 1H7M) as the searching model. The statistics for crystallization and structure determination are summarized in Table 3.4. All mutant proteins resemble closely the structure of wild-type *T. celer L30e* (Figure 3.7). The root-mean-square deviations

Table 3,4. Statistics for crystal structure determination of E6A, E6AR92A, E90AR92A, K46AE62A

	E6A	E6AR92A	E90AR92A	K46AE62A
Summary of crystallization conditions:				
	23% PEG 8000, 50 mM KH_2PO_4, pH 4.7, 289 K	25% PEG 20000, 0.1 M Tris, pH 8.0, 289 K	18% PEG 3350, 0.1 M sodium citrate, pH 6.0, 289 K	5% PEG MME 2000, 0.1 M MES, pH 6.0, 289 K
Diffraction data collection statistics:				
X-ray source	Cu Kα	Cu Kα	Cu Kα	Cu Kα
Space group	P32	P32	$P2_1$	$P2_1$
Resolution (Å)	24.1-2.1 (2.2-2.1)	36.5-1.8 (1.9-1.8)	31.7-1.8 (1.9-1.8)	32.4-2.0 (2.1-2.0)
Molecules per asymmetric unit	2	2	1	1
Unit cell dimension (Å)	a, 63.8; b, 63.8; c, 48.3	a, 64.2; b, 64.2; c, 48.4	a, 23.9; b, 53.1; c, 33.5	a, 23.6, b, 53.9; c, 34.3
Unit cell angles (deg.)	α, 90.0; β, 90.0; γ, 120.0	α, 90.0; β, 90.0; γ, 120.0	α, 90.0; β, 108.6; γ, 90.0	α, 90.0; β, 109.2, γ, 90.0
Multiplicity	2.8 (2.8)	2.7 (2.7)	3.0 (3.0)	2.7 (2.7)
Completeness (%)	99.9 (100.0)	100.0 (100.0)	98.1 (96.3)	99.5 (100.0)
Mean I/σ (I)	8.0 (5.4)	11.4 (4.9)	14.8 (3.7)	7.9 (4.8)
R_{merge} (%)	9.6 (17.3)	6.4 (15.3)	7.3 (24.4)	7.7 (17.4)
Unique reflections	12821 (1870)	20739 (3047)	7279 (1034)	5520 (791)
Structural refinement statistics:				
R-factor / R_{free} (%)	17.8 / 24.0	19.0 / 23.5	16.5 / 20.4	19.6 / 25.3
r.m.s.d. from idea values:				
Bond distances (Å)	0.007	0.006	0.027	0.007
Bond angles (deg.)	1.019	1.014	1.940	1.008
Ramachandran pot analysis:				
Preferred region (%)	98.4	97.3	99.0	97.9
Allowed region (%)	1.6	2.7	1.0	2.1
Outliers (%)	0.0	0.0	0.0	0.0
C^{α} r.m.s.d. with wild-type T. celer L30e	0.46	0.43	0.49	0.42
Values in parentheses are for the highest-resolution shell.				

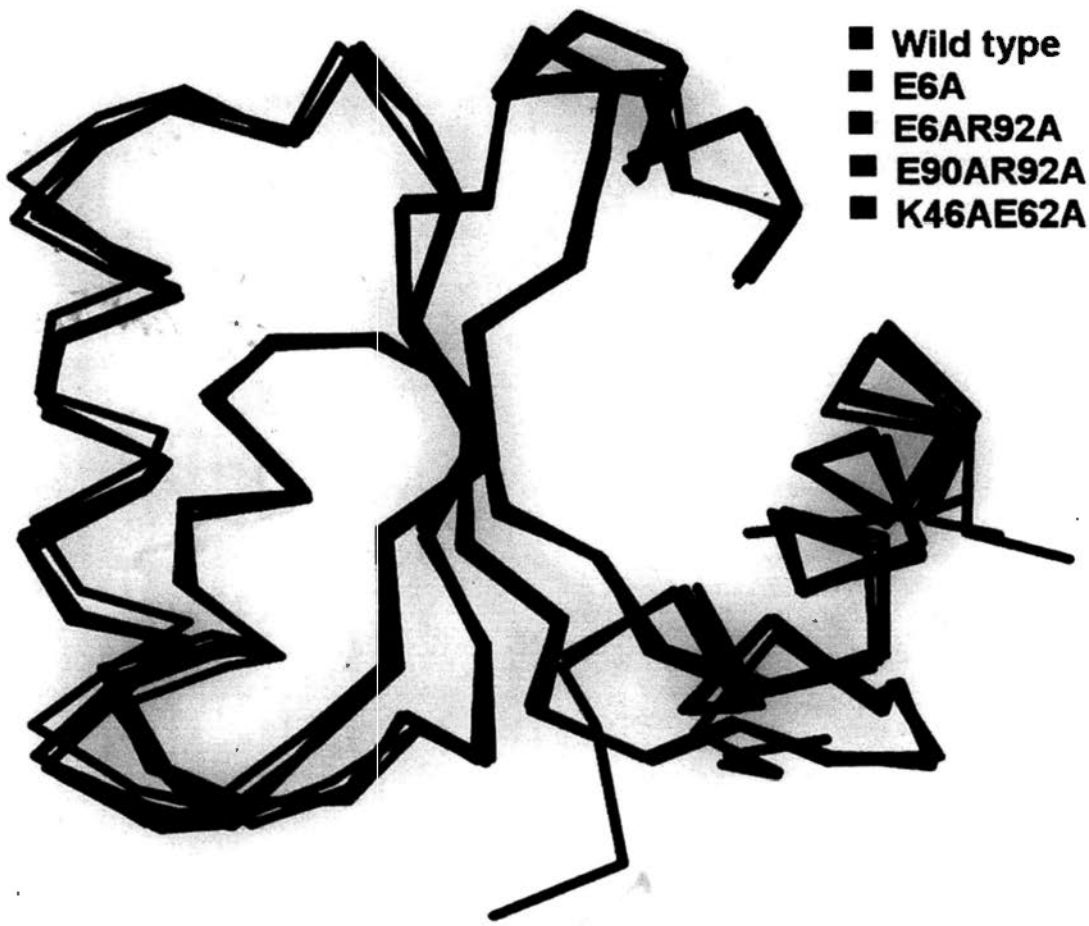


Figure 3.7. Crystal structures of some single and double mutants. Crystal structures of single mutant E6A (magenta), double mutants E6AR92A (red), E90AR92A (blue), and K46AE62A (green) are superimposable upon that of wild-type T. celer L30e (black).

(r.m.s.d.) between C^{α} atoms of mutant and wild-type *T. celer L30e* were less than 0.5 Å, suggesting there were no major structural change in these mutants.

3.3 Discussion

3.3.1 Electrostatic interaction of salt bridge contributes to thermal stability of protein

Our data here demonstrated that the coupling free energy of salt bridges were stabilizing by about 2 – 3 kJ mol^{-1} through double mutant cycles analysis. The data were validated by confirming the reversibility of urea-induced two-state unfolding, and no structural changes for the mutants used.

Electrostatic interaction should be dominant in the $\Delta\Delta G_u^{coupling}$ of the two salt bridges found in this study because only the charged group of the residues were in close proximity and the contact surface areas of salt bridges were small. The contact surface areas of E6 and R92, as well as that of K46 and E62 were 12.9 $Å^2$ and 10.4 $Å^2$ respectively, which account for less than 10% of the total ASA of the residues. A more direct evidence to show the dominance of electrostatic interaction comes from the consistence of $\Delta\Delta G_u^{coupling}$ found by double mutant cycles with using charge-to-Ala mutants and charge-to-Met mutants for the same ion pair. The $\Delta\Delta G_u^{coupling}$ found by using charge-to-Ala mutants in double mutant cycles should include both electrostatic and hydrophobic interactions between the two charged

residues. If non-charged long side-chain residues (Met and Gln in this study) were used for substitution of one or both charged residues for reserving the hydrophobic interaction, hydrophobic interaction cannot be cancelled out in double mutant cycle and lead to a smaller resulting $\Delta\Delta G_u^{coupling}$. In this study, the small difference of $\Delta\Delta G_u^{coupling}$ values between different double mutant cycles used (0.2 ± 0.4 kJ mol^{-1} for E6/R92; 0.5 ± 0.6 kJ mol^{-1} for K46/E62) indicated electrostatic interaction dominant the interaction in salt bridges.

Although the coupling free energy found by double mutant cycle cannot be directly translated to free energy of unfolding of a protein because of the importance of long range interaction on protein stability[54,75], it is an excellent way to dissect the interaction energy of a salt bridge out for a more specific study on its effect on protein stability. Double mutant cycle has been widely utilized for studying ion pairs of which the coupling free energies and the separation distances were summarized in Table 3.5. 21 out of 26 ion pairs reported were classified as salt bridges, of which the separation distances were less than or equal to 4.0 Å. The $\Delta\Delta G_u^{coupling}$ values of salt bridges were found to be highly variable ranging from 0.4 kJ mol^{-1} to 14.3 kJ mol^{-1}, with the lower quartile, median and upper quartile of 2.1 kJ mol^{-1}, 2.7 kJ mol^{-1} and 3.6 kJ mol^{-1} respectively. The $\Delta\Delta G_u^{coupling}$ values of two salt bridges found here fall in the inter-quartile range, indicating they are good representatives of salt bridges.

Table 3.5. The contribution of ion pairs to protein stability found by double mutant cycles

Ion pair	$\Delta\Delta G_u^{coupling}$ (kJ / mol)	Separation distance[a] (Å)
Barnase[72,130-132]		
D8 / R110	3.5	2.8
D12 / R110	1.4	3.1
R69 / D93	14.3	2.8
D12 / R16	0.9	5.1*
E28 / K32	0.3	5.2*
GCN4-p1[133]		
K8 / E11	1.2	2.9
E11 / K15	2.1	2.9
λ repressor[134]		
D14 / R17	3.4	2.7
NTL9[74]		
Nt / D23	7.1	2.8
Protein G[135,136]		
R6 / E53 (T44A)	2.5	3.3*
R6 / E53 (T44R)	2.7	3.3*
E6 / R53	3.1	2.5*
K6 / E53	1.6	3.9*
R44 / E53	2.2	2.8*
R44 / E53 (I6R)	2.4	2.8*
Rubredoxin[43]		
K6 / E49	1.2	4.2
Nt / E14	6.3	2.8
Ssh10b[137]		
E36 / K66	6.0	3.0
E54 / R57	2.4	2.6
T4 lysozyme[138]		
D116 / R119	0.4	2.8
T. celer L30e		
E6 / R92	1.9	3.6
E90 / R92	0.4	10.8
K46 / E62	3.3	3.4
Ubiquitin[75]		
K11 / E34	3.6	3.4
E11 / K34	3.8	3.4*
Zinc finger[139]		
K3 / E14	0.8	7.0

[a] PDB entries used for finding separation distances: 1A2P, 1BRF, 1H0X, 1H7M, 1LMB, 1MEY, 1PGA, 1UBQ, 2HBB, 2LZM, 3K7Z. Separation distances estimated by virtual mutagenesis are indicated by asterisks.

The contribution of salt bridge to thermal stability has been being challenged by large desolvation penalty[37,38,57] and the entropic cost of fixing two charged side-chains[43,44,46]. A theoretical calculation based on a simple electrostatic model has predicted the desolvation cost outweighs the favourable electrostatic interactions of a salt bridge and the salt bridge alone would destabilize the native protein by 3.2 kJ mol^{-1} [129]. Despite the presence of desolvation penalty and entropic cost of salt bridge formation, many experimental evidences supporting the stabilizing role of salt bridge in thermal stability of protein have been reported[48,50-52]. Our findings also support the conclusion that the electrostatic interaction of salt bridge stabilizes thermal stability of protein.

3.3.2 Electrostatic interaction contributes to thermal stability by up-shifting of thermophilic *T. celer L30e* protein stability curve

In this study, our data also showed the protein stability curve of thermophilic *T. celer L30e* is largely up-shifted when compared with that of its mesophilic yeast homolog (Figure 3.4A). Also, we have shown the $\Delta\Delta G_u^{coupling}$ of the two salt bridges were always stabilizing by about 2 - 3 kJ mol^{-1} at temperature ranging from 25 °C to 75 °C, without showing any trend with respect to temperature.

Large up-shift of thermophilic protein stability curve is one of the common

observations when compared to that of their mesophilic homologues. Mathematically speaking, the effect of a positive constant (a horizontal line above x-axis) on a curve is to up-shift the whole curve by that positive constant (Figure 3.8). Since the $\Delta\Delta G_u^{coupling}$ of salt bridges were found to be temperature independent, and always stabilizing by about 2 – 3 kJ mol^{-1} per one salt bridge according to double mutant cycles, the effect of salt bridge interactions on protein stability was therefore to improve protein stability at every temperature by up-shifting the whole protein stability curves by the total average sum of coupling free energy of those salt bridges (5.9 kJ mol^{-1}). Although $\Delta\Delta G_u^{coupling}$ of each salt bridge only has a moderate stabilizing effect, summing up the $\Delta\Delta G_u^{coupling}$ of the two salt bridges on *T. celer L30e* can already account for 35% of its total up-shift of protein stability curve (5.9 kJ mol^{-1} out of 16.8 kJ mol^{-1}).

Another evident to show salt bridge is important for the up-shift of the protein stability curve comes from the study on ΔC_p values. Our previous research showed that electrostatic interaction contributes to the reduced ΔC_p[99], which can explain why only those salt bridge disrupting mutants have a significant increase in ΔC_p values (in average 1.1 ± 0.1 kJ mol^{-1} K^{-1}). Our novel analysis of ΔC_p showed the interaction of salt bridge can account for almost all the reduced ΔC_p in the mutants. In average,

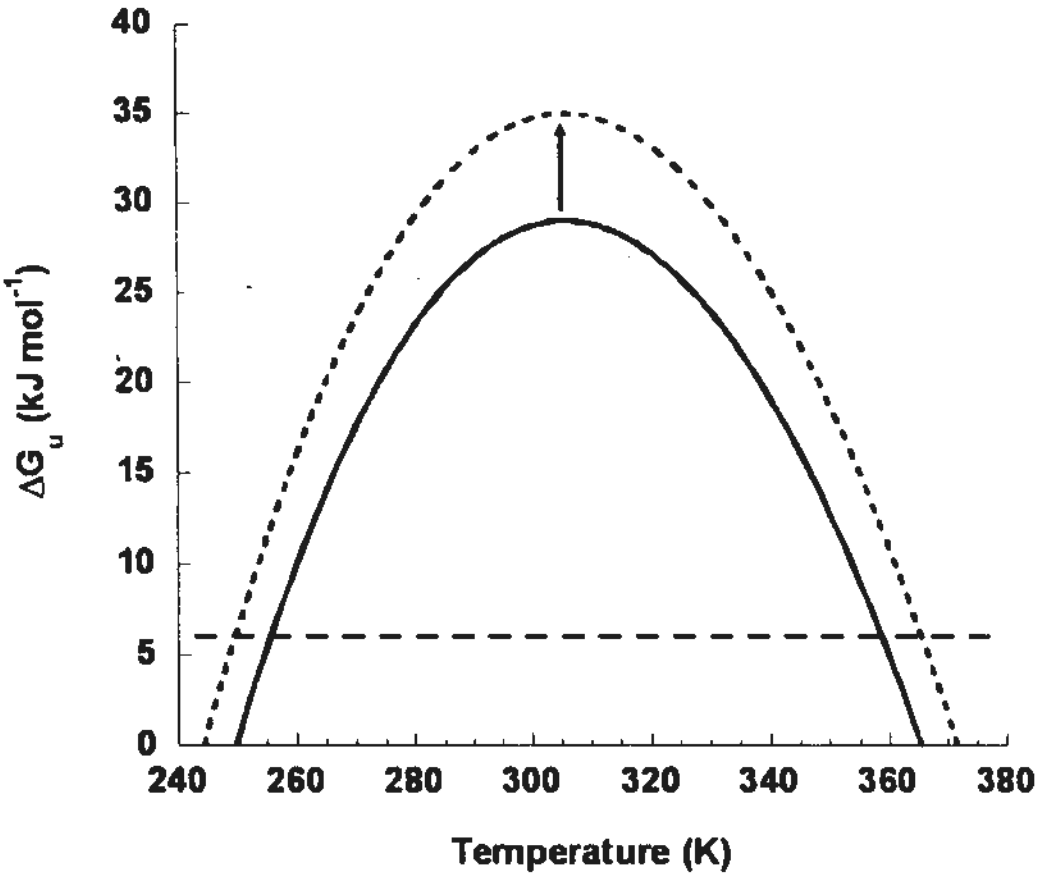


Figure 3.8. The up-shifting effect of salt bridge coupling free energy. Protein stability curves were simulated using the Gibbs-Helmholtz equation, with setting the ΔC_p to 5.3 kJ mol^{-1} K^{-1}. The effect of temperature independent salt bridge coupling free energy (dashed horizontal line) on hypothetical protein without stabilized by the salt bridge (continuous line) is to up shift the whole protein stability curve by the same amount of the coupling free energy (dotted line, up-shifted by 6.0 kJ mol^{-1}). The up-shift of the protein stability curve is indicated by the arrow.

interaction of each salt bridge accounted for 1.0 ± 0.1 kJ mol^{-1} K^{-1} reduction of ΔC_p. On the contrary, individual long range ion pair (E90/R92) seemed to have no effect on ΔC_p value. Structural analysis of T. celer and yeast L30e revealed the former has 6 more long range ion pairs (ion pair within 8 Å), but only two more salt bridges (ion pair within 4 Å). Although the number of salt bridge does not differ much among the homolog proteins, 2 more salt bridges can already account for 77% of the total reduction of ΔC_p in thermophilic homolog (2.0 ± 0.1 kJ mol^{-1} K^{-1} out of 2.6 ± 0.6 kJ mol^{-1} K^{-1}). Our data has clearly demonstrated that the majority of the reduction of ΔC_p values in thermophilic proteins comes from a few salt bridges. In theory the protein stability curve is not necessarily up-shifted even salt bridges have reduced the ΔC_p value, provided that ΔH_m value has increased largely enough (Figure 3.9). However, single or double mutations which broke a salt bridge seem not to have an observable effect on ΔH_m (Table 3.1). Thus, the reduction of ΔC_p by salt bridge interaction would therefore lead to the up-shift of protein stability curve.

To sum up, our data on the effect of salt bridge has clearly demonstrated the effect of salt bridge interactions could lead to the large up-shift of the protein stability curve, which is in fact the common observation that thermophilic proteins have largely up-shifted protein stability curves when compared to their mesophilic homologues. Be reminded that thermophilic proteins have in general elevated surface

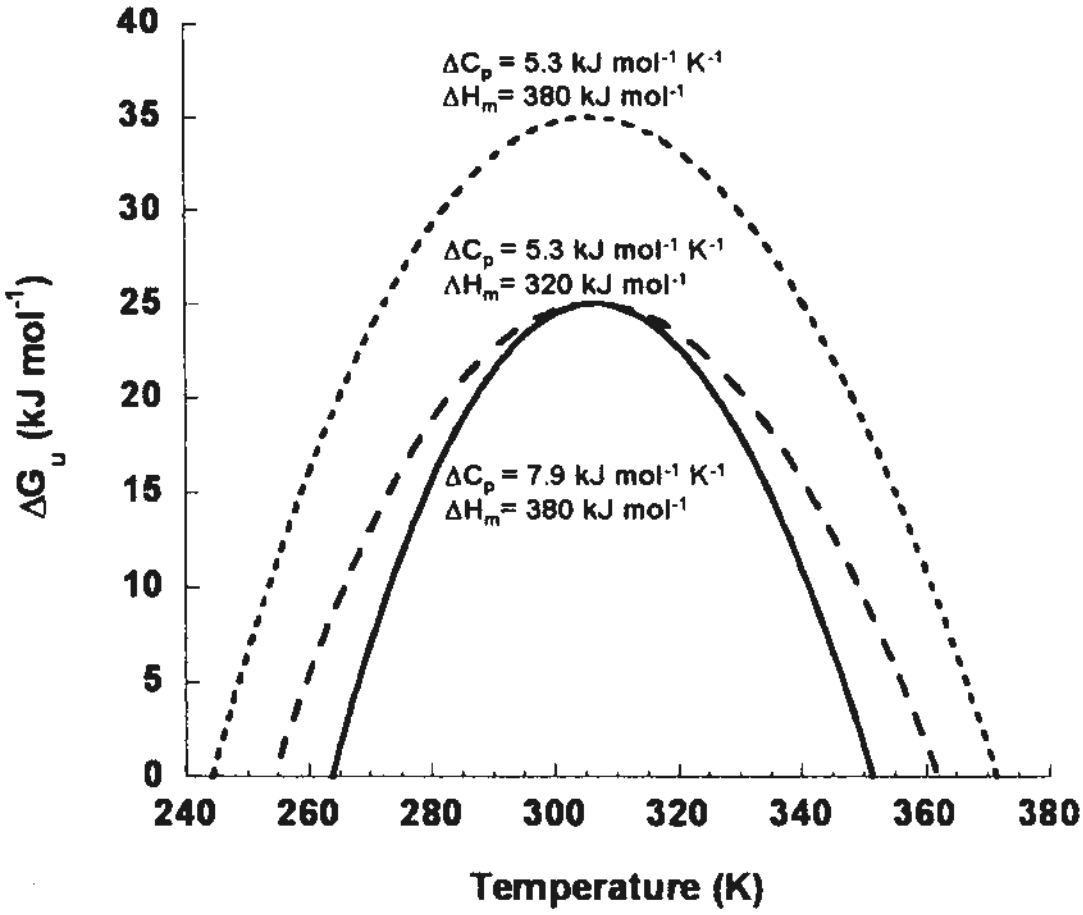


Figure 3.9. The effect of reduced DCp value on the protein stability curve. Protein stability curves were simulated using the Gibbs-Helmholtz equation. ΔC_p and ΔH_m of hypothetical mesophilic protein (continuous line) were set to 7.9 kJ mol^{-1} K^{-1} and 380 kJ mol^{-1}. Reducing ΔC_p to 5.3 kJ mol^{-1} K^{-1} with keeping ΔHm unchanged gave an largely up-shifted curve (dotted line), while a broadened curve without up-shifting (dashed line) would be resulted if ΔH_m were set free for fitting and decreased to 320 kJ mol^{-1}.

charges when compared to their mesophilic homologues. The finding in this study evidently supports the hypothesis that the increased electrostatic interaction of thermophilic proteins contributes to their unusually high thermostability. Our analysis on $\Delta\Delta G_u^{coupling}$ and $\Delta\Delta C_p^{coupling}$ also clearly demonstrated the common strategy to achieve the unusually high thermostability of thermophilic proteins by largely up-shift the protein stability curve which is mainly due to the salt bridge electrostatic interactions.

3.3.3 Electrostatic interaction of salt bridge is temperature independent

Conflicting reports on the role of electrostatic interaction on protein stability are found in the literatures. Electrostatic interaction has been reported to be stabilizing in many cases[48,50-53], but it has also been reported to have little contribution to protein stability[37,38,43] or even destabilize proteins[44-46] because of the large desolvation penalty and entropic cost. One explanation for these contradicting findings comes from theoretical models[140]. Elcock A.H. has argued that salt bridge should be more stabilizing at high temperatures solely due to the decrease in dehydration penalty of salt bridge formation. In his theoretical calculation, the interaction energy of a salt bridge (3.4 Å) is 3.3 kJ mol^{-1} more stabilizing by raising the temperature by 75 °C from 25 °C to 100 °C. However, our study has experimentally shown that no

observable increasing trend for the $\Delta\Delta G_u^{coupling}$ of salt bridge as temperature increased. The temperature independency of $\Delta\Delta G_u^{coupling}$ was also observed in salt bridges in another thermophilic protein Ssh10b[137]. The failure of Elcock's theoretical model in predicting the trend of $\Delta\Delta G_u^{coupling}$ with increasing temperatures reveals that some energetic considerations might be leaved out in the theoretical model. One possible unconsidered factor is the entropic cost of fixing interacting side chains, which did not take part in the calculation of salt bridge interaction energy. Nevertheless, our experimental data has clearly shown that the $\Delta\Delta G_u^{coupling}$ of salt bridge is temperature independent and is resilient to temperature increases. Although the absolute $\Delta\Delta G_u^{coupling}$ value of salt bridge does not increase with temperature, the contribution of salt bridge to protein stability still becomes more important at high temperatures. Both mesophilic and thermophilic proteins usually have their maximal protein stability at temperature (T_s) around room temperature[100]. Further increase in temperature results in the decrease of protein stability. The contribution of salt bridge however remains constant at high temperatures. Therefore, proportionally salt bridges contribute more at high temperature. For example, the two salt bridges E6/R92 and K46/E62 contribute by about 17% at 25 °C, and the percentage contribution increased to 28% at 75 °C.

3.3.4 Concluding remarks

Our data has demonstrated the electrostatic interaction between salt bridges stabilize the protein by about 2-3 kJ mol^{-1} per salt bridge at temperatures ranging from 25 °C to 75 °C, showing increasing temperature will not greatly increase the stabilizing effect of salt bridges as predicted in theoretical calculation[140]. Instead, the temperature independency of salt bridge interactions showed that electrostatic interaction between ion pairs will lead to the up-shift of the protein stability curves. According to our novel double mutant cycle of ΔC_p analysis, we have also shown that salt bridge interactions dominant the effect of reduced ΔC_p, which will also lead to the large up-shift of protein stability curves. Taking together, improving electrostatic interaction of salt bridges can improve the protein stability at all temperatures by largely up-shift the protein stability curve, which is the common strategy for thermophilic proteins, which have improved surface charge-charge interactions, to achieve their unusually high thermostability.

Chapter 4: Guanidine HCl-induced denatured *T. celer L30e* has no observable residual structures

4.1 Introduction

Characteristics of native (folded) state of proteins have been extensively investigated in thermodynamic studies[5,25,26,141,142]. Less attention has been paid to the effect of denatured state on protein stability, although improving global protein stability by manipulating denatured state has also been investigated[84-86,143-145]. In principle, denatured state should be equally important for determining the free energy of unfolding (ΔG_u) of the proteins as ΔG_u is the difference between the free energy of denatured and native state. Unlike the free energy of native state, that of denatured state refers to the average free energy level of an ensemble of inter-converting conformations, which are termed as "random coil conformations". The population of these similar free energy level random coil conformations should be normally distributed according to Boltzmann distribution. In general, residual structures referred to any preferred conformations that have significantly increased population in the denatured state, usually due to electrostatic interactions, hydrogen bonds, or other interactions present in the denatured state. Although it has been reported that the interaction in denatured state only has little effect on protein stability of small

CspB protein[146], many experimental studies have demonstrated the important role of residual structures in protein stability[89,91]. Recently, it has also been suggested the high thermostability of thermophilic proteins (largely reduced ΔC_p) is originated from the residual structures in the denatured state[147]. Therefore, to investigate if any residual structures contribute to the high thermostability and the largely reduced ΔC_p of *T. celer L30e*, residual structures were tried to be detected by the rigidity of backbone (backbone amide intensities along the sequence), structural effect on chemical shifts (secondary shifts of backbone and nearby atoms), and the combination of the two backbone torsion angles ϕ_i and ψ_i (normalized $\sigma_{NN(i-1,i)}$, and $\sigma_{N\alpha(i,i)}$: $\sigma_{\alpha N(i-1,i)}$ ratio). Our data showed there were no observable residual structures in the completely denatured protein, and therefore residual structures should not be the major contributor for the thermophilic properties of *T. celer L30e*.

4.2 Result

4.2.1 Wild-type *T. celer L30e* denatured in the presence of 6.0 M Guanidine hydrochloride

$^1H^{15}N$-HSQC[148] spectra of *T. celer L30e* at pH 4.0 was obtained with or without the presence of 6.0 M guanidine HCl at 298 K (Figure 4.1). The $^1H^{15}N$-HSQC spectrum with no guanidine HCl clearly has a much wider dispersion in ^1H dimension, with 1H chemical shifts range from 6.5 ppm to 10.0 ppm. This high

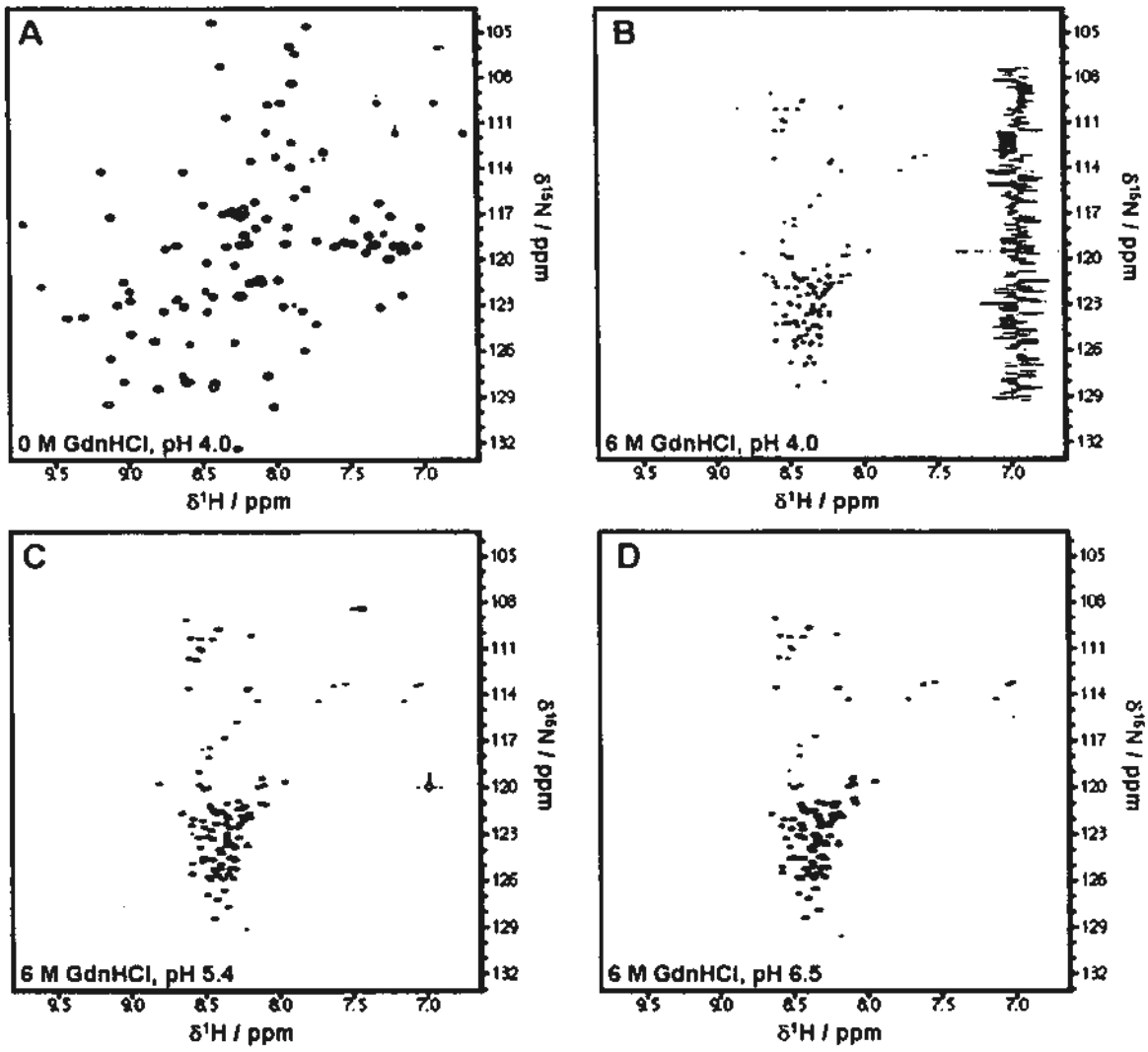


Figure 4.1. 2D 15N-HSQC spectra of native and denatured *T. celer L30e*. 2D ^{15}N-HSQC spectra of wild-type *T. celer L30e* at 298 K in 10 mM citrate/phosphate buffer, at (A) pH 4.0, (B) pH 4.0 with 6.0 M guanidine HCl, (C) pH 5.4 with 6.0 M guanidine HCl, and (D) pH 6.5 with 6.0 M guanidine HCl. Be noted that samples with 6.0 M guanidine-HCl (B, C, D) had much narrower dispersion in ^{1}H dimension.

dispersion of proton chemical shift clearly showed that *T. celer L30e* in the low ionic strength NMR condition at pH 4.0 at 298 K remained native. On the other hand, the proton chemical shift dispersion in the presence of 6.0 M guanidine hydrochloride was much narrower, ranging from 7.5 ppm to 8.7 ppm. The characteristic narrow proton chemical shift dispersion that is close to the random coil values evidently suggested *T. celer L30e* was completely denatured in the presence of 6.0 M guanidine HCl.

To investigate the effect of pH on the denatured state, $^{1}H^{15}N$-HSQC spectra of *T. celer L30e* at pH ranging from 4.0 to 6.5 in the presence of 6.0 M guanidine HCl were also acquired (Figure 4.1). When compared to the spectrum at pH 4.0, spectra at pH 5.4 and pH 6.5 also had the same narrow proton chemical shift dispersion, suggesting *T. celer L30e* also completely denatured at high pH. In addition, the coordinates of the peaks at different pH were very similar, suggesting no major difference in denatured state at different pH.

4.2.2 Resonance assignment of denatured *T. celer L30e*

Backbone assignments of ^{13}C, ^{15}N doubly labeled denatured *T. celer L30e* at pH 5.4 at 298 K were obtained by using triple-resonance experiments. HNCACB[149] was the most important for doing the backbone assignment. According to the phase and magnitude of the resonance peaks, HNCACB alone could provide us the $^{13}C^{\alpha}$ and

$^{13}C^{\beta}$ chemical shift values for i and i – 1 residues[150]. Therefore, the reside types for i and i – 1 residues were determined, and the residues could be sequentially assigned. However, the narrow dispersion in protein chemical shifts, as well as the very close $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts (close to random coil values) for the same type of residue, induced ambiguity in the assignment due to the overlapping of resonance peaks. To this ends, a pair of complementary triple-resonance experiments, namely ^{15}N-TOCSY-HSQC[151] and HC(C)(CO)NH-TOCSY[152], where the aliphatic side chain ^{1}H chemical shifts of i and i – 1 were correlated to that of backbone amide respectively, were done to provide independent linkage for sequential assignment. The same problem in HNCACB was also faced by this pair of experiments. This pair of experiments did help determining the residue type of i and i – 1 residues, but since the aliphatic ^{1}H chemical shifts of all residues were close to the random coil shifts, not much extra sequential linkage could be provided. In contrast, another triple resonance experiment ^{1}H^{15}NHSQC-NOESY-HSQC[153] provided us with unambiguously rich information on sequential assignment, although no information for residue type was provided. In this experiment, ^{1}H^{15}N-HSQC-NOESY-HSQC amide NOEs were correlated with amide chemical shifts and the amide resonances were dispersed into two nitrogen dimensions. Using mixing time of 120 ms, nitrogen resonance peaks of the i – 1, i, and i + 1 residues were clearly observed for each

backbone amide. Since nitrogen has very large dispersion (~ 30 ppm), the three nitrogen resonance peaks for each residue were in the most cases well separated. Therefore, this $^{1}H^{15}N$-HSQC-NOESY-HSQC could be used for the unambiguous conformation of the sequential linkage assigned according to HNCACB and the pair of TOCSY experiments (Figure 4.2). With the experiments mentioned above, the sequence and residue type of all the residues in the denatured *T. celer L30e* were successfully assigned (Figure 4.3).

4.2.3 Small variation of amide signal intensities along the sequence of wild-type *T. celer L30e*

Backbone amide resonance single intensities of all residues were calculated from $^{1}H^{15}N$-HSQC, and the intensities were plotted against the amino acid position (Figure 4.4). It was found that the intensities near both N-terminal and C-terminal were a little bit higher than those in the middle of the protein sequence. The farther away from the terminal, the smaller the intensity the residue has. Therefore, a rough "V-shape" curve with the minimum at around position 50 (the middle of the protein sequence), was obtained. However, although intensities near terminals were higher, the intensity difference along the sequence was not significant. According to Huber's methods, no residue could be regarded as outliners according to their amide single intensity.

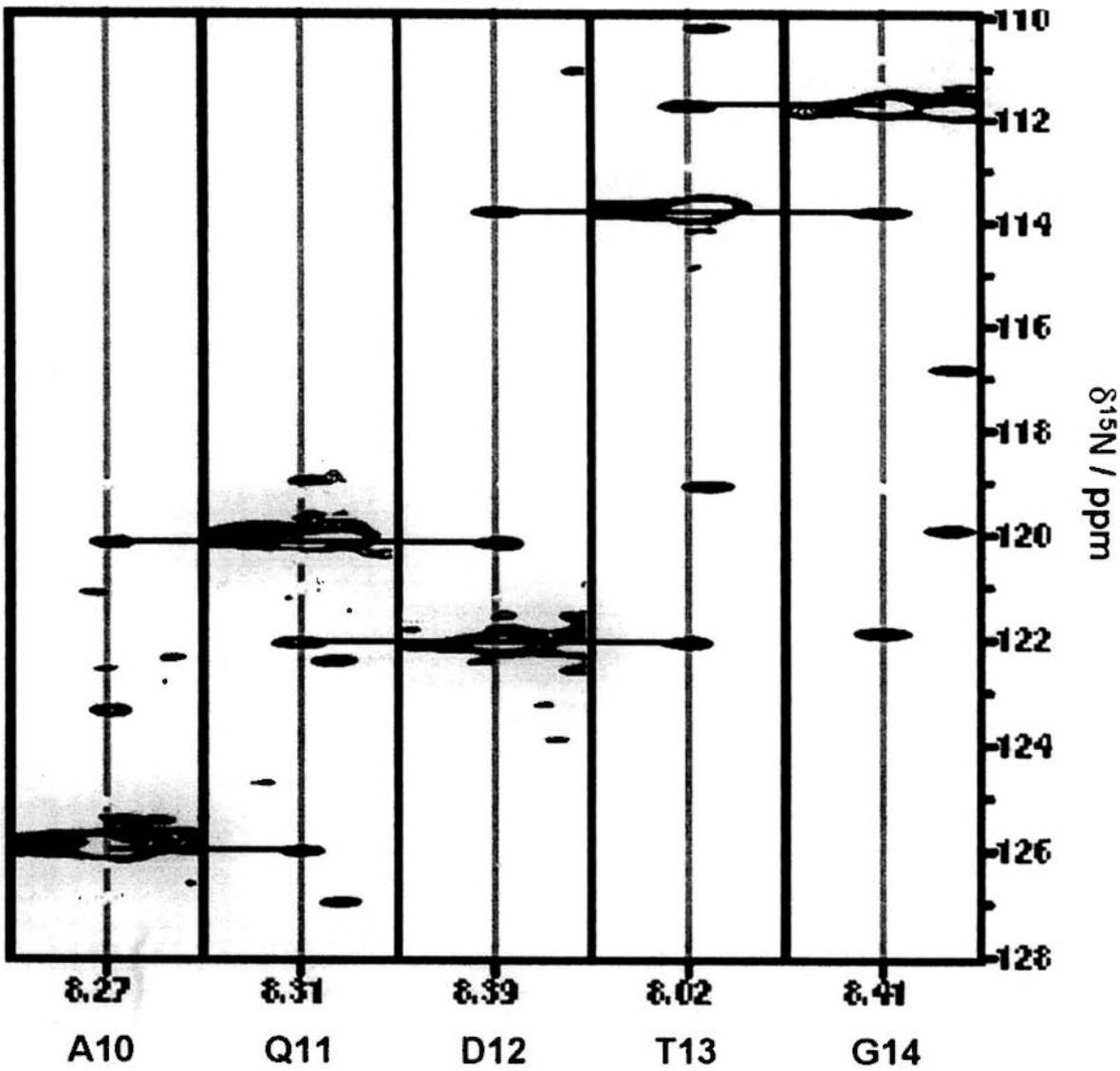


Figure 4.2. Sequential connectivities of 3D ^{15}N-HSQC-NOESY-HSQC spectrum. Selected pairs of ^{15}N^{15}N strips from the 3D ^{15}N-HSQC-NOESY-HSQC spectrum of unifomly ^{15}N^{13}C-labelled wild-type T celer L30e in 6.0 M guanidine HCl, 10 mM citrate/phosphate buffer at pH 5.4 at 298 K, which illustrate main-chain sequential connectivities for residues Ala-10 to Gly-14. Sequential connectivities are indicated by horizontal lines between strips.

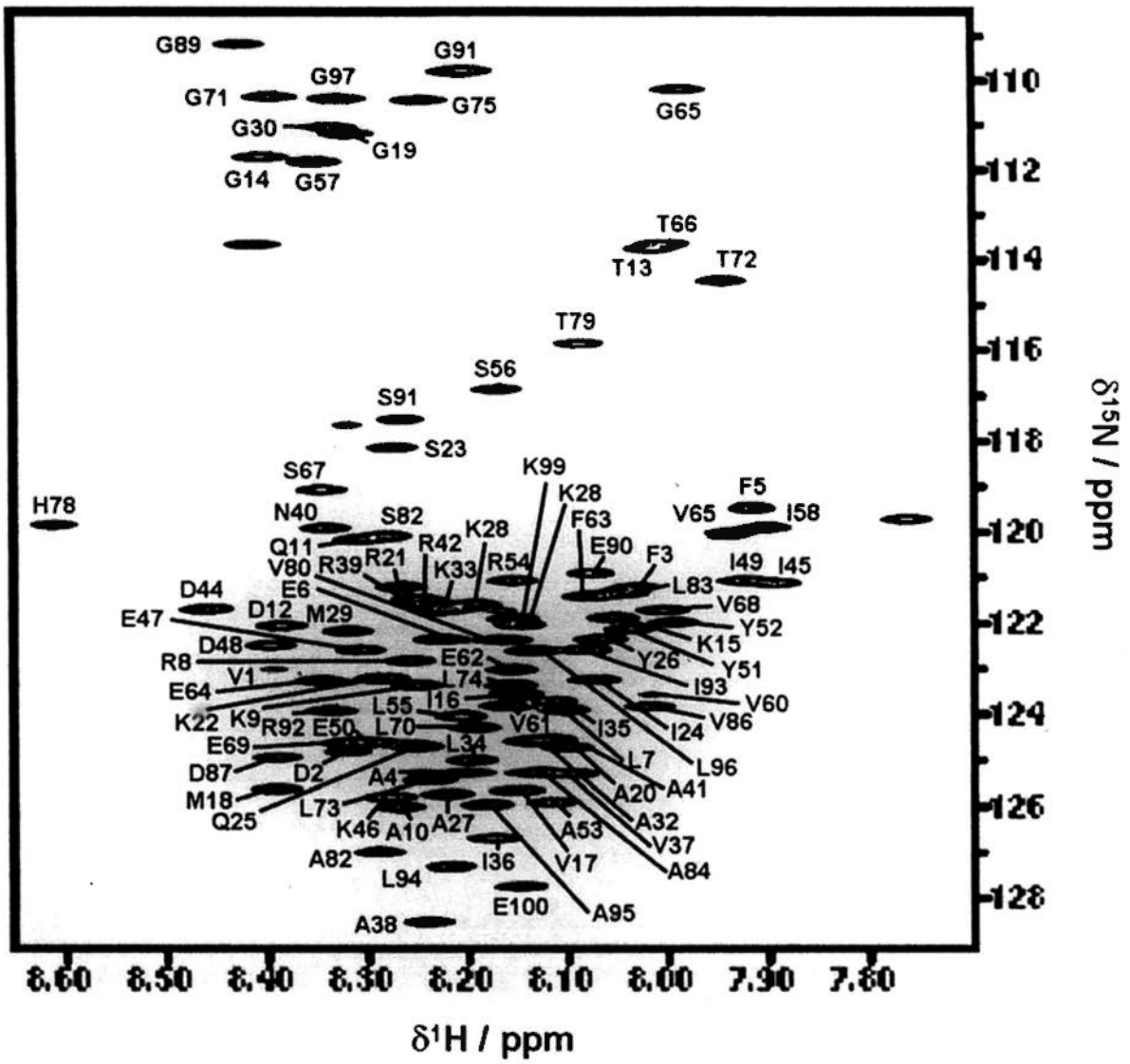


Figure 4.3. Backbone assignment of denatured *T. celer L30e*. $^{1}H^{15}N$-HSQC spectrum of the *T. celer L30e* at 298 K at pH 5.4. Assignments of 96 backbone amide signals are denoted by the one-letter code and the residue number.

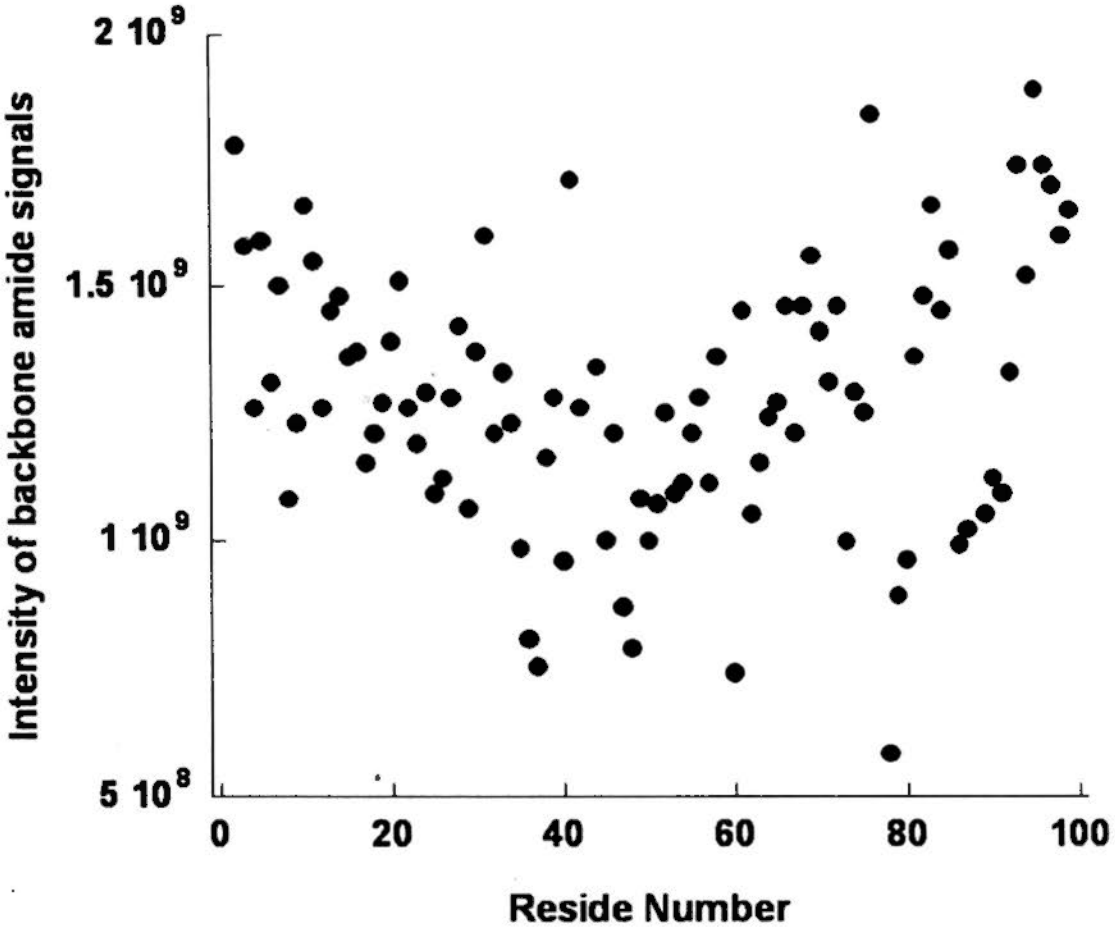


Figure 4.4. Backbone amide intensities from 1H15N-HSQC spectrum. Backbone amide intensities in arbitrary units from a $^{1}H^{15}N$-HSQC spectrum of the wild-type *T. celer L30e* in 6.0 M guanidine HCl at 298 K at pH 5.4 were plotted against sequence position. The peak intensities were calculated by fitting Lorentzian peak shape to amide signals and integrating the peak volume.

4.2.4 Chemical shifts of denatured wild-type *T. celer L30e* deviate little from random-coil values

Chemical shifts of amino acids, especially for those of the backbone and the nearby atoms, can reflect the secondary structures in the protein[154-158]. In our study, the chemical shift of ^{1}H, ^{15}N, CO, C^{α}, H^{α}, C^{β} were collected from various NMR experiments ($^{1}H^{15}N$-HSQC, NHCO, ^{15}N-TOCSY-HSQC, HC(C)(CO)NH-TOCSY , and HNCACB). These experimental chemical shifts values were compared with the sequence-corrected random coil shifts[159-161], and the deviation from sequence-corrected random coil shifts of ^{1}H, ^{15}N, CO, C^{α}, H^{α}, C^{β} for all residues were summarized in Figure 4.5. Huber's method was applied for identifying outliner chemical shifts. It was found that almost all of these experimental chemical shifts were within the random coil range defined by Huber's method. The exceptions included Lys-28 and Arg-54 in ^{1}H chemical shifts, Glu-100 in ^{15}N chemical shifts, Arg-42, Ile-58, Arg-76, and Glu-100 in C^{α} chemical shifts, Pro-43, and Glu-100 in H^{α} chemical shifts, as well as Met-0 and His-78 in C^{β} chemical shifts.

4.2.5 Using NOE as a probe for residual structures in denatured wild-type *T. celer L30e*

The inter-atomic distance between the backbone ^{1}H of i and i – 1 residues ($d_{NN(i-1,i)}$) is all dependent on the combination of the two torsion angles about a single

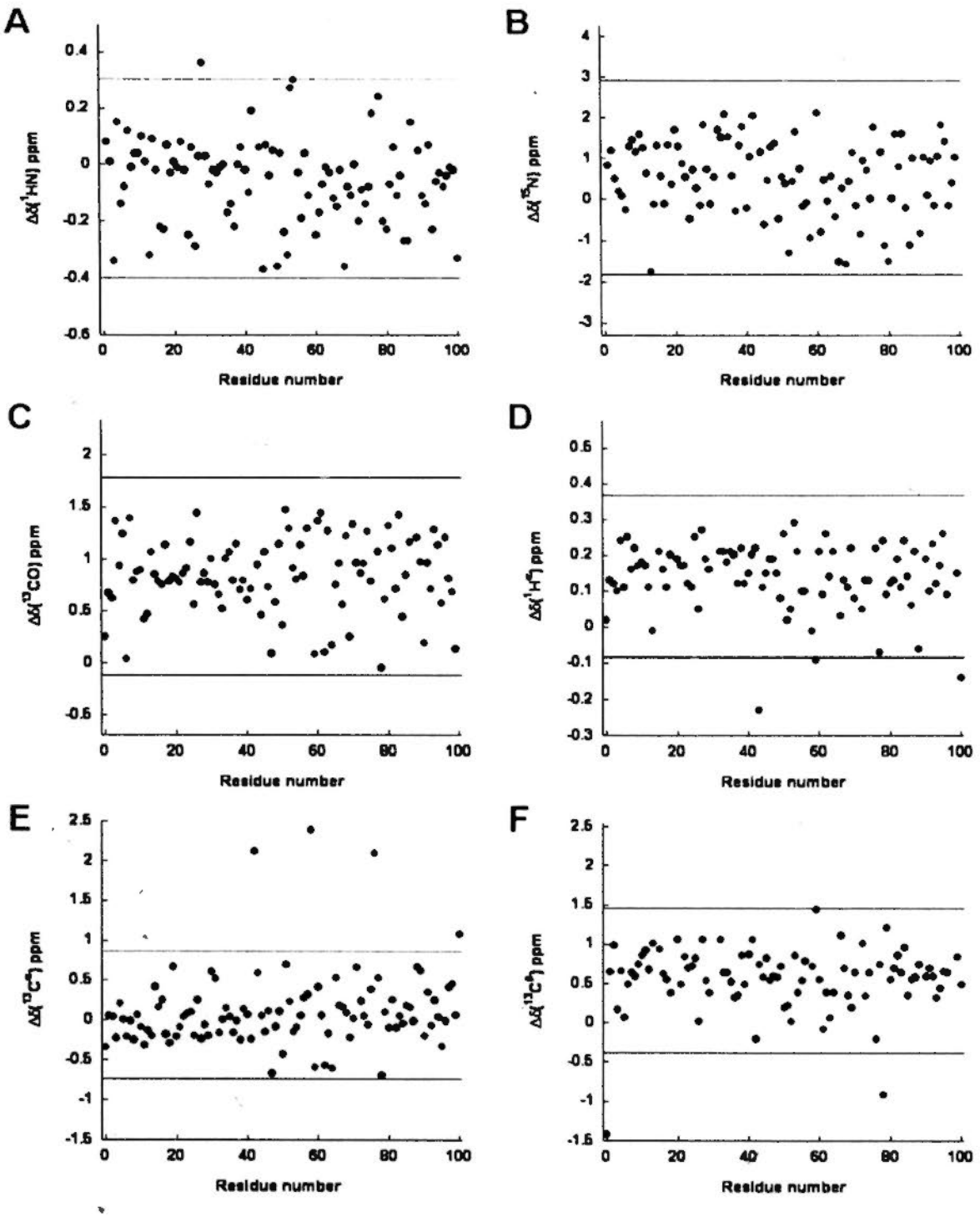


Figure 4.5. Chemcial shift deviations from random coil shifts. Chemical shift deviations (Δδ) for resonances of backbone atoms (A) ^{1}HN, (B) ^{15}N, (C) ^{13}CO, (D) $^{1}H^{\alpha}$, and (E) $^{13}C^{\alpha}$, as well as nearby atoms (F) $^{13}C^{\beta}$ of wild-type *T. celer L30e* with 6.0 M guanidine HCl at 298 K at pH 5.4. Sequence-corrected random coil shifts were used for the deviation calculation. Outliners of deviations were found by Huber's method.

bond in backbone ϕ_i and ψ_i. For L-amino acid, ϕ_i and ψ_i will be highly restricted if the residue is located in secondary structures like regular α-helix or β-strand[162,163]. Consequently, the average d_{NN} of a residue located in regular α-helix (2.8 Å) is shorter than that in β-strand (4.3 Å)[164]. Without residual structure in the denatured state, proteins are expected to have an ensemble of inter-converting random coil structures, in which polyproline II has been reported as an important conformation[165-169]. Random coil structures, including polyproline II, are predominantly populated by conformers in the β-region of ϕ_i and ψ_i conformational space[170]. Therefore, $d_{NN(i-1,i)}$ is a good parameter to indicate if a residue is involved in any residual structure in denatured protein. Thus, distance sensitive NOE should be a good choice for probing any residual structures (α-helix structures) in the denatured state protein. In another words, a residue should be involved in residual structures if the signal intensity of $d_{NN(i-1,i)}$ NOE ($\sigma_{NN(i-1,i)}$) of the residue is "extraordinarily large" when compared to that of the rest of the protein (it is unlikely that most of the residues in the denatured state are involved in residual structures). For a fair comparison, $\sigma_{NN(i-1,i)}$ of each residue was normalized with the corresponding diagonal peaks in $^1H^{15}N$-HSQC-NOESY-HSQC spectrum by calculating the ratio of the their intensities, which were obtained by calculating the Lorentzian-shape fitted volumes of the resonance peaks. In our study, the normalized $\sigma_{NN(i-1,i)}$ values for all

residues in *T. celer L30e* at pH 5.4 at 298 K were calculated (Figure 4.6) It was found that the $\sigma_{NN(i-1,i)}$ values in general were a little bit higher than those near N- and C-terminals. However, the differences were not significant as no outliners were detected by Huber's method. Also, the absolute value of the ratio is very small (less than 0.05).

In addition to $d_{NN(i-1,i)}$, the separation distance between H" and ^{1}H in the same residue ($d_{N_\alpha(i,i)}$), as well as the separation distance between "H in residue i – 1 and ^{1}H in i residue ($d_{\alpha N(i-1,i)}$), are also dependent on the backbone ϕ_i and ψ_i torsion angles. For $d_{N_\alpha(i,i)}$, the value is in average a bit higher in β-strand structure (2.8 Å) than in α-helix structure (2.6 Å). Whereas for $d_{\alpha N(i-1,i)}$, the value is much smaller in β-strand structure (2.2 Å) than in α-helix structure (3.5 Å)[164]. Therefore, for the same argument as the case of $\sigma_{NN(i-1,i)}$, the $\sigma_{N_\alpha(i,i)}$ to $\sigma_{\alpha N(i-1,i)}$ of a residue would be "extraordinarily large" if it is involved in residual structures. Similarly, for fair comparison, $\sigma_{N_\alpha(i,i)}$ and $\sigma_{\alpha N(i-1,i)}$ were first normalized with the corresponding diagonal peaks in ^{15}N-TOCSY-HSQC and HC(C)(CO)NH-TOCSY respectively before calculating the ratio (Figure 4.7). In our study, more than half (60 out of 100) of the $\sigma_{N_\alpha(i,i)}$: $\sigma_{\alpha N(i-1,i)}$ ratio were discarded because of serious overlapping problem for the diagonal or cross peaks in the two TOCSY-type experiments. Nevertheless, there is no obvious trends that any parts of the protein sequence have "extraordinarily

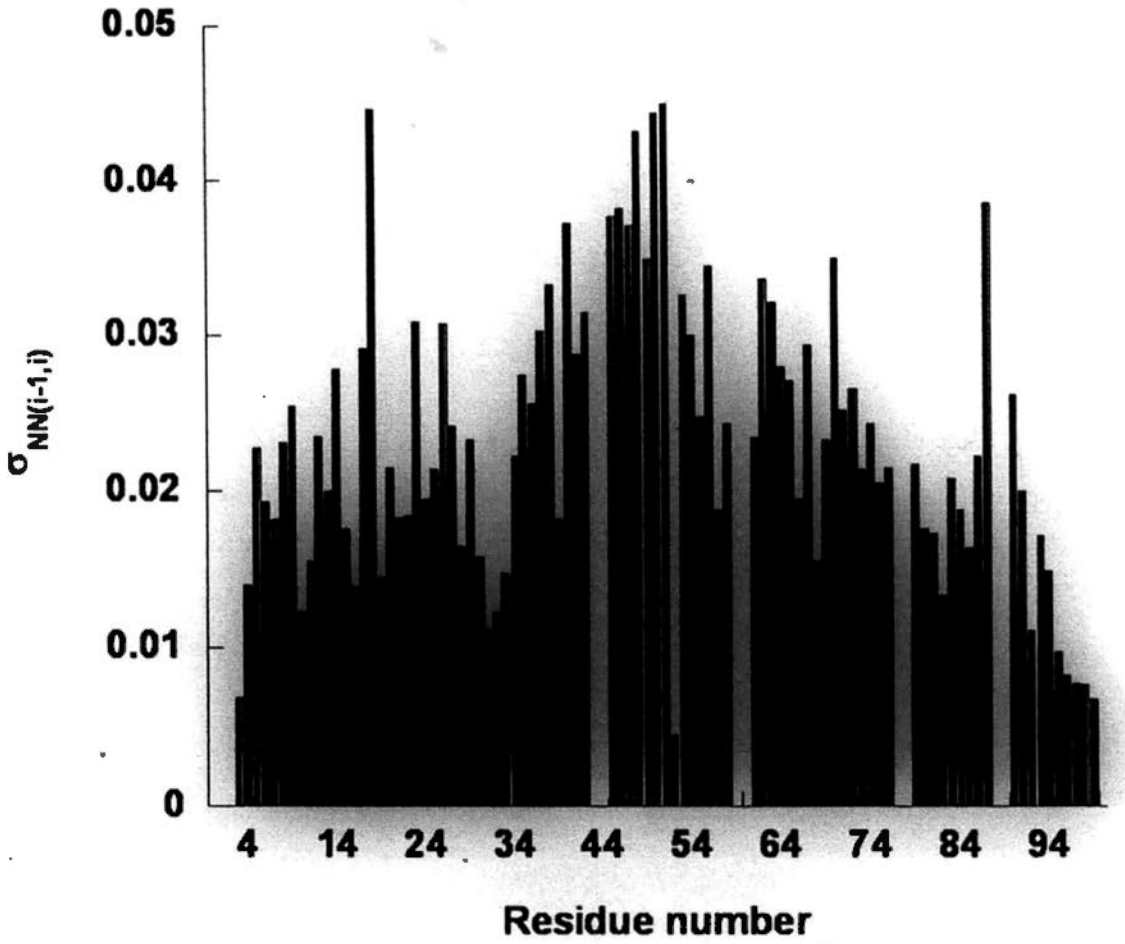


Figure 4.6. Backbone amide $^{15}N^{15}N$ intensity ratios. Backbone amide $^{15}N^{15}N$ intensity ratios ($\sigma_{NN(i-1,i)}$) observed for wild-type *T. celer L30e* in 6.0 M guanidine HCl at 298 K at pH 5.4. $\sigma_{NN(i-1,i)}$ were determined as the intensity ratio of $HN_{(i-1)}$-$HN_{(i)}$ NOE to diagonal peak based on 3D $^{1}H^{15}N$ HSQC-NOESY-HSQC spectrum.

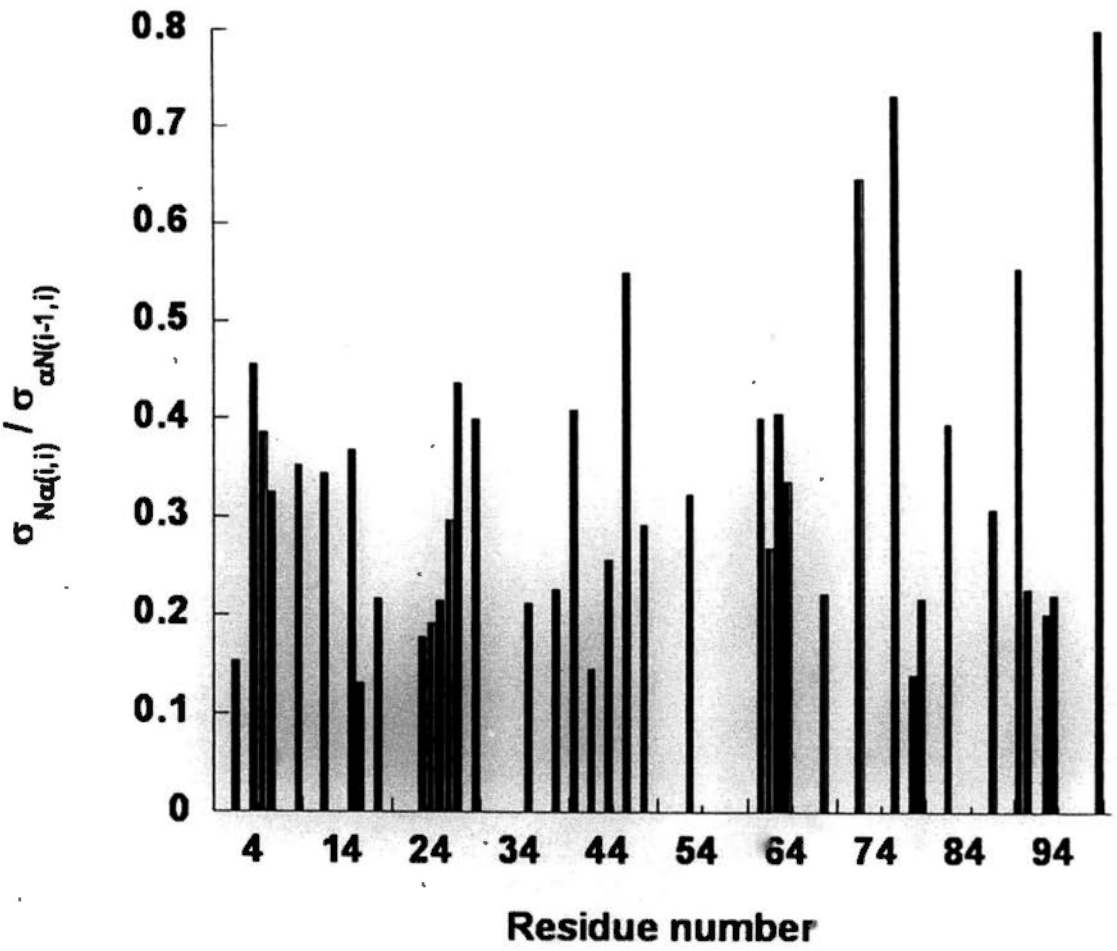


Figure 4.7. Backbone $^1H^{\alpha}{}^1H^{\alpha}$ intensity ratios. Backbone $^1H^{\alpha}{}^1H^{\alpha}$ intensity ratios ($\sigma_{N\alpha(i,i)} / \sigma_{\alpha N(i-1,i)}$) observed for wild-type *T. celer L30e* in 6.0 M guanidine HCl at 298 K at pH 5.4. $\sigma_{N\alpha(i,i)} / \sigma_{\alpha N(i-1,i)}$ values were determined as the intensity ratio of $H^{\alpha}_{(i)}$-$HN_{(i)}$ to $H^{\alpha}_{(i-1)}$-$HN_{(i)}$ NOEs measured in 3D ^{15}N TOCSY-HSQC spectrum and HC(C)(CO)NH-TOCSY spectrum.

large normalized $\sigma_{N\alpha(i,i)} : \sigma_{\alpha N(i-1,i)}$ ratio.

4.3 Discussion

4.3.1 *T. celer L30e* has no observable residual structures in the presence of 6.0 M Guanidine HCl

Unlike native state that has a energetically minimized stable structure, denatured state of a protein is an ensemble of inter-converting random coil conformations, of which their free energy levels are expected to be comparable. With similar free energy level, none of these inter-converting conformations should be majorly populated according to the Blotzmann distribution. On the other hand, polypeptide will have conformational preference if interactions like salt-bridges and hydrogen bonds are present in the denatured state, as these interactions will alter the free energy levels of some conformations. In other words, residual structures resist the change in conformation which results in a higher rigidity of the backbone. In this study, the rigidity of the backbone in denatured state was traced by the signal intensity of backbone amide in $^{1}H^{15}N$-HSQC experiment, as the signal intensity would be higher (and narrower line-width) if the backbone is more flexible in general. Therefore, the overall similar backbone amide signal intensities throughout the sequence strongly suggested *T. celer L30e* has no residual structures in denatured state, although the middle of the sequence may be a little bit more rigid than the two

terminals.

Although there is no preferred conformations if the denatured protein is purely random coil, Proproline II conformation was suggested to be an important type of conformation for the denatured proteins. Polyproline II type conformations have a characteristic that the combination of two backbone torsion angles (ϕ_i and ψ_i) are similar to that of the β-strand structures. The fact that random coil structures are predominantly populated by conformers in the β-region ϕ_i and ψ_i conformational space has been supported by other studies. These findings lead to one general principle for the detection of residual structures in denatured protein. Any residues having ϕ_i and ψ_i deviated from those of β-strand structure are regarded as involved in residual structures. Obviously, this definition of residual structure cannot distinguish any β-strand structures from random coil conformations. However, no report suggests β-strand structures are present in the denatured state. Although ϕ_i and ψ_i cannot be directly measured, their combination can be reflected in $d_{NN(i-1,i)}$, $d_{N_\alpha(i,i)}$, and $d_{\alpha N(i-1,i)}$, which can in turn be measured by intensities of NOEs as discussed above. Under this manner, the information from backbone atoms is enough for investigating residual structure. This is understandable as afterall, it is the polypeptide conformation being studied. Using NOEs as the probe for residual structures, our data strongly suggested *T. celer L30e* did not have any observable

residual structures in the denatured state.

Chemical shifts have long been known to be affected by secondary structure[171-174] as well as the nearby residues in the sequences[159-161]. Thus, residual structures can also be detected if the chemical shifts, especially for those of the backbone atoms, are deviated from the random coil shifts. The random coil shifts can be obtained from model peptide[159-161], or from the loop regions of native proteins[174]. In our study, the deviations of chemical shifts for ^{1}H, ^{15}N, H^{α}, C^{α}, C^{β} and CO from the sequence-corrected random coil shifts did not give consistent evidence to show that any of the residues in denatured *T. celer L30e* were involved in residual structures. However, the C-terminal residue Glu-100 has been shown to have significant deviation in ^{15}N, C^{α}, and H^{α} chemical shifts, and the N-terminal residue Met-0 has shown to have significant deviation in C^{β} chemical shift. This is unlikely for Met-0 and Glu-100 to have residual structures as this contradicts to the finding that the amide signal intensity is a bit higher at two terminals. Instead, the large deviations were likely because of the well known terminal effects which were not taken into account in our study. For the deviations in C^{α}, Arg-42, Ile-58, and Arg-76 were found to have large deviations. Be noted all these three residues are preceding a proline residue. In the sequential correction of random coil C^{α} shifts, the corrections for residues preceding a proline residue were as large as 2.0 ppm[161]. This correction

may not be accurate as the correction values are based on short model peptides only. Not only the corrections for the preceding residue may not be accurate, the random coil shifts for proline residue taken from short model peptides may not be accurate enough too. In the H^{α} chemical shift, Pro-43 was found to be significantly smaller than the random coil shift. The ^{1}H chemical shifts of the other three proline residues in the sequence, Pro-59, Pro-77 and Pro-88 also almost reached the lower boundary of the "random coil range". Therefore, other estimations of random coil shift like estimating from loop region of native protein[174] may be a better method for obtaining more accurate random coil shift. Other residues with a large deviation detected are all charged residues (Lys-28, Arg-54, His-78). Although their large deviations were not consistent in every backbone atoms, this observation gives us a cue that charge-charge interaction may be present in the denatured state, which may in turn affect the protein stability[88-91].

4.3.2 Concluding remarks

In this study, we have demonstrated *T. celer L30e* had completely denatured in the presence of 6.0 M guanidine HCl, and the pH values did not have large effect on its denatured state. Also, with the success of sequential assignment of all residues, any residual structures present in the denatured state were tried to be detected with the rigidity of backbone (backbone amide intensities along the sequence), structural

effect on chemical shifts (secondary shifts of backbone and nearby atoms), and the combination of the two backbone torsion angles ϕ_i and ψ_i (normalized $\sigma_{NN(i-1,i)}$, and $\sigma_{N\alpha(i,i)}$: $\sigma_{\alpha N(i-1,i)}$ ratio). No consistent evidence was found to show *T. celer L30e* has form residual structures in denatured state. However, it is possible that electrostatic interactions are affecting the "stability" of denatured state as the charge density in the sequence is unusually high (see Chapter 6), although no observable residual structures were confidently detected. Nevertheless, this study evidently suggested the high thermostability and largely reduced ΔC_p should not be mainly contributed by the residual structure of *T. celer L30e*.

Chapter 5: De-crystallization of wild-type *T. celer L30e* in low ionic strength NMR conditions

5.1 Introduction

It is well established that some proteins are readily crystallized under numerous solution conditions, whereas others appear to crystallize strictly in exact conditions. Undoubtedly it is a great advantage for structural determination if the protein can be easily crystallized in various conditions. However, crystallization can be confrontational to biophysical and functional analysis if the proteins crystallize in assays, especially to those which high protein concentration is needed, such as NMR. The effect of crystallization in assays does not differ much from that of amorphous precipitation. Both of them would result in the great reduction in protein concentration, but crystallization seems to have a larger effect. Having all extrinsic conditions being the same, the solubility of a protein solution with an amorphous solid phase was shown to be higher than that with a crystalline solid phase[175-177].

Assay conditions are usually very restricted. Very often, biologically relevant low ionic strength environment will be chosen for performing assays[178-180]. Especially for the studies which are sensitive to ionic strength such as the investigation of protein electrostatic properties, low ionic strength in assays should

be particularly well controlled. Unfortunately, low ionic strength conditions are also widely used for crystallization because of the salting in effect [181-184]. The success in crystallizing chymotrypsinogen without precipitants in low ionic strength condition [185] revealed the possibility that proteins crystallize in low ionic strength assay conditions.

Crystallization in low ionic strength assay conditions can seriously hinderthe functional characterizations of proteins. Different strategies were applied to prevent any unwanted crystallization. One of the straight forward strategies is to improve protein solubility by adding salt and additive in a trial and error manner, so that the original crystallization contacts are disrupted and protein crystallization is prevented. This strategy is usually effective, but it will inevitably change the ionic strength of buffer and therefore affect protein properties such as screening out of long range electrostatic interactions[51,146]. Removing surface charges by mutagenesis is another method for reducing the chance of crystallization [186,187]. However, the outcome is not guaranteed and sometimes removing surface charges may contrarily promote crystallization[188,189].

Since high charge density and increased surface charge-charge interactions are crucial to protein crystal stability[186,187], the crystallization problem should be more vexing to thermophilic proteins of which the surface has a high charge density as a

result of the compositional bias to charged amino acid[56]. A thermophilic protein *Thermococcus celer* ribosomal protein L30e serves as a good example to illustrate this point. *T. celer L30e* is a 100-residue thermophilic protein with very high surface charge density. There are 13 acidic and 14 basic residues unevenly distributed on its surface[190,191], which would enhance its crystal stability and promote crystallization. We have shown that wild type *T. celer L30e* would crystallize in low ionic strength conditions in this study. As discussed in chapter 3 and chapter 6, as well as in previous studies[54,99], electrostatic interaction of *T. celer L30e* is one of the important factors for its high protein stability. However, crystallization of *T. celer L30e* in low ionic strength conditions fatally hindered the investigation of its electrostatic contribution to protein stability by estimating the pK_a values of its surface charge through NMR experiments. Obviously, removing surface charges, and increasing protein solubility by adding salt or additive should be strictly avoided, as these kinds of treatment would alter the original electrostatic interactions for sure.

To our knowledge, no rational strategy has been reported for tackling unwanted crystallization in low ionic strength conditions to date. To this end, using *T. celer L30e* as an example, we present a general method for rational de-crystallization in low ionic strength conditions without affecting the buffer condition and the protein properties by charge-conserved Arg-to-Lys mutations.

To investigate the de-crystallizing power, the solubility and crystallizability in low ionic strength conditions of five single and a quintuple Arg-to-Lys variants were designed. In addition to the crystallizability, the structural, thermodynamic, and electrostatic properties of the variant were also examined. Our results supported the conclusion that this rational de-crystallization approach successfully prevented unwanted crystallization in low ionic strength conditions without affecting buffer conditions and protein characteristics. In addition to de-crystallization strategy, this study also demonstrated the important role of Arg in crystallization, providing a direction for improving crystallization by simple mutations of surface charges.

5.2 Results

5.2.1 Wild-type *T. celer L30e* crystallizes in low ionic strength conditions

To demonstrate the crystallization of wild type *T. celer L30e* in low ionic conditions (10 mM citrate/phosphate buffer), protein with high concentration (10 mg ml^{-1}) was dialyzed in low ionic strength condition at pH 2.5, 4.0, and 6.5. No precipitation could be observed at pH 2.5 and 4.0, whereas a lot of white precipitates were formed at pH 6.5 after dialysis. After centrifuged down the white precipitates, the residual protein concentration of the supernatant was found to be less than 1 mg ml^{-1}, indicating the white precipitates are wild-type *T. celer L30e*. Careful

observation under light microscope revealed the white precipitates were in fact microcrystals (Figure 5.1). It was found that microcrystals could readily re-dissolve in the same buffer with 1 M NaCl, but was insoluble in the original low ionic strength buffer at pH 6.5 even the solvent volume was increased 10 folds.

Different from amorphous precipitation, the crystalline solubility (residual protein concentration in supernatant) should only be mildly affected by initial protein concentration[177]. To test if the microcrystal formation be dependent on initial protein concentration, the same set of dialysis experiment was repeated using protein with initial concentration of 0.5 and 1 mg ml^{-1}. The same result was obtained as expected. Microcrystals formed at pH 6.5, but not at lower pH of 2.5 and 4.0. The only difference was that much less precipitates were observed at pH 6.5, which can be trivially explained by using less protein in the experiment. This preliminary result indicated low protein concentration (0.5 mg ml^{-1}), which is common for various biological assays, is already enough to trigger the crystallization of wild-type *T. celer L30e* in low ionic strength condition at neutral pH.

5.2.2 Multiple Arg-to-Lys substitution improves protein crystalline solubility significantly

To test for the de-crystallizability of charge-conserved Arg-to-Lys substitution, five single lysine-substituted variants with substitution site at Arg-8, Arg-21, Arg-42,

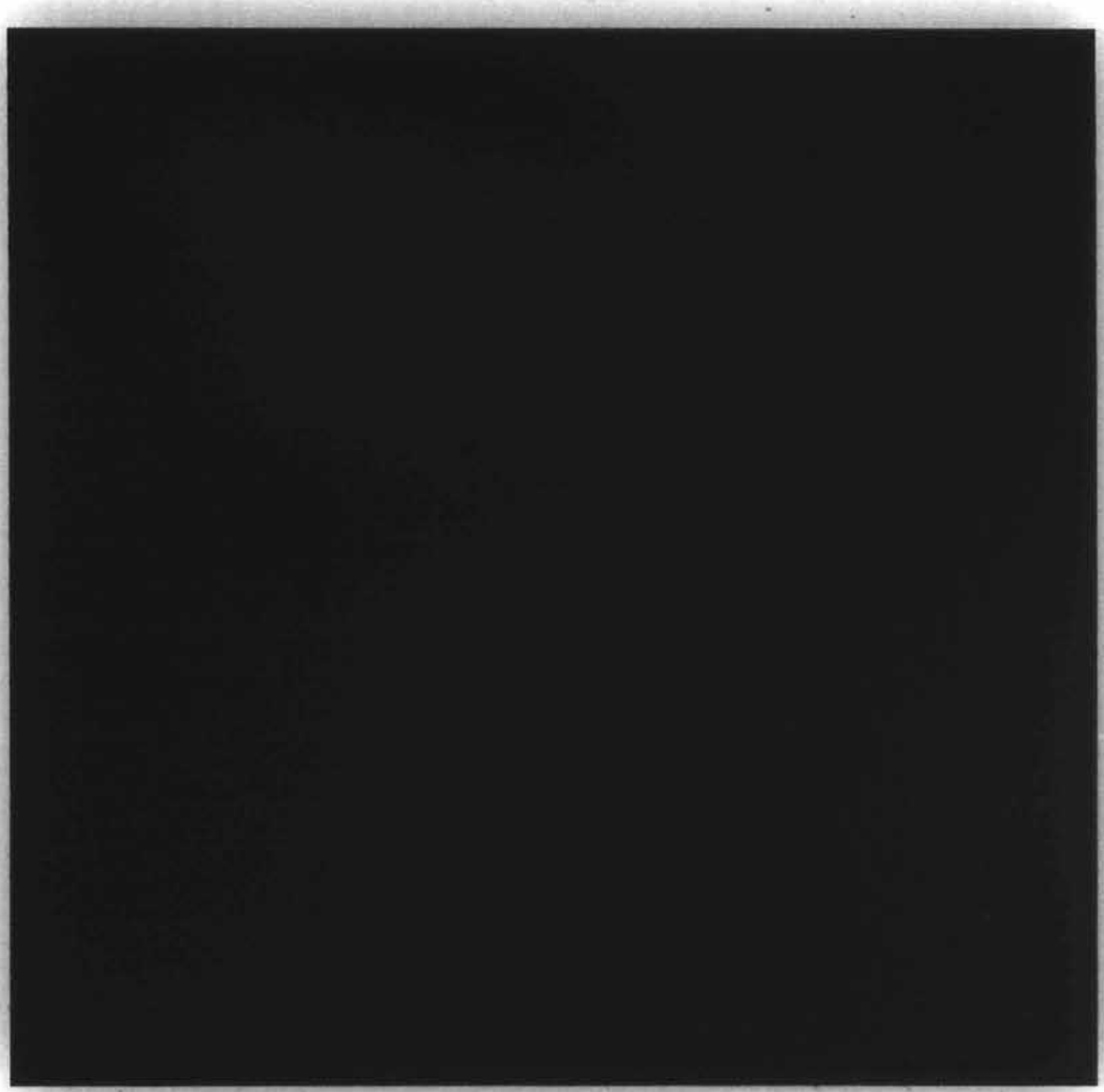


Figure 5.1. Microcrystals observed under light microscope. White precipitates obtained from dialysis of 10 mg ml^{-1} wild-type T. celer L30e in low ionic strength condition (10 mM citrate/phosphate buffer) at neutral pH of 6.5 were carefully observed under light microscope with polarizer equipped.

Arg-54, or Arg-76, as well as a quintuple Arg-to-Lys variant with all the mentioned five arginine substituted with lysine, were generated (Figure 5.2). The crystalline solubility of wild-type *T. celer L30e* was then compared with that of six Arg-to-Lys variants by two independent methods, namely dialysis approach and concentration approach (Figure 5.3). The results from two different approaches are consistent, except the solubilities found by concentration approach were higher than those found by dialysis approach by about 0.07 mg ml^{-1} in average, which was likely due to super-saturation of the protein sample. Nevertheless, both approaches pointed out that all five single Arg-to-Lys variants had their solubility increased mildly by about 0.2 to 0.3 mg ml^{-1} when compared to that of wild-type *T. celer L30e* (0.1 mg ml^{-1}). Although the absolute change in solubility were little, but the substitution has already improved the solubility by about 2 to 3 folds.

The most encouraging result came from the quintuple Arg-to-Lys variant. No white precipitants (microcrystals) were formed in both approaches. In the concentration approach, the protein sample was concentrated to about 28.2 mg ml^{-1} and still no precipitation formed (only 9.1 mg ml^{-1} in dialysis approach because it was limited by the experimental design). To demonstrate this high protein concentration was not the result of super-saturation, the concentrated protein sample

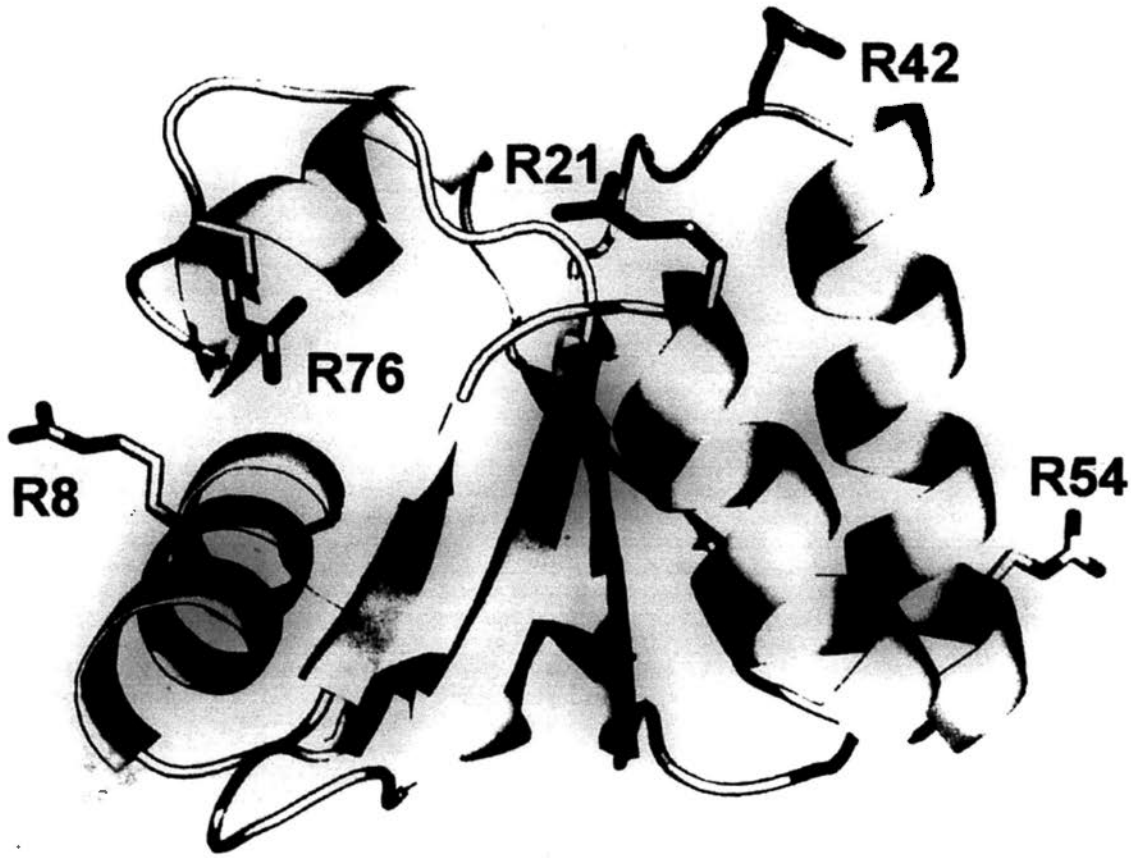


Figure 5.2. Location of the five substituted arginine residues. The five arginine residues were substituted by lysine in single and the quintuple Arg-to-Lys variants. The five substituted arginine residues were shown in sticks representation, and labeled with one-letter-code and residue number.

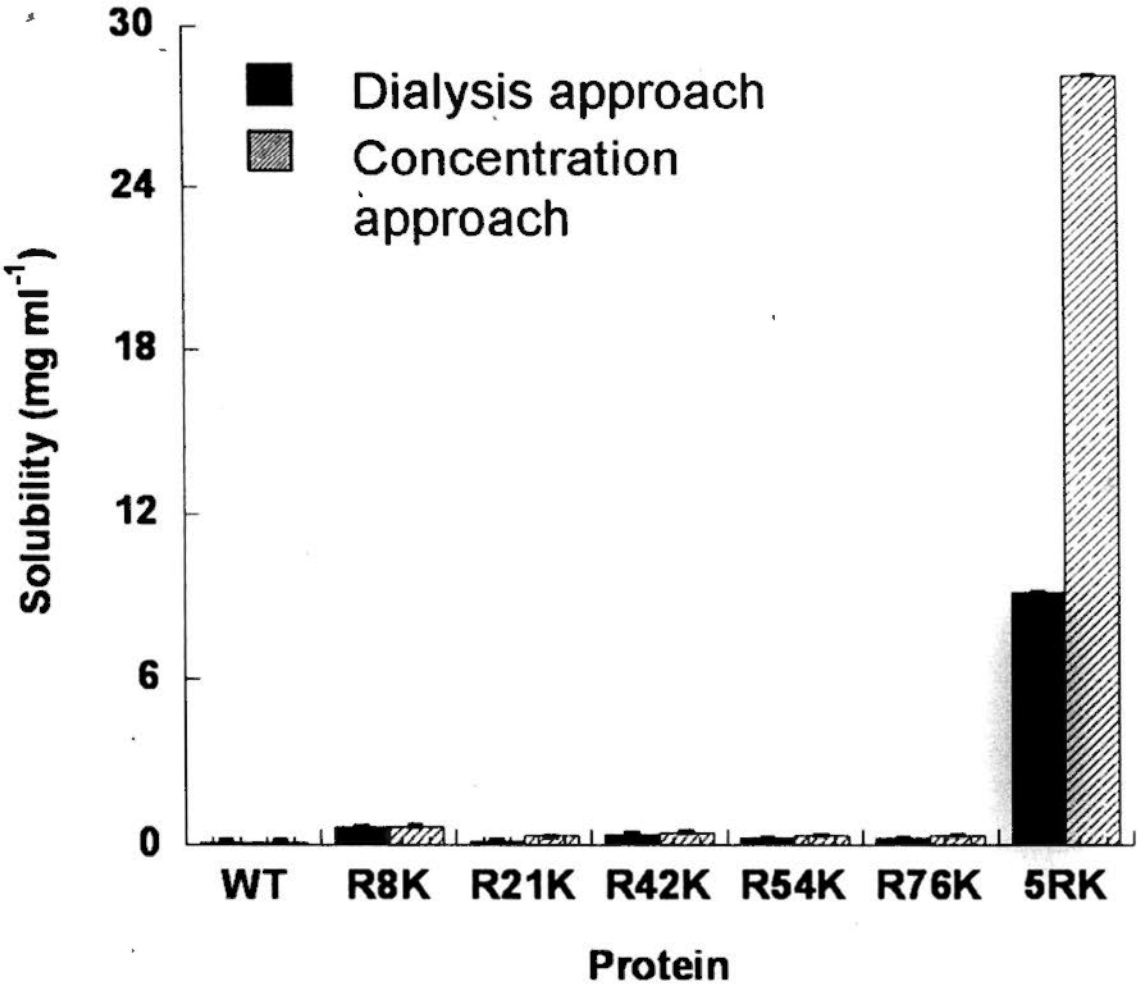


Figure 5.3. Crystalline solubility of *T. celer L30e* and its variants. Determination of crystalline solubility of wild-type *T. celer L30e* (WT), single Arg-to-Lys variants (R8K, R21K, R42K, R54K, and R76K), and the quintuple Arg-to-Lys variant (5RK) by dialysis approach (filled) and concentration approach (shaded). Be noted that the quintuple Arg-to-Lys variant had an extraordinarily high crystalline solubility when compared to that of wild-type protein and other singly substituted variants.

was equilibrated at 4 °C for a week. It was found that neither precipitate nor microcrystal was formed.

5.2.3 Arg-to-Lys substitution does not induce structural changes

Crystal structure of wild *T. celer L30e* at low ionic strength condition without precipitant has been solved in this study (Table 5.1). It showed no major structural differences from the previously solved crystal structures with using PEG as precipitant (PDB code: 17HM) [99,190], having the r.m.s.d. between C^{α} atoms of 0.44 Å (Figure 5.4).

To investigate if substitution of Arg with Lys would lead to any structural changes, crystal of the five single Arg-to-Lys variants were grown in low ionic strength conditions (Table 5.1). The crystal structures of all five single Arg-to-Lys variants could overlay very well to that of wild-type *T. celer L30e*, with the r.m.s.d. between C^{α} atoms smaller than 0.50 Å, indicating the substitutions did not induce major structural changes (Figure 5.4).

Since the quintuple Arg-to-Lys variant did not crystallize in low ionic strength condition without precipitant, sparse matrix screening was performed using commercial kits. The quintuple variant was found to crystallize only in one out of 396 conditions tested, where the crystallization condition had very high ionic

Table 5.1. Statistics for crystal structure of wild-type *T. celer L30e* and its variants

	Wild-type without precipitant	Quintuple Arg-to-Lys variant	R8K	R21K
Summary of crystallization conditions:				
	10 mM citrate / phosphate buffer pH 6.5, 289 K	10 mM citrate / phosphate buffer pH 6.5, 289 K	10 mM citrate / phosphate buffer pH 6.5, 289 K	10 mM citrate / phosphate buffer pH 6.5, 289 K
Diffraction data collection statistics:				
X-ray source	Cu Kα	Cu Kα	Cu Kα	Cu Kα
Space group	$P2_1$	$P22_12_1$	$P2_12_12_1$	$P2_12_12_1$
Resolution (Å)	31.9-2.4 (2.2-2.1)	39.4-2.4 (2.5-2.4)	34.7-1.9 (2.0-1.9)	33.7-1.8 (1.9-1.8)
Molecules per asymmetric unit	1	2	1	1
Unit cell dimension (Å)	a, 24.2; b, 53.3; c, 33.7	a, 30.5; b, 61.4; c, 102.8	a, 34.5; b, 42.6; c, 54.4	a, 34.5; b, 42.8; c, 54.5
Unit cell angles (deg.)	α, 90.0; β, 109.1; γ, 90.0	α, 90.0; β, 90.0; γ, 90.0	α, 90.0; β, 90.0; γ, 90.0	α, 90.0; β, 90.0; γ, 90.0
Multiplicity	3.5 (3.5)	14.7 (14.7)	3.3 (3.1)	6.9 (6.6)
Completeness (%)	99.3 (98.8)	100.0 (100.0)	99.3 (97.0)	99.5 (98.7)
Mean I/σ (I)	6.7 (3.9)	15.4 (8.7)	10.0 (4.5)	9.7 (4.6)
R_{merge} (%)	12.6 (26.4)	12.6 (26.8)	8.4 (22.6)	12.2 (29.4)
Unique reflections	3199 (456)	8086 (1149)	6656 (919)	7894 (1108)
Structural refinement statistics:				
R-factor / R_{free} (%)	17.7 / 27.8	18.2 / 25.0	15.6 / 21.2	15.1 / 19.0
r.m.s.d. from idea values:				
Bond distances (Å)	0.006	0.007	0.005	0.004
Bond angles (deg.)	0.893	1.069	0.816	0.793
Ramachandran pot analysis:				
Preferred region (%)	94.7	98.4	96.7	96.7
Allowed region (%)	5.3	1.6	3.3	3.3
Outliers (%)	0.0	0.0	0.0	0.0
C^α r.m.s.d. with wild-type T. celer L30e	0.44	0.47	0.40	0.39
Values in parentheses are for the highest-resolution shell.				

	R42K	R54K	R76K	
Summary of crystallization conditions:				
	10 mM citrate / phosphate buffer pH 6.5, 289 K	10 mM citrate / phosphate buffer pH 6.5, 289 K	10 mM citrate / phosphate buffer pH 6.5, 289 K	
Diffraction data collection statistics:				
X-ray source	Cu Kα	Cu Kα	Cu Kα	
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_1$	
Resolution (Å)	34.6-2.3 (2.4-2.3)	33.6-2.3 (2.4-2.3)	53.3-2.2 (2.3-2.2)	
Molecules per asymmetric unit	1	1	1	
Unit cell dimension (Å)	a, 34.6; b, 43.1; c, 54.4	a, 34.6; b, 42.7; c, 54.6	a, 24.4; b, 53.3; c, 34.0	
Unit cell angles (deg.)	α, 90.0; β, 90.0; γ, 90.0	α, 90.0; β, 90.0; γ, 90.0	α, 90.0; β, 109.3; γ, 90.0	
Multiplicity	2.8 (2.8)	6.7 (6.7)	4.8 (4.8)	
Completeness (%)	99.8 (99.6)	100.0 (100.0)	99.8 (99.5)	
Mean I/σ (I)	7.8 (4.4)	11.3 (6.6)	19.4 (5.4)	
R_{merge} (%)	10.4 (24.4)	12.4 (27.4)	6.6 (26.7)	
Unique reflections	3905 (540)	3910 (547)	4232 (606)	
Structural refinement statistics:				
R-factor / R_{free} (%)	15.4 / 23.6	17.1 / 25.5	19.9 / 26.9	
r.m.s.d. from idea values:				
Bond distances (Å)	0.005	0.005	0.008	
Bond angles (deg.)	0.803	0.831	1.096	
Ramachandran pot analysis:				
Preferred region (%)	95.8	95.9	94.8	
Allowed region (%)	4.2	4.1	5.2	
Outliers (%)	0.0	0.0	0.0	
C^{α} r.m.s.d. with wild-type T. celer L30e	0.40	0.41	0.43	
Values in parentheses are for the highest-resolution shell.				

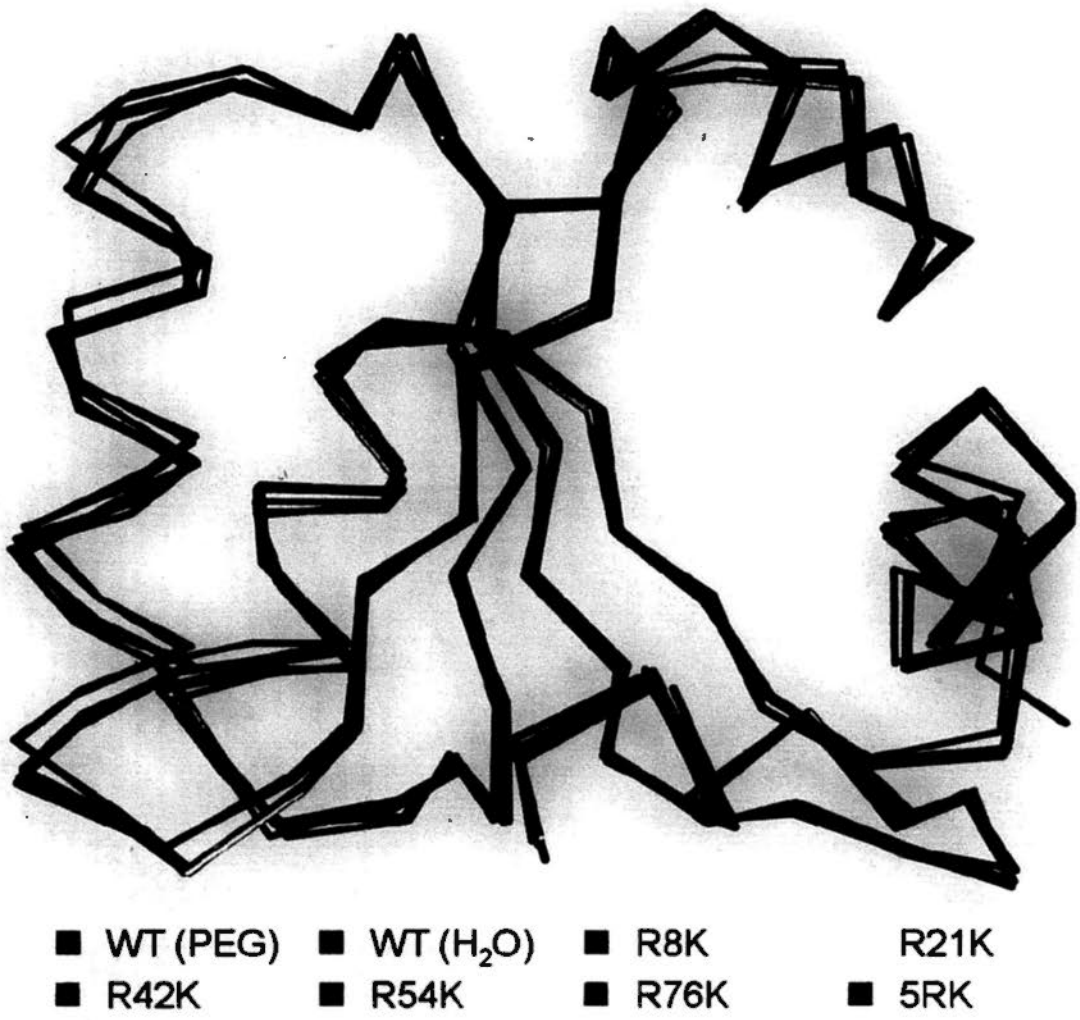


Figure 5.4. Crystal structures of wild-type T. celer L30e and its variants. Crystal structure of wild-type T. celer L30e of which the crystals were grown without precipitants (WT (H2O), grey), as well as the crystal structures of R8K (green), R21K (yellow), R42K (magenta), R54K (cyan), R76K (orange), and the quintuple Arg-to-Lys variant (5RK, red) are superimposable upon that of wild-type T. celer L30e of which the crystals grown with precipitant PEG (black) (PDB: 1H7M).

strength. The crystal structure of the quintuple variant in high ionic strength condition was solved and no major structural change was found (the r.m.s.d. between C^{α} atoms was 0.47 Å) (Figure 5.4). The fact that quintuple variant adopts essentially the same fold as the wild type *T. celer L30e* in low ionic strength is supported by the spectroscopic evidence that their far-UV CD spectra in low ionic strength are almost identical in shape (Figure 5.5).

5.2.4 Arg-to-Lys substitution changes the crystal packing

Although no major structural changes were observed, the crystal packings of Arg-to-Lys variants in low ionic strength condition without precipitant were different from that of wild-type *T. celer L30e*. The crystal contacts at ariginine position (8, 21, 39, 42, 54, 76, and 92 in sequence) in low ionic strength condition and at neutral pH were investigated and summarized in Table 5.2.

The space group of wild-type *T. celer L30e* crystal without adding PEG as precipitant was found to be $P2_1$. In this space group, it was found that Arg-8, Arg-54, and Arg-76 were involved in crystal contact by forming salt bridges (< 4.0 Å with unlike charges in other molecules). While Arg-54 and Arg-76 only formed salt bridges in crystal contacts, Arg-8 was also found to form hydrogen bonds in crystal contact. For Arg-21, although it did not form strong salt bridges in crystal contact, it was in close proximity with the negatively charged surface of other molecule formed

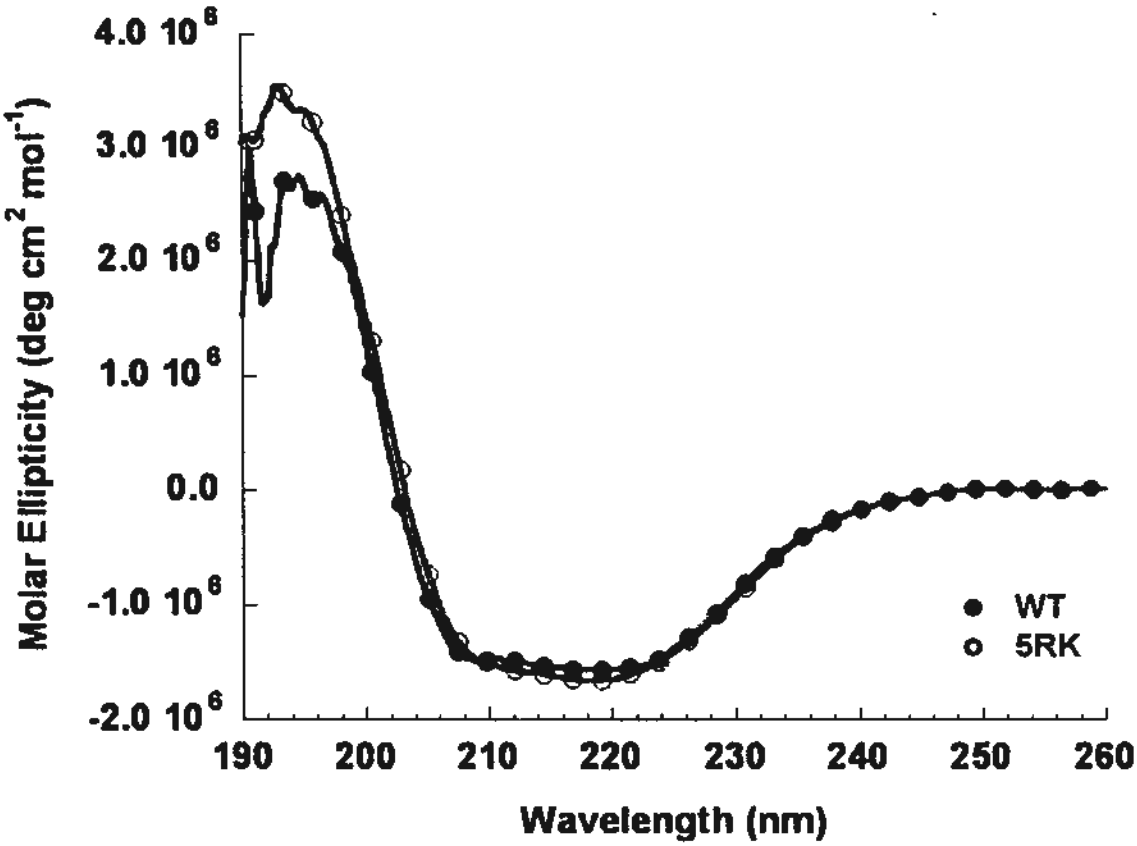


Figure 5.5. Far-UV CD spectra of wild-type *T. celer L30e* and its variant. Far-UV CD spectra of wild-type *T. celer L30e* and the quintuple Arg-to-Lys variant in low ionic strength condition. The far-UV CD spectra from 260 nm to 190 nm of wild-type *T. celer L30e* (filled circle) and the quintuple Arg-to-Lys variant (open circle) with the same protein concentration were obtained in 10 mM citrate/phosphate buffer pH 6.5 at 298 K.

Table 5.2. Summary of crystal contacts of *T. celer L30e* and its variants.

Protein	Residue number						
	8	21	42	54	76	39	92
WT	SB: D48 HB: Y52		HB: L94	SB: D2	SB: E62 SB: E64		
R8K			HB: L94	SB: D2	SB: E62 SB: E64 HB: Y61	SB: D12 HB: Q11	
R21K	SB: D48	SB: E50	HB: L94	SB: D2	SB: E62 SB: E64 HB: Y61	SB: D12 HB: Q11	
R42K	SB: D48 HB: Y52			SB: D2	SB: E62 SB: E64 HB: Y61	SB: D12 HB: Q11	
R54K	SB: D48		HB: L94		SB: E62 SB: E64 HB: Y61	SB: D12 HB: Q11	
R76K	SB: D48 HB: Y52		HB: L94	SB: D2	SB: E62		
5RK	HB: Y52				SB: D12	SB: E6	SB: D48

The type of crystal contacts were indicated by SB (salt bridge) or HB (hydrogen bond) before colon, while the interacting residues in other molecule were indicated by one-letter-code and the residue number.

by Glu-47, Glu-69, Asp-87, and Glu-90. Virtual mutagenesis by adopting different rotamers showed that Arg-21 has potential to form salt bridges with Glu-47 and Asp-87 (separation distance are 5.8 Å and 8.0 Å respectively). Instead of having strong electrostatic interaction, Arg-42 was found to form several hydrogen bonds to backbones of other molecules. No clear crystal contact can be observed for Arg-39 and Arg-92. R76K was also found to have the same space group of $P2_1$ with similar unit cell dimensions. The crystal contacts are basically the same. The salt bridges originally formed by Arg-76 (with Glu-62 and Glu-64) were retained by Lys-76.

However, the space groups of R8K, R21K, R42K and R54K were changed to $P2_12_12_1$, and the size of unit cell was increased. The crystal contacts in this space group are basically the same as those in $P2_1$ space group. However, according to the structures, all the substituted sites were found not to be able to form back salt bridges and hydrogen bonds, except Lys-21 could form a new salt bridge with Glu-50 to compensate the original electrostatic interactions with other unlike charges. Arg-39, which does not contribute to crystal contact in $P2_1$ space group, was found to form a new crystal contact by forming salt bridge with Asp-12. Other differences in crystal contacts were small. Arg-76 was found to form hydrogen bonds with Tyr-61 in addition to the original salt bridges. Also, Arg-21 was found to face a new negatively charged surface, having other acidic charged residues in close proximity (Asp-2 and

Glu-64, with separation distance of 5.8 Å and 4.3 Å respectively).

For the quintuple Arg-to-Lys variant, it crystallized in a new space group of $P22_12_1$. As expected, since five original crystal contact involving arginine residues were substituted, the overall crystal contacts differed much from those in $P2_1$ and $P2_12_12_1$ space group. In this space group, the Lys-8 and Lys-76 were found to form hydrogen bond and salt bridge respectively. Other substituted lysine residues, Lys-21, Lys-42, and Lys-54 were found not to contribute to crystal contact. The remaining retained arginine residues, Arg-39, and Arg-92, which did not form crystal contact in wild-type *T. celer L30e* ($P2_1$) originally, were found to form salt bridges and hydrogen bonds in crystal contacts.

5.2.5 Arg-to-Lys substitution does not affect thermodynamic stability of *T. celer L30e*

To explore the effect of Arg-to-Lys substitution on thermodynamic stability of proteins, the melting temperature (T_m) value and the Gibbs free energy of unfolding (ΔG_u) at pH 6.5 of wild-type *T. celer L30e* were compared with those of the quintuple Arg-to-Lys variant (Figure 5.6).

T_m values, which indicate the thermal stability of proteins, in low ionic strength condition were found by CD spectrum of temperature-induced denaturation of the

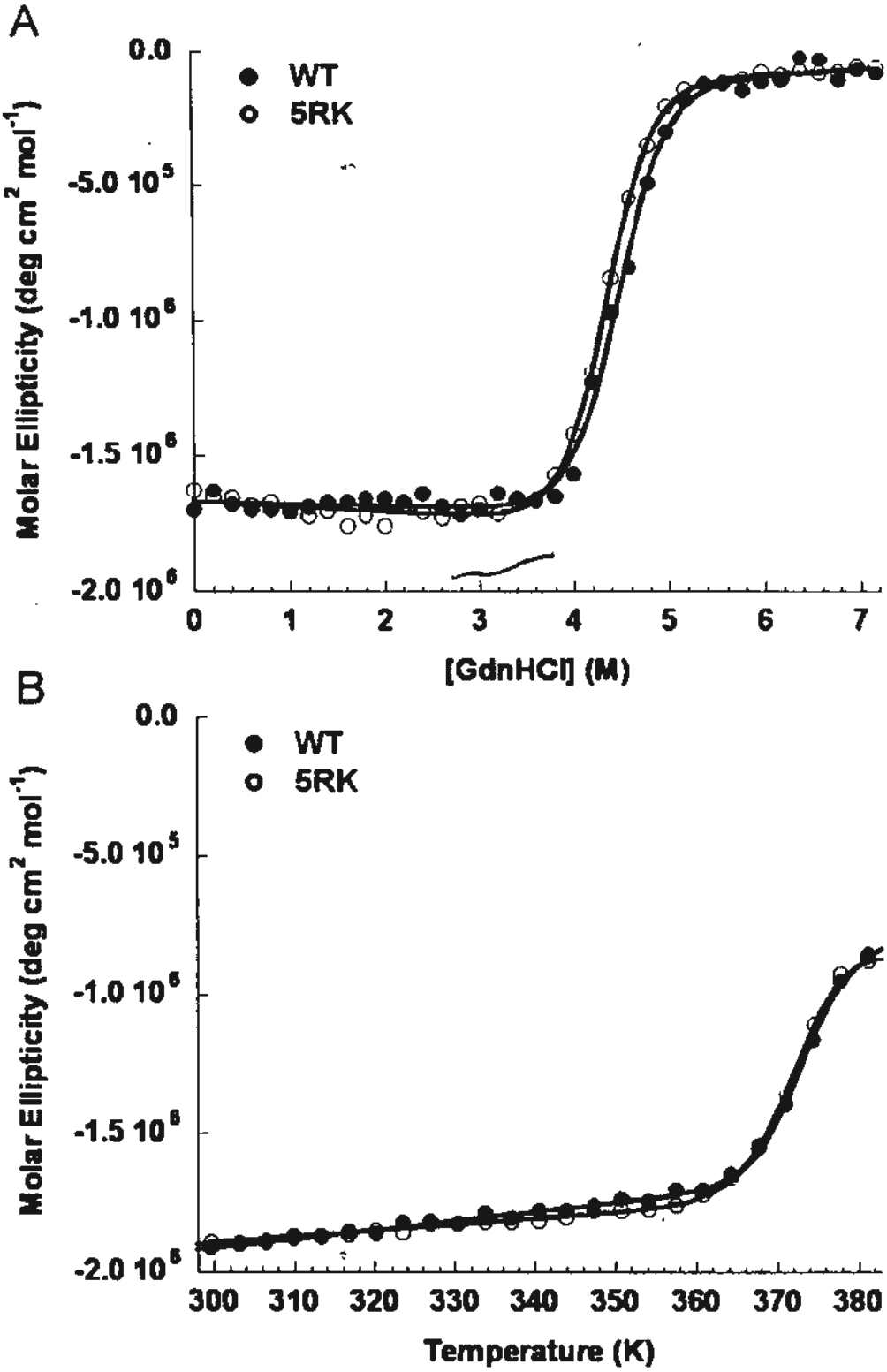


Figure 5.6. Chemical and thermal denaturation of *T. celer L30e* and its mutants. (A) Urea-induced denaturation curves and (B) thermal denaturation curves of wild-type *T. celer L30e* (filled circle) and the quintuple Arg-to-Lys variant (open circle) obtained in 10 mM citrate/phosphate buffer, pH 5.4, at 298 K.

wild-type protein and the variant. The T_m values of wild-type *T. celer L30e* and its quintuple Arg-to-Lys variant were found to be 372.8 ± 0.1 K and 373.2 ± 0.1 K respectively, showing that the quintuple substitution virtually had no effect on T_m value (ΔT_m equals to 0.4 ± 0.0 K). Both of the proteins were shown to be thermophilic proteins.

Similarly, ΔG_u at 25 °C in low ionic strength condition of both proteins were found by CD spectrum of guanidine-induced denaturation. The ΔG_u values of the wild-type *T. celer L30e* and the quintuple Arg-to-Lys variant were 42.9 ± 0.4 kJ mol^{-1} and 43.5 ± 0.4 kJ mol^{-1} respectively. The small difference of the two ΔG_u (0.6 ± 0.6 kJ mol^{-1}) indicates the quintuple substitution virtually has no influence on conformation stability of proteins.

5.2.6 Charge-conserved Arg-to-Lys substitution does not affect the salt dependency of T_m value

The pK_a values of the side chain of Arg and Lys are expected to be around 10.0 to 12.5. With this high pK_a values, it is expected that substituting ariginine by lysine residue would retain the overall charges of the protein. To investigate the effect of charge-conserved Arg-Lys substitution on the overall electrostatic interactions, salt dependencies of thermostability at pH 6.5 of wild-type *T. celer L30e* and all the single and quintuple Arg-to-Lys variants were determined (Figure 5.7). All proteins

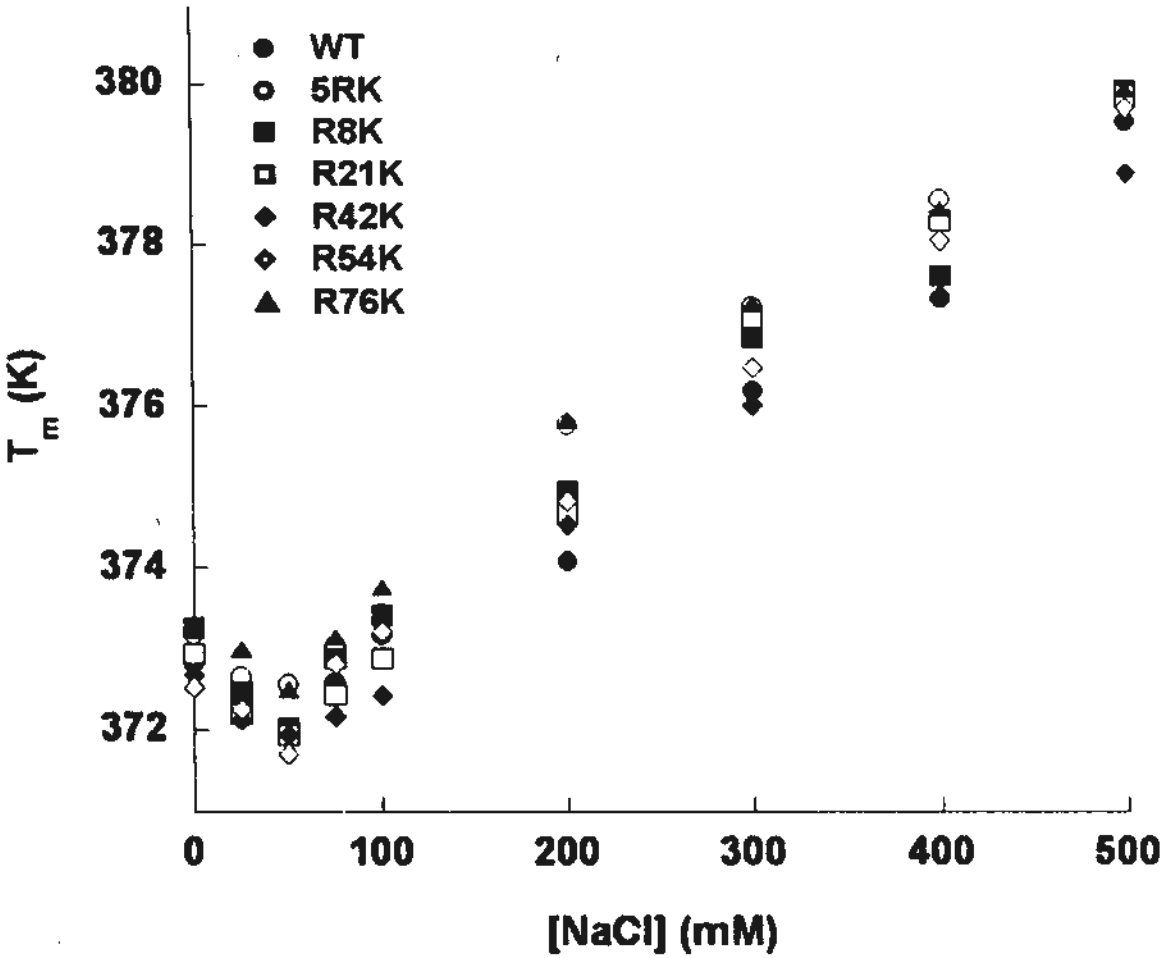


Figure 5.7. Salt dependency of melting temperature. Salt dependency of melting temperature (T_m) of wild-type *T. celer L30e* (filled circle), the quintuple Arg-to-Lys variant (open circle) , R8K (filled square), R21K (open square), R42K (filled diamond), R54K (open diamond), and R76K (filled triangle) in 10 mM citrate/phosphate buffer at pH 6.5 at 298 K. Be noted the virtually identical salt dependency of T_m indicates charge-conserved Arg-to-Lys substitutions did not affect surface charge-charge interaction, which was shown to have large effect on T_m[54].

had an almost identical salt dependency of melting temperatures (difference between T_m values is about 1 to 2 K at every examined NaCl concentration), which is a summation of screening of electrostatic interactions and the stabilizing Hofmeister effect. The T_m values were decreased in 0-50 mM of NaCl due to the screening of electrostatic interactions, but increased at higher concentrations of salt where the Hofmeister effect dominated.

5.2.7 The quintuple Arg-to-Lys variant has high tolerance to temperature- and pH-induced denaturation

To test the tolerance of the quintuple Arg-to-Lys variant to temperature- and pH-induced denaturation in a two-dimensional manner, T_m values at pH ranging from 1.0 to 6.5 were found by CD spectrum (Figure 5.8). It was found that T_m values decrease with the decrease in pH from 6.5 to about 2.0, and the decreasing rate of T_m increases when the pH value is linearly decreased. It was because when pH decreases, a larger population of acidic residues shifted to the protonated state in equilibrium and the balance of charge of *T. celer L30e* broke down. With diminished favorable electrostatic interactions between acidic and basic residues, the unfavorable charge-charge interactions between basic residues dominated and therefore the protein is destabilized in lower pH. However, when pH further decreases from 2.0 to 1.0, the T_m values contrarily increased. This observation demonstrated the

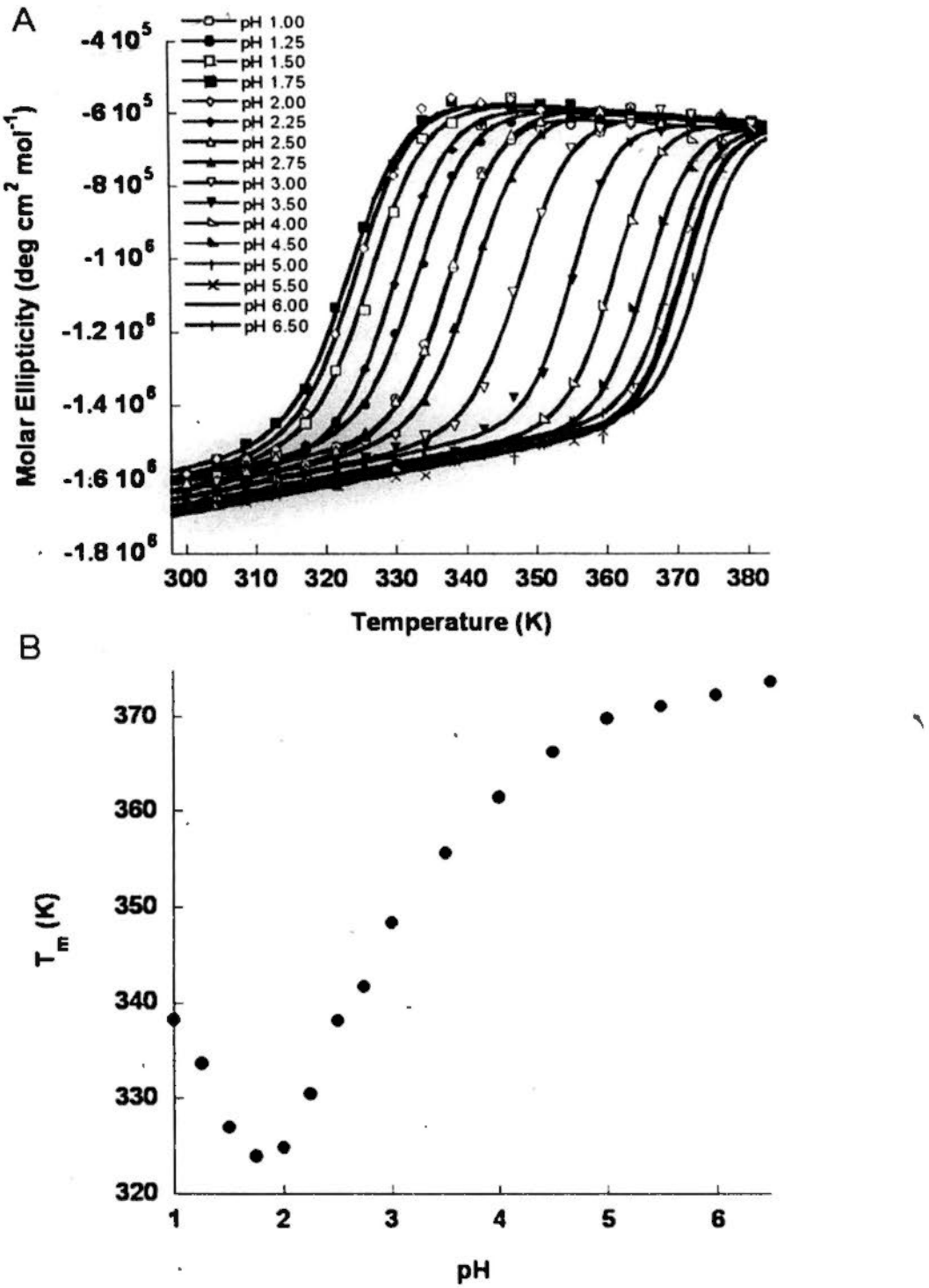


Figure 5.8. pH dependency of melting temperatures. pH dependency of melting temperature of the quintuple Arg-to-Lys variant in NMR conditions. (A) Thermal denaturations of the quintuple Arg-to-Lys variant at pH 1.00 (open circle), 1.25 (filled circle), 1.50 (open square), 1.75 (filled square), 2.00 (open diamond), 2.25 (filled diamond), 2.50 (open triangle), 2.75 (filled triangle), 3.00 (open inverted triangle), 3.50 (filled inverted triangle), 4.00 (open right angled triangle), 4.50 (filled right angled triangle), 5.00 (plus sign), 5.50 (cross), 6.00 (horizontal dash), and 6.65 (vertical dash) were performed. (B) T_m values found from thermal denaturation were plotted against pH to obtain the pH dependency of T_m.

Hofmeister effect. As the pH value of 10 mM citrate/phosphate buffer in acidic conjugation is around 2.4, strong acid (HCl in our study) have to be added in order to achieve lower pH. Therefore, when pH is lower than about 2.4, the salt (Cl^-) concentration will increase with decreasing pH, and thus the inevitable Hofmeister effect will increase with decreasing pH. Nevertheless, our data clearly demonstrated the quintuple Arg-to-Lys variant can remain stable, which indicated by the large Tm values at various pH (T_m value become lowest of 50.9 °C at pH 1.75), at pH ranging from 1.0 to 6.5 in the low ionic strength condition

5.3 Discussions

5.3.1 Crystallization of *T. celer L30e* in low ionic strength condition is electrostatic driven

It has demonstrated that wild-type *T. celer L30e* can be crystallized in low ionic strength condition at neutral pH 6.5 without precipitant (Figure 5.1). We have also shown a low protein concentration of 0.5 mg ml^{-1} (and probably lower than 0.1 mg ml^{-1} as indicated by the protein concentration of the supernatant after removing microcrystals) is already enough to trigger the crystallization.

It has been reported that surface charges play an important role in crystal stability[186,187], and therefore thermophilic protein with unusually high surface charge density, like wild-type *T. celer L30e*, is highly suspectable to be able to crystallize in

low ionic strength condition without the help of any precipitant. To show the role of electrostatic interaction in low ionic strength crystallization, the pH dependency of crystallization was found. Having checked *T. celer L30e* did not denatured at pH ranging from 1.0 to 6.5, the observation that *T. celer L30e* only crystallized at neutral pH 6.5 but not the acidic pH 2.5 and 4.0 clearly showed that the crystallization is electrostatic driven. *T. celer L30e* has an overall balanced net charge unevenly distributed on the surface[190,191]. At neutral pH, all acidic and basic residues were charged. As the screening effect was negligible in the low ionic strength condition (< 50 mM NaCl or equivalent, see Figure 5.7), the inter-molecular charge-charge interaction was maximized for crystallization. However, when the pH was low (< 4.0), larger population of acidic residues would be protonated, weakening the inter-molecular electrostatic interaction and therefore the protein is unable to crystallize. The important role of electrostatic interaction in crystallization was further supported by the fact that the microcrystals formed in low ionic strength condition were readily re-dissolved at high ionic strength condition (1 M NaCl), keeping other physical and chemical conditions the same.

In fact, the crystallization of wild-type *T. celer L30e* in low ionic strength condition is an extreme example for "salting in". Since the inter-protein electrostatic interaction was very strong, water molecules failed to screen out the inter-protein

interaction even when protein concentration was low (0.5 mg ml^{-1}). However, as the quintuple Arg-to-Lys variant, which has the same net charge and charge distribution as the wild-type protein has, did not crystallize in low ionic strength condition, it is shown that the strong inter-protein electrostatic interaction is essential, but not sufficient for this "salting in" crystallization. The quintuple Arg-to-Lys variant has rationally designed for disrupting the crystal contacts in the wild-type protein. Therefore, we can conclude that high specificity of electrostatic interaction is another factor that is essential for the "salting in" crystallization.

5.3.2 Arg is preferred in crystallization because of its high propensity of forming salt bridge and hydrogen bond

Detail analysis of the crystal structures revealed most of the arginine residues were involved in forming crystal contacts in wild-type *T. celer L30e*. Although no structural changes were observed in the crystal structures of all Arg-to-Lys variants, the crystal packing in low ionic strength condition without precipitant was found to be influenced by charge conserved Arg-to-Lys substitution. Single Arg-to-Lys substitution at crystal contact was shown to be enough to trigger the protein to adapt an alternative crystal packing. Our data clearly demonstrated Arg played an important role in crystallization.

Undoubtedly, electrostatic interaction played an important role in crystallization

in low ionic strength conditions as discussed above. Although both Arg and Lys carry a positive charge, in general Lys residues seldom appear in crystal contacts, whereas Arg residues are prominently used[192]. In fact, substitution of lysine residues with arginine has been suggested for improving crystallization [192,193]. There must be some intrinsic differences between arginine and lysine residue for the former to be more favourable to crystallization. Side chain entropy should be excluded from consideration because both arginine and lysine have very similar side chain conformational entropy of about 7.9 kJ mol^{-1} [194].

Another possible reason why arginine is more favourable to crystallization is that instead of having a small amine group in lysine residue, it has a large planar guanidinium group, which is much more readily to form salt bridge and hydrogen bond in crystal contacts. The strong ability of arginine residue for forming salt bridges and hydrogen bonds was also observed in wild type *T. celer L30e*. While Arg-21, Arg-54, and Arg-76 mainly formed electrostatic interactions in crystal contacts, Arg-42 contributed to crystal stability by forming hydrogen bonds. Arg-8 was found to stabilize the crystal by forming both hydrogen bond and salt bridge. Except for Arg-76, when one of these crystal contacts involving Arg is replaced by Lys, the stability of crystal with original packing (space group of $P2_12_12_1$) decreased and therefore protein crystallized in a new packing ($P2_12_12_1$) instead. In this new crystal

packing, although the substituted Lys in general could not form back salt bridges or hydrogen bonds (substituted Lys-21 was the only exception that it could form a new salt bridge with Glu-50, which did not interact with Arg-21 in wild-type protein), the loss of crystal stability was compensated by the new crystal contact formed by Arg-39 (salt bridges with Asp-12), which did not form any crystal contact in wild-type *T. celer L30e* crystal. The reason for no packing rearrangement observed in R76K is not because Arg-76 contributes only very litle to crystal stability, but because the substituted Lys-76 can retain the original salt bridge formed. Therefore, the crystal stability in the original crystal packing ($P2_1$) was not affected much by the Arg-to-Lys substitution.

Since all the 5 original crystal contact involving Arg in the quintuple Arg-to-Lys variant were substituted, the crystal contacts and space group ($P22_12_1$) were changed significantly as expected. One interesting observation is that the remaining arginine residues in the variant protein, Arg-39 and Arg-92, which did not form any crystal contacts in wild-type crystal packing ($P2_1$), formed salt bridge and hydrogen bonds in the new crystal packing. It seems that *T. celer L30e* has a very high preference for using Arg in crystal contact as much as possible. When one or more crystal contact involving Arg are replaced (even by charge conserved Lys), protein will then adapt a new crystal packing so that the remaining Arg will form new crystal contact by

forming salt bridge or hydrogen bonds. The Arg-to-Lys-induced packing rearrangement indicates clearly the importance of Arg in forming crystal contact not only because of its charge, but also its high propensity for forming salt bridge and hydrogen bond.

5.3.3 Multiple Arg-to-Lys substitutions successfully de-crystallize *T. celer L30e* in low ionic strength condition without changing its structural, thermodynamic, and electrostatic properties

By replacing five Arg residues in wild-type *T. celer L30e* with Lys, the protein solubility was increased for at least more than 50 folds, and no microcrystals were formed in low ionic strength condition. The quintuple Arg-to-Lys variant was found to have no major structural changes when compared with the crystal structure of wild-type *T. celer L30e*. Also the guanidine-induced ΔG_u, T_m, as well as the pH dependency of T_m of the quintuple Arg-to-Lys variant in low ionic strength condition were found to essentially not differ from those of wild-type protein.

Unwanted protein crystallization in assays will be an insurmountable difficulty for protein characterization. In our study of electrostatic contribution, the crystallization of protein in low ionic strength NMR condition made the wild-type *T. celer L30e* impossible to be used as the protein model for pK_a approach. Since our

purpose is to study the electrostatic interaction of the protein, general methods like removing charged residues[177] or changing buffer conditions (ionic strength)[195] that would inevitably influence protein charge-charge interaction must be avoided. As discussed above, the rational de-crystallization of electrostatic driven *T. celer L30e* crystal was achieved by decreasing the propensity of forming salt bridges and hydrogen bonds, keeping the surface charge unchanged. To our knowledge, this study is the first report to present a rational strategy to overcome this problem without changing assay condition and affecting protein properties.

To be a good substitute of wild-type *T. celer L30e* for pK_a shift approach, the quintuple Arg-to-Lys substitutions should not induce any changes in structure, thermodynamic stability, and electrostatic interaction in addition to the number of charges when compared to wild-type protein. The crystal structure in high ionic strength together with the far-UV CD spectrum in low ionic strength of the quintuple Arg-to-Lys variant evidently suggested those Arg-to-Lys mutation did not have any effect on overall protein structure. One reason why the electrostatic interaction of *T. celer L30e* is interesting is because of its high thermostability. The comparable magnitude of ΔG_u values obtained from guanidine-induced denaturation and T_m values obtained from thermal-induced denaturation showed clearly the quintuple Arg-to-Lys variant retains the high thermostability. Lastly, virtually the same salt

dependency of T_m value of wild-type protein, and its single and quintuple Arg-to-Lys variants indicated none of the single Arg-to-Lys mutation has any effect on overall protein charge-charge interaction, i.e. the same salt dependency of T_m value of the quintuple Arg-to-Lys variant as that of wild-type protein was not due to the averaging effect of several largely charged salt dependencies from different single Arg-to-Lys variant. In addition, the very high tolerances to pH- and temperature-induced denaturation make the quintuple Arg-to-Lys variant very suitable for pK_a shift approach study.

5.3.4 Concluding remarks

Wild-type *T. celer L30e* was shown to be crystallized in low ionic strength NMR condition in neutral pH 6.5, which hindered its electrostatic study by pK_a approach. The pH and salt dependencies of the crystallization indicated this "salting in" crystallization is electrostatic driven. Although changing the buffer condition (ionic strength) and protein surface charges sound to be a good approach for de-crystallization, they are highly unappreciated in our electrostatic study. Therefore, according to the wild-type protein crystal structure, a charge-conserved quintuple Arg-to-Lys variant was rationally designed for preventing the "salting in" crystallization by altering surface propensity for forming salt bridges and hydrogen bonds. The solubility of the quintuple Arg-to-Lys variant was shown to be largely

increased by at least 50 folds, and the "salting in" crystallization was prevented. In addition, the quintuple Arg-to-Lys variant exhibits virtually the same structural, thermodynamic, and electrostatic properties as those of wild-type *T. celer L30e*. Together with its high tolerances to pH- and temperature-induced denaturation, the quintuple Arg-to-Lys variant was an excellent protein model for electrostatic study by pK_a shift approach.

In addition to report the first rational de-crystallization in low ionic strength condition without adding precipitant as a general method, we have also showed that Arg is highly preferable for the crystallization of *T. celer L30e*. This observation also provides us with hints on rational design for improving protein crystallizability.

Chapter 6: The contribution of global electrostatic interactions to protein stability

6.1 Introduction

Electrostatic interaction has long been suggested to play a crucial role in thermostability[94,95]. Continuous effort has been paying for investigating the relationship between charge-charge interaction and protein stability by mutagenesis. Many studies have been reported to support the role of electrostatic interaction in protein stability[50,52]. Our previous study on charge-to-Ala substitution of *T. celer L30e* also demonstrated most of the charged residues more or less stabilize the protein[54]. However, the effect of removing charge-charge interaction by mutagenesis on protein stability is rather unpredictable[43,47,51,196]. It is because electrostatic interaction can be either attractive or repulsive, not to mention the large desolvation and entropic penalty of their formation as discussed in chapter 1. More importantly, unlike hydrogen bonds of which the formation is in a one-to-one and close-proximity manner, a single charge can form electrostatic interaction between more than one charge which separated in a relatively long distance[55]. Therefore, the overall charge distribution should be taken into account when estimating the stabilizing effect of a charge.[66,75]. The optimization of charge distribution is especially important to

thermophilic proteins, as the usage of amino acid of thermophilic proteins is bias to the charged residues[56,97].

Unlike double mutant cycle which focuses on the interaction of a chosen ion pair, pK_a shift approach provides us a good way to investigate the complicated electrostatic interactions of a charge to the rest of the protein (see Chapter 1). Although pK_a values can be estimated by mutagenesis[76,78-80], finding chemical shifts of a charge at different pH by NMR is the preferred approach, because of its extremely high resolution and sensitivity to titrational event of charges. In fact, interactions of charged residues have been widely studied by investigating their pK_a values and titrational behavior[81-83,114,197-200]. One difficulty for analyzing pK_a values comes from the non-ideal Henderson-Hasselbalch (HH) behavior of the chemical shift titration curves obtained from NMR experiments, which may be due to strongly coupled electrostatic interactions for charges with comparable pK_a values, or changes in local chemical environment that are unrelated to protonation of the target charges such as structural rearrangement. Usually, non-HH titrational behavior is expected for charged residues for buried charges in active sites[113,114] because of the high charge density due to the low dielectric constant environment and the close proximity of charged residues in active sites. In another words, one interpretation for many charges to have non-HH titration curves is the presence of complicated charge-charge

interactions networks among them. Different models have been developed to describe non-HH titration curves, such as microscopic model[111], global fitting of titrational event (GloFTE) based on pH dependent protonation state populations[114], the decoupled sites representation[200]. Each model has its own advantages and weaknesses. Therefore, careful analysis is needed for getting information from the non-HH titration curves.

Being a thermophilic protein, *T. celer L30e* also has increased number of surface charges and charge clusters when compared to its mesophilic yeast homolog. In chapter 3, we have showed the pair-wise salt bridge interaction does stabilize the protein. However, the stabilizing role of their global electrostatic interaction of these surface charges remains unclear because of the possible complicated repulsion among them. To this ends, the quintuple Arg-to-Lys variant of *T. celer L30e*, will be used as the protein model to address this question by using pK_a shift approach. The Arg-to-Lys variant has its structural, thermodynamic, and electrostatic properties virtually identical to the wild type protein, but a largely elevated solubility in various low ionic strength conditions which were required in NMR experimental conditions (see Chapter 5). We have demonstrated that the high surface charge density would result in complicated repulsive electrostatic interactions among surface charges, which is indicated by non-HH titration curves for most of the Asp and Glu. With

careful analysis of microscopic fitting and GloFTE of non-HH titration curves, a set of reliable and largely down-shifted pK_a values of various acidic residues were obtained, showing that the protein had an optimized surface charge pattern for stabilizing the protein. To gain the knowledge of how temperature affects the contribution to protein stability of global electrostatic interaction, the electrostatic contributions were investigated at both low (298 K) and high (333 K) temperatures in this study.

6.2 Results

6.2.1 Arg-to-Lys variant of *T. celer L30e* has a very high surface charge density

T. celer L30e is a 100-residue thremophilic protein with very high surface charge density. It has an overall balanced charge of +1 at neutral pH, having 13 acidic residues and 14 basic residues unevenly distributed on the protein surface[190].

Based on the crystal structure of quintuple Arg-to-Lys variant of *T. celer L30e*, the undefined region at the flexible C-terminal was rebuilt by MODELLER. Using 8.0 Å as the cutoff distance, 2 large charge clusters and a surface of scattering isolated charges were identified (Figure 6.1). The first charge cluster is located in the region near N- and C-terminals, consisting N-terminal, Asp-2, Glu-6, Lys-8, Lys-9, Asp-12, Glu-90, Arg-92, Lys-99, Glu-100, and C-terminal (Figure 6.1, 0° view). The

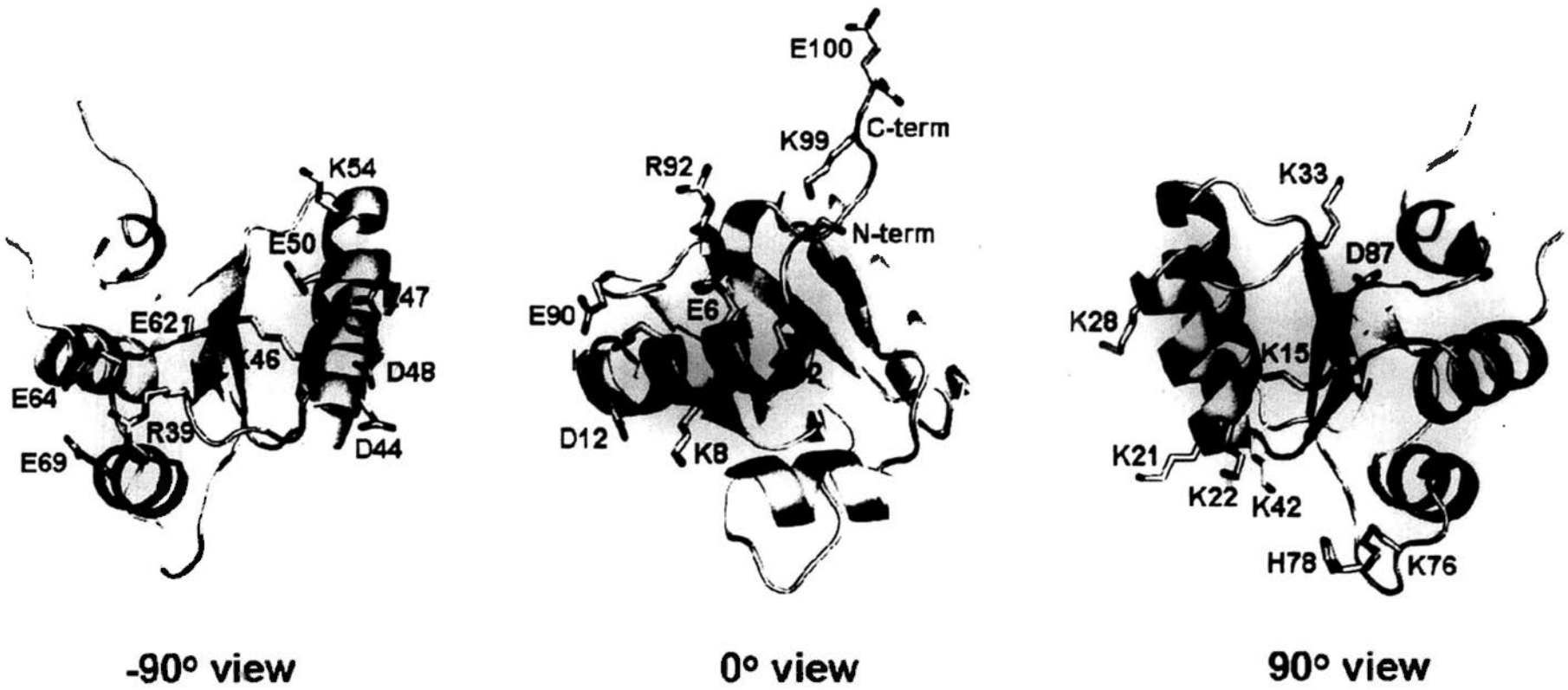


Figure 6.1. Surface charge distribution of the quintuple Arg-to-Lys variants. The side chain of charged residues (Asp, Glu, Arg, Lys, His) as well as the N- and C-terminals were indicated by one-letter-code and residue number, and shown as sticks presentation in -90°, 0° and 90° views. The charged oxygen atoms were colored red, and the charged nitrgen atoms were colored blue.

second charge cluster involves Arg-39, Asp-44, Lys-46, Glu-47, Asp-48, Glu-50, Lys-54, Glu-62, Glu-64, and Glu-69 in helix-3, strand-3, and helix-4 (Figure 6.1, -90o view). The surface of scattering isolated charges (Lys-15, Lys-21, Lys-22, Lys-28, Lys-33, Lys-42, Lys-76, His-78, and Asp-87) is basically the RNA binding site of *T. celer L30e* (Figure 6.1, 90o view)[190].

6.2.2 Backbone sequential assignments of native Arg-to-Lys variant at 298 K and 333 K

Backbone assignments of doubly labeled native Arg-to-Lys variant at pH 6.5 at 298 K and 333 K were obtained by using triple-resonance experiments. The backbone assignments were mainly relied on HNCACB[149]. The intensities of $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ peaks of HNCACB have 180° phase shifted, allowing us to distinguish them unambiguously. Therefore, the residue type of each peak on ^{15}N-HSQC plan was obtained based on the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shift values[150]. HNCACB also provided us with the chemical shifts of $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ of i – 1 residue, of which the peaks of i – 1 residue can be distinguished from that of i residue by having a much smaller intensity. Thus, the residue type of the i – 1 residue was also obtained. Spin systems could then be linked by these sequential connectivities (Figure 6.2). More than 90% of residues could be sequentially assigned without ambiguity. The

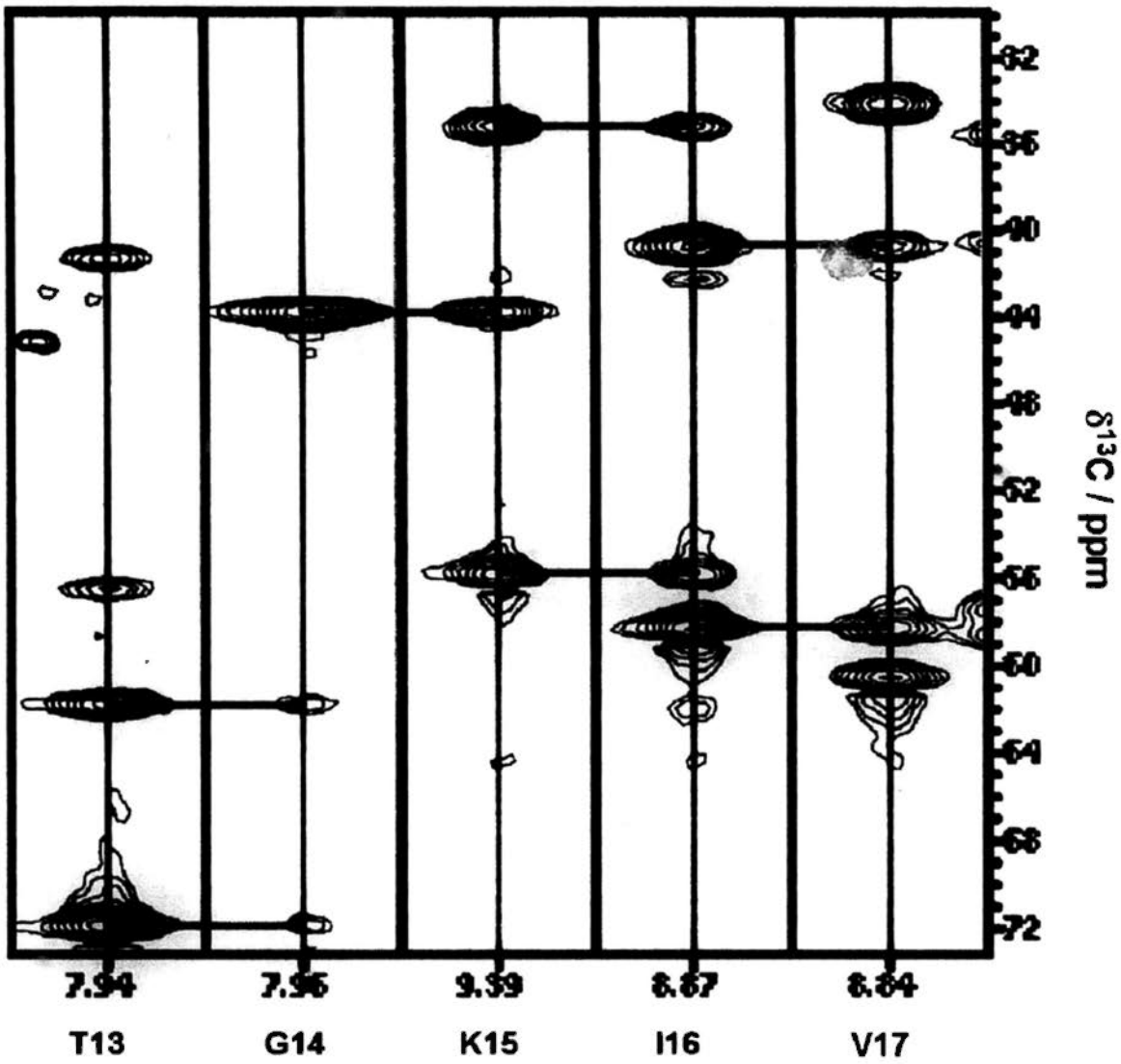


Figure 6.2. Sequential connectivities of 3D HNCACB. Selected pairs of $^{1}H^{13}C$ strips from the 3D HNCACB spectra of unifomly $^{15}N^{13}C$-labelled quintuple Arg-to-Lys variant in low ionic strength condition (10 mM citrate/phosphate buffer) at pH 6.5 at 298 K, which illustrate main-chain sequential connectivities for residues Thr-13 to Val-17. Sequential connectivities are indicated by horizontal lines between strips. Black and red peaks represent peaks with positive and negative amplitude.

ambiguous backbone assignments were confirmed by a pair of complementary triple-resonance experiments, namely ^{15}N-TOCSY-HSQC[151] and HC(C)(CO)NH-TOCSY[152], where the aliphatic side chain ^{1}H chemical shifts of i and i – 1 were correlated to that of backbone amide respectively. This pair of experiments provided additional connectivities for backbone sequential assignment and all backbone resonances were unambiguous (Figure 6.3). In addition, solution structure of wild-type *T. celer L30e* was solved in previous study[191]. The high similarity of 2D 15N-HSQC for wild-type *T. celer L30e* and Arg-to-Lys variant provided us with independent proof of the sequential assignment.

6.2.3 Side-chain resonance assignments for Asp and Glu of native Arg-to-Lys variant at 298 and 333 K

Assignments of side chain ^{1}H and ^{13}C resonances of Arg-to-Lys variant at 298 K and 333 K were obtained based on triple-resonance experiments of ^{15}N-TOCSY-HSQC[151], HC(C)(CO)NH-TOCSY and (HC)C(CO)NH-TOCSY[152], as well as HCCH-TOCSY[201].

The assignment of side chain ^{1}H resonances was straight forward. Using ^{15}N-TOCSY-HSQC only could already assign all aliphatic side-chain ^{1}H resonances of Asp and Glu. The assignment was double confirmed by HC(C)(CO)NH-TOCSY spectrum. For the side chain ^{13}C resonance assignment, the signal intensity of

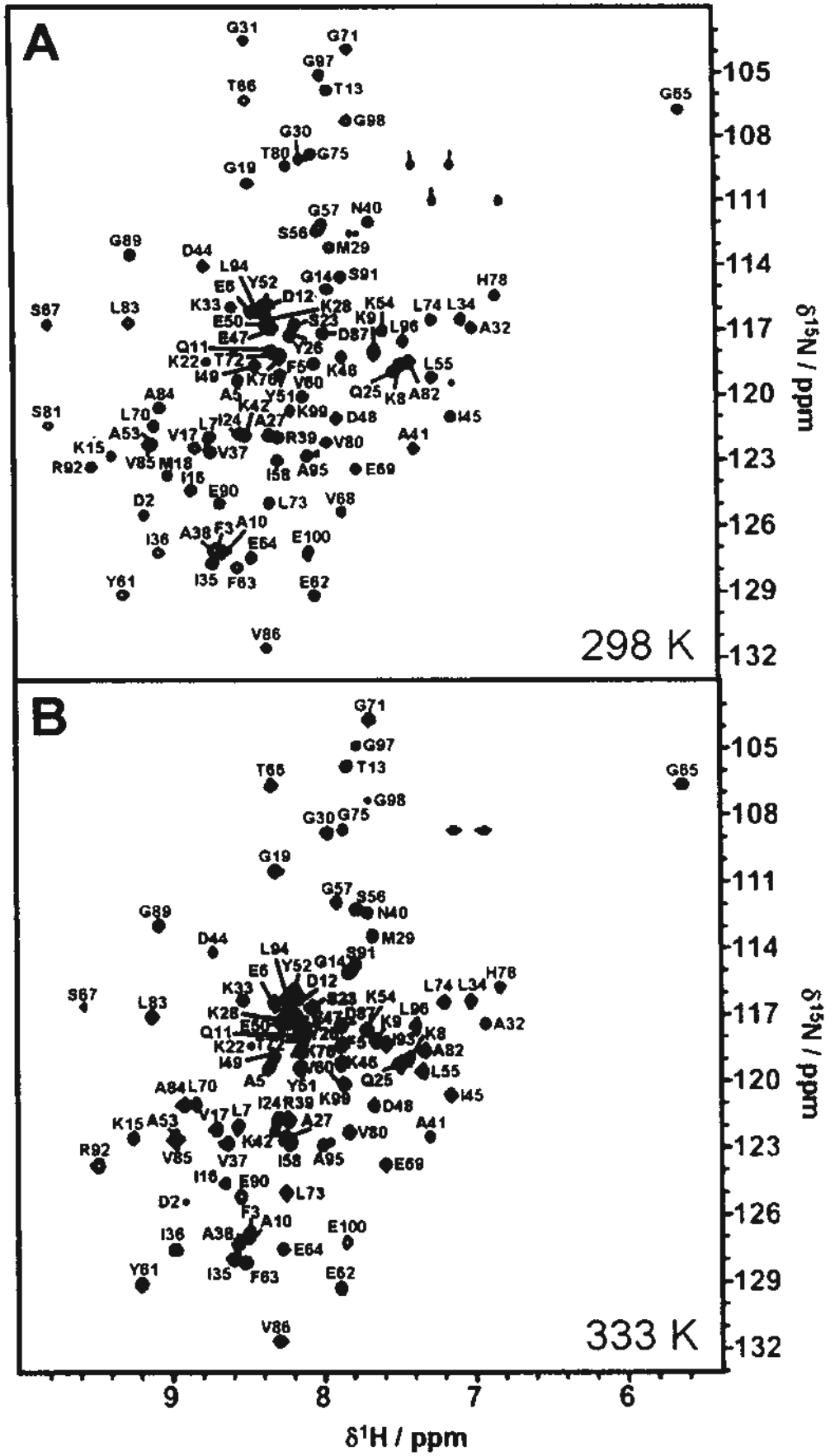


Figure 6.3. Backbone assignment of the quintuple variant. $^{1}H^{15}N$-HSQC spectra of the quintuple Arg-to-Lys variant at (A) 298 K and (B) 333K at pH 6.5. Assignments of 96 backbone amide signals are denoted by the one-letter code and the residue number.

(HC)C(CO)NH-TOCSY spectrum obtained were weak, and only the resonances of $^{13}C^{\alpha}$ showed relatively clear peaks in the spectrum. These problems limited the reliability of using (HC)C(CO)NH-TOCSY on determining the side chain carbon chemical shifts. These problems could be overcome by optimizing the experimental condition. Obviously, the week signal to noise ratio can be improved by increasing the number of scans for each increment (NS = 4 for the obtained spectrum). The length of the mixing time (12 ms for the obtained spectrum) should also be optimized to prevent any disappearance of side chain signal. By carrying out a series of experiments with successively longer mixing times ranging, for example, from 12 ms to 120 ms, the connectivity pathways through the whole spin system can be traced out, and the mixing time can be optimized. Nevertheless, with the help of powerful HCCH-TOCSY, all $^{13}C^{\beta}$ and $^{13}C^{\gamma}$ resonances of Asp and Glu respectively could be assigned unambiguously.

Having assigned $^{1}H^{\beta}$ and $^{1}H^{\gamma}$ resonances of Asp and Glu respectively, carboxyl carbon resonances of Asp and Glu could also be assigned by a modified 2D H(CA)CO[202], which optimized the one-bond correlation of carboxyl carbon and CH_2 group (by setting parameter d21 = 9 ms in Bruker standard pulse sequence HCACOGP3D). At 298 K, Asp and Glu peaks in H(CA)CO spectrum had a good dispersion and did not overlap with Asn, Gln, and Gly peaks at pH 6.5. Therefore,

$^{13}C^{\delta}$ and $^{13}C^{\varepsilon}$ resonances of all Asp and Glu at pH ranging from 1.0 to 6.5 at 298 K were assigned unambiguously (Figure 6.4). The assignment at pH 6.5 at 298 K was double confirmed by a 3D HACACO, which additionally correlated resonances of one-bond carbon to that of the carboxyl carbon. As the pH decreased from 6.5 to 1.0, $^{13}C^{\delta}$/$^{13}C^{\varepsilon}$ resonances of Asp/Glu were shifted upfield (Figure 6.5). The peaks became less disperse at lower pH, but all the resonances could still be assigned unambiguously. Again, the assignment at low pH 2.5 was double confirmed by a 3D HACACO. Similarly, $^{13}C^{\delta}$ and $^{13}C^{\varepsilon}$ resonances of all Asp and Glu at pH ranging from 1.0 to 6.5 at 333 K were also assigned by 2D H(CA)CO (Figure 6.4).

6.2.4 pH titration of Asp and Glu of native Arg-to-Lys variant at 298 K and 333 K.

The side chain carboxyl carbon chemical shifts of Asp and Glu at both 298 K and 333 K were collected and plotted as a function of pH (Figure 6.6). The changes of chemical shift of Asp and Glu from pH 1.0 to pH 6.5 were about 2 to 5 ppm, which were similar to the reported values in other researches[203-206]. At 298 K, the pH dependency of chemical shift of some acidic residues such as Glu-47, Asp-48, Glu-50, and Glu-62 has more than one obvious transition in pH range of 1.0 to 6.5, indicating those pH dependency of chemical shift cannot be explained by classical Henderson-Hasselbalch (HH) equation. The pH dependencies of chemical shift at

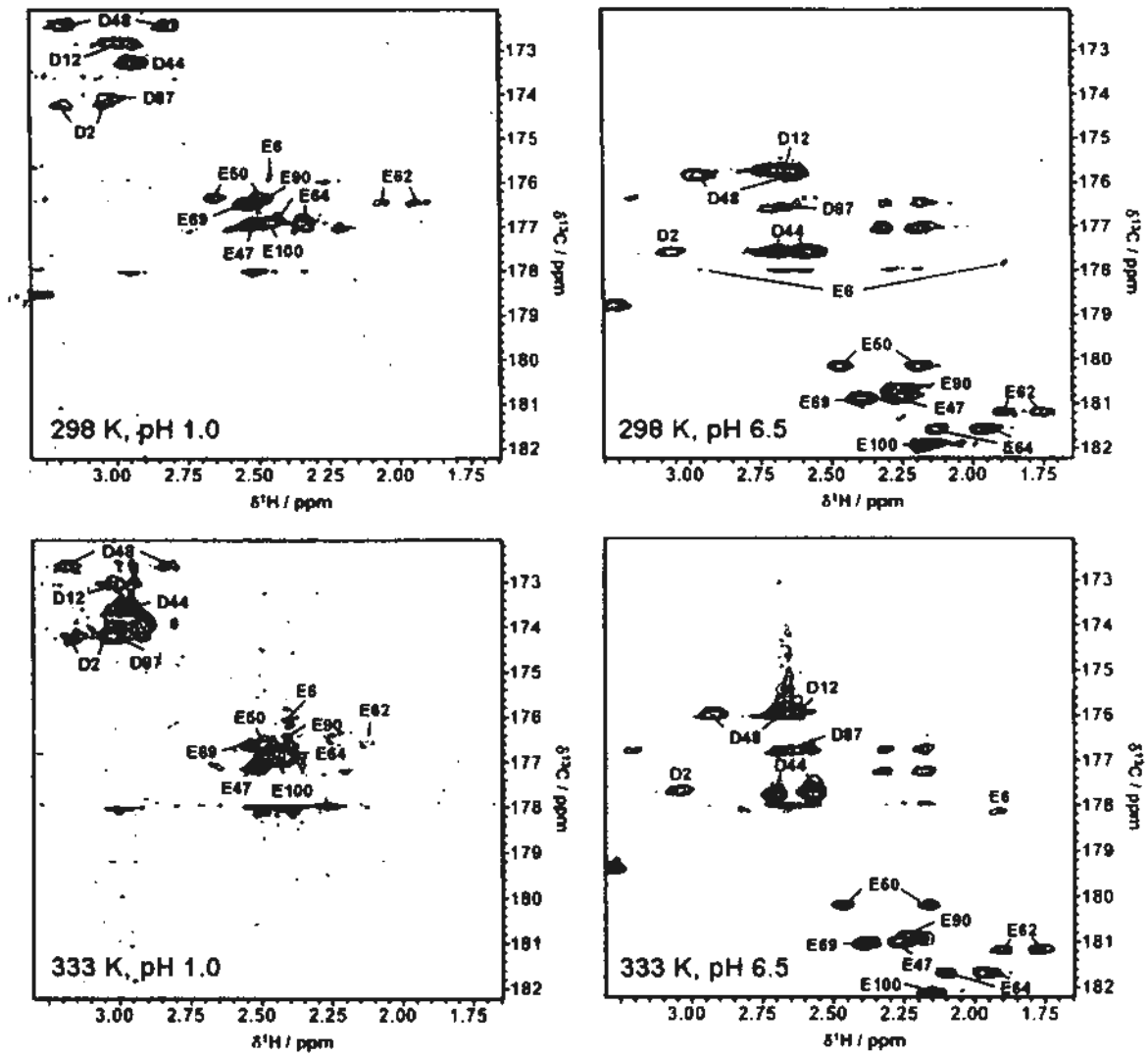


Figure 6.4. Side-chain assignment of acidic residues. Modified 2D $^{1}H^{13}C$ H(CA)CO spectra of the quintuple Arg-to-Lys variant of *T. celer L30e* at (A) pH 1.0 at 298 K, (B) pH 6.5 at 298 K, (C) pH 1.0 at 333 K, and (D) pH 6.5 at 333 K. The region shown contains the cross peaks for $^{1}H^{\beta}(^{13}C^{\beta})^{13}C^{\gamma}$ and $^{1}Hg(^{13}C^{\gamma})^{13}C^{\delta}$ connectivities of all Asx and Glx residues in the protein. Peak assignments for the side chain carboxyl groups of all Asp and Glu are denoted by one-letter code and residue number.

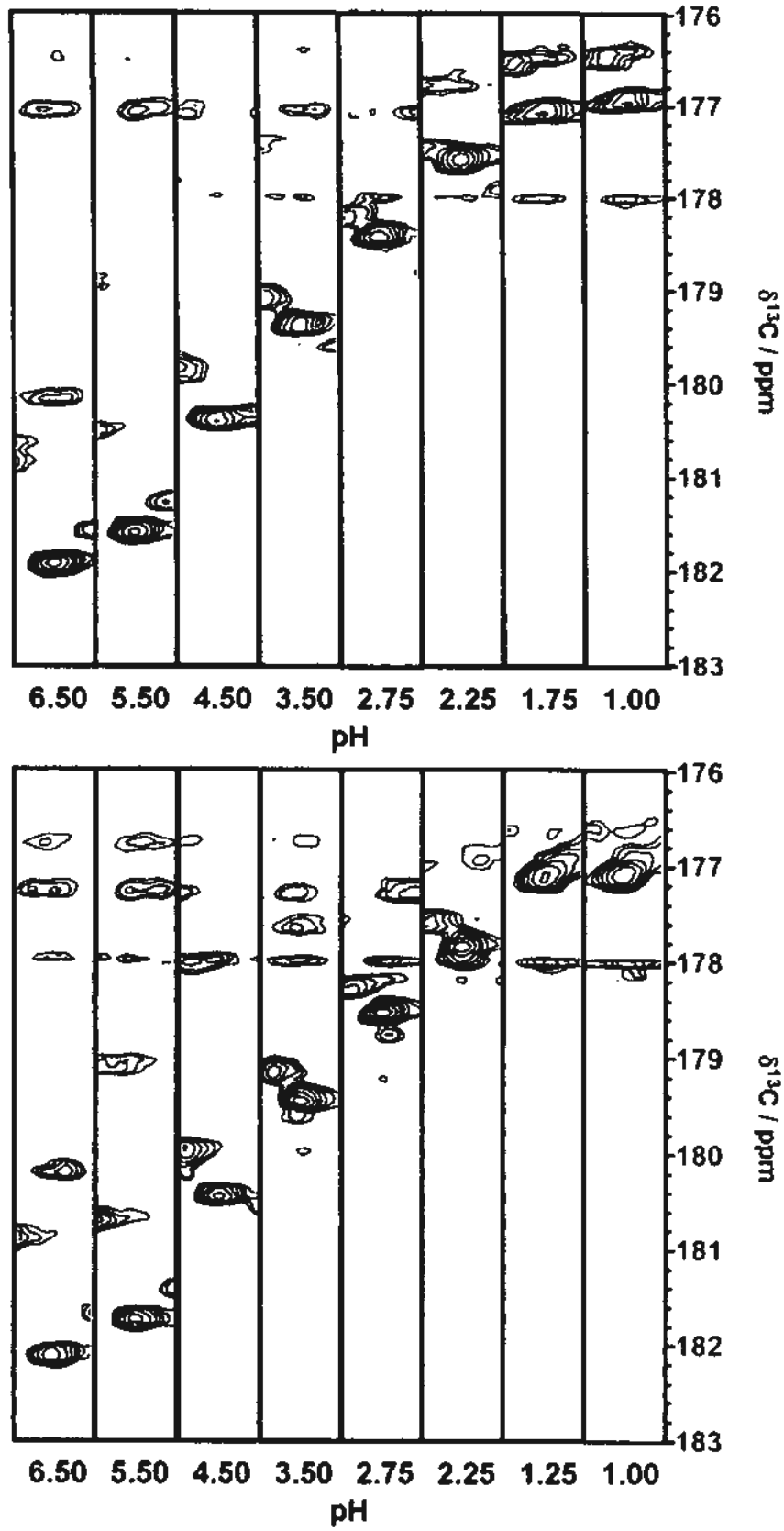
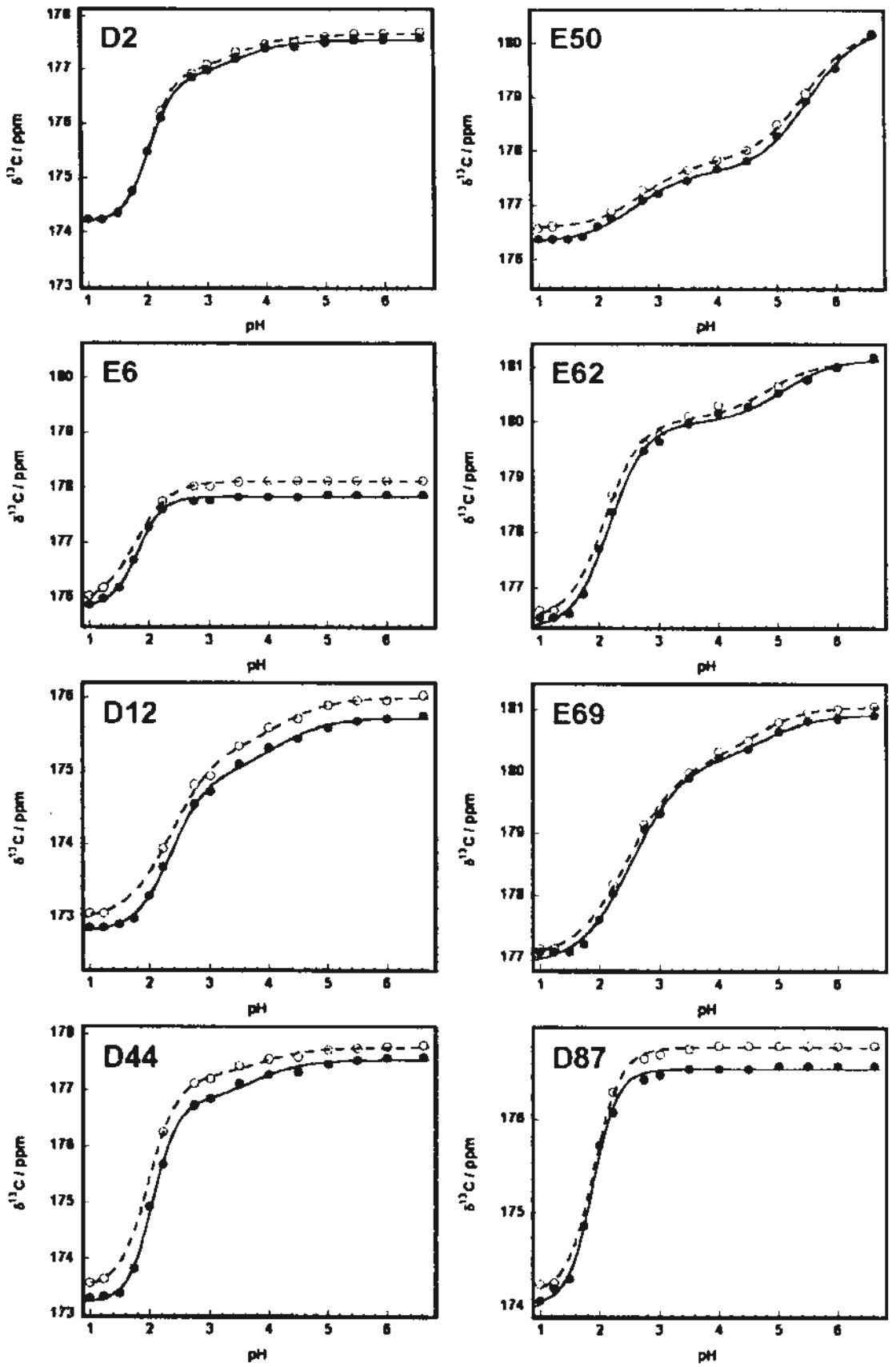


Figure 6.5. pH dependency of Glu-100 $\delta^{13}C^{\delta}$. pH titration of Glu-100 of the quintuple Arg-to-Lys variant at (A) 298 K and (B) 333 K. The cross peaks of Glu-100 of modified 2D $^{1}H^{13}C$ H(CA)CO were centered in ^{1}H dimension in all 0.1-ppm-width $^{1}H^{13}C$ strips.

A



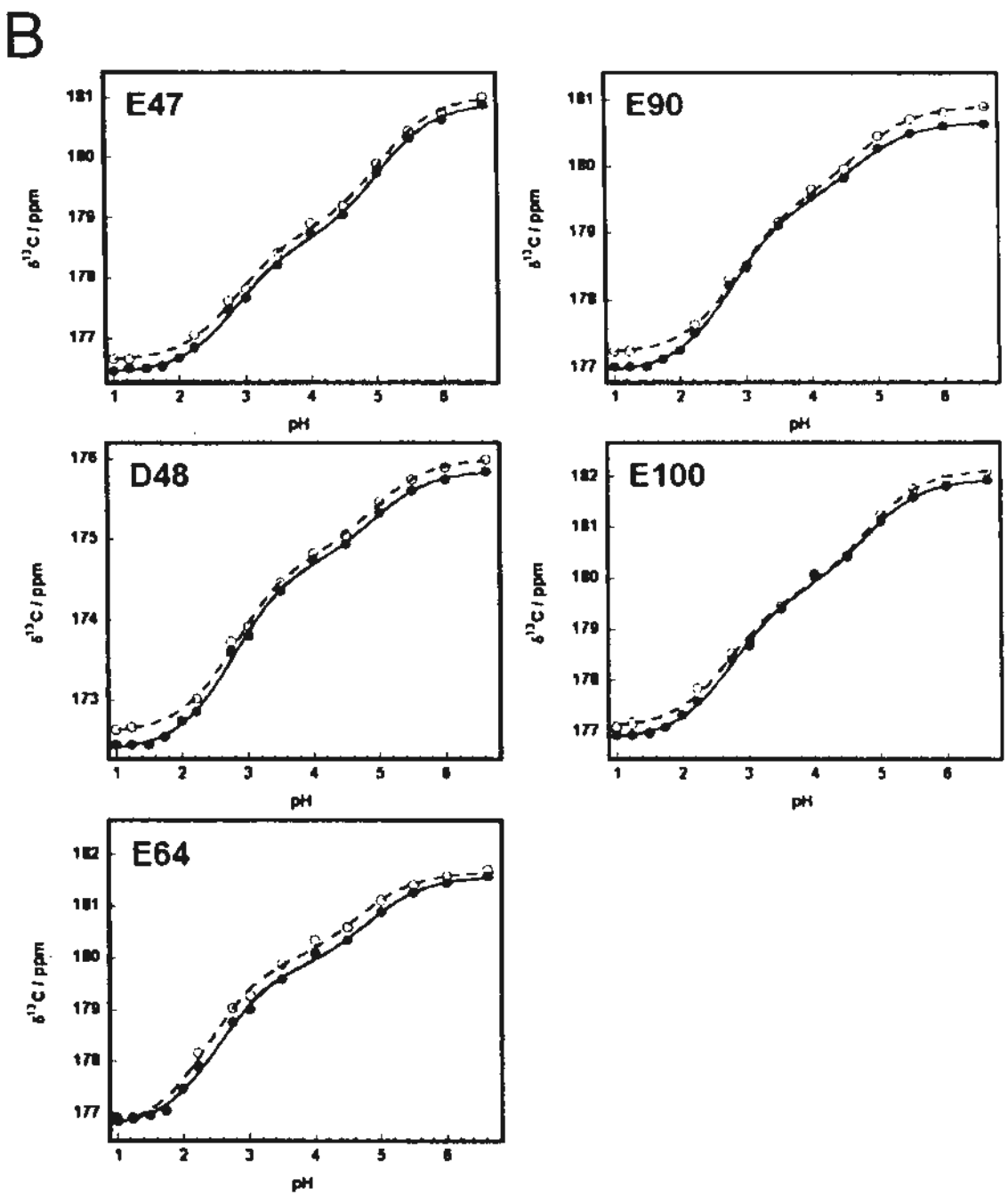


Figure 6.6. Microscopic pK_a fitting of carboxyl chemical shifts. Microscopic pK_a fitting of (A) residues that have reliable pK_a values and (B) residues that have no reliable pK_a values at 298 K (filled circle, continuous line) and at 333 K (open circle, dashed line). The transitions fitted for a curve were assumed independent to each other. The choice of number of transition fitted was validated by F-test and AIC (see Chapter 2).

333 K were apparently identical to those at 298 K, except having an upfield shifted chemical shift by about 0.2 – 0.3 ppm for every measured pH.

6.2.5 Fitting microscopic pK_a values of Asp and Glu of native Arg-to-Lys variant

pK_a values of Asp and Glu were estimated by microscopic pK_a model using modified Henderson-Hasselbalch equations, assuming all transitions of a residue are independent. To determine the number of transition for a residue, the pH dependencies of the residues were fitted from at least 1 transition up to 4 transitions. According to F-test statistics and AIC values, $\delta^{13}CO$ of Glu-6 and Asp-87 were best fitted to the standard one-transition HH equation with Hill's coefficient of about 2, whereas the titration curves of the others had two observable transitions (Figure 6.6). As reflected in the small χ^2 values in fitting (0.01 – 0.02), all pH dependencies of chemical shift were fitted very well to the microscopic pK_a model, suggesting all the observable transitions were well described by the fitted titration curves, and the obtained pK_a values should be accurate. The fitted pK_a values, and their absolute and percentage contributions were summarized in Table 6.1.

Although the assumption that all transitions of a residues are independent takes an advantage of being inclusive of every possible transitions in the titration curves, making data quality be the only limiting factor for the reflection of real pK_a values,

Table 6.1. pK_a values and changes of carboxyl ^{13}C chemical shift of acidic residues

Residue	$pK_{a(major)}$	$pK_{a(minor)}$	C_{major} (%)[b]	$\Delta\delta^{13}C_{total}$ (ppm)	$\Delta\delta^{13}C_{major}$ (ppm)[c]
D2*	2.0 (2.0)	3.5 (3.4)	74.5 (72.5)	3.4 (3.5)	2.5 (2.5)
D12*	2.4 (2.4)	4.1 (4.2)	70.6 (76.0)	2.9 (3.0)	2.0 (2.3)
D44*	2.0 (2.0)	3.7 (3.8)	79.8 (85.6)	4.3 (4.2)	3.4 (3.6)
D48	2.8 (2.8)	4.9 (4.9)	64.7 (63.1)	3.4 (3.3)	2.2 (2.1)
D87*	1.9 (1.9)		100 (100)	2.5 (2.6)	2.5 (2.6)
E6*	1.8 (1.8)		100 (100)	2.0 (2.1)	2.0 (2.1)
E47	5.0 (5.0)	2.9 (2.9)	50.4 (51.8)	4.4 (4.4)	2.2 (2.3)
E50*	5.5 (5.5) 5.1[a] (4.9[a])	2.5 (2.4)	70.0 (70.0)	3.8 (3.6)	2.7 (2.5)
E62*	2.2 (2.1) 2.4[a] (2.3[a])	5.2 (5.1)	74.8 (76.7)	4.7 (4.6)	3.5 (3.5)
E64	4.8 (4.7)	2.5 (2.4)	62.0 (63.0)	4.7 (4.9)	2.9 (3.1)
E69*	2.5 (2.5)	4.8 (4.6)	80.6 (76.4)	3.8 (3.9)	3.1 (3.0)
E90	4.7 (4.7)	2.8 (2.9)	66.0 (63.0)	3.6 (3.7)	2.3 (2.3)
E100	2.8 (2.7)	4.8 (4.8)	56.0 (53.0)	5.0 (5.0)	2.8 (2.7)

Residues which have reliable pK_a values (percentage contribtion ≥ 70 % and $\Delta\delta^{13}C_{major} \geq 2.0$) are indicated by asterisks. Values in parentheses are data obtained at 333 K, while other data are collected at 298 K.

[a] The pK_a values were obtained by GloFTE by pKaTool[113] where Glu-50 and Glu-62 were fitted together as an isolated charge group.

[b] The percentage contributions of the major transition were found by fitting the pH dependency of carboxyl ^{13}C chemical shift to microscopic pK_a model.

[c] The absolute change of chemical shift due to major transition ($\Delta\delta^{13}C_{major}$) was obtained by multiplying the total absolute change of chemical shift ($\Delta\delta^{13}C_{total}$) to the percentage contribution.

fitting in this manner suffers from difficult assignment of the transitions fitted. This is because it is impossible to distinguish whether a transition is due to titrational events directly, or due to non-titrational (but pH dependent) events, such as structural rearrangement or other effects that influence the sensitive chemical shifts but not the protonation/deprotonation state. Since the chemical shifts obtained in this study came directly from the titratable group (side-chain carbonxyl carbon of Asp and Glu), it is confident that the major transition of the titration curve is mainly due to the protonation/deprotonation of the corresponding residue.

Another problem for assigning transitions comes from the uniqueness of the fitted pK_a values. This problem is not only specific to microscopic pK_a model, but also associates with GloFTE. The uniqueness problem has two levels, namely fitting uniqueness and physical meaning uniqueness (Figure 6.7). Fitting uniqueness associates with the mathematical problem of determining if the fitted pK_a values are the only solution to describe the titration curve. Whereas physical meaning uniqueness refers to the problem of determining if the fitted pK_a values have a unique particular physical meaning. The physical meaning has ambiguity when it is an averaged value of more than one comparable pK_a values from different residues. It is not difficult to imagine different combination of pK_a values can be averaged into the same number. If a transition contributes largely to the change of chemical shift in

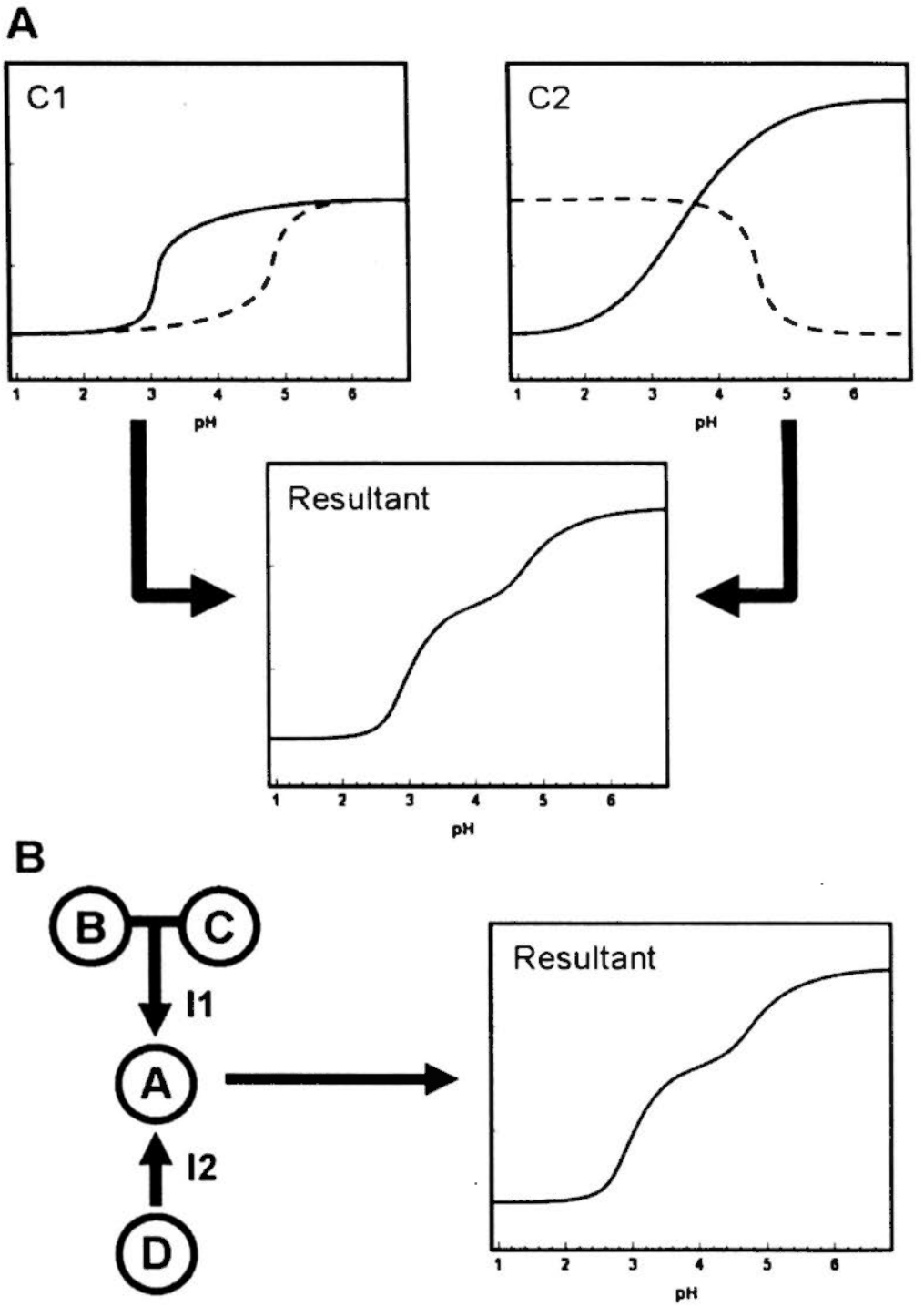


Figure 6.7. The fitting and physical meaning uniqueness problems. Schematic diagrams that illustrate (A) the fitting uniqueness problem and (B) the physical meaning uniqueness problems. For the fitting uniqueness problem, two different combinations (C1 and C2) of two transitions (continuous and dashed lines) can produce the same resultant non-HH titration curves. Be noted that, mathematically speaking, the fitting uniqueness problem is also valid for perfect HH titration curves. For the physical meaning uniqueness problem, the extra transition in the resultant titration curve of atom A can either due to the combinatory interacting effect of atom B and C (I1), or due to the interaction of atom D (I2).

terms of absolute value (≥ 2 ppm) as well as in terms of percentage (≥ 70%), the corresponding fitted values should be reliably reflecting the true pK_a value of the residues due to its titrational event, because a large change in chemical shift is highly resistant to changes and the fitting uniqueness problem will be diminished. For the physical meaning uniqueness problem, since the crystal structure of Arg-to-Lys variant was solved (see Chapter 5), the physical meaning of transition can be estimated from the structure unless there is any pH dependent structural change and rearrangement, or any non-titrational event that affects chemical shifts but not protonation/deprotonation, occurred.

Summing up the criteria, pK_a value fitted from the major transition which has contributes more than or equal to 2 ppm change and 70% of total change of chemical shift is robust enough to be safely regarded as the reliable pK_a values of the residue due to its titratable event. The titration curves with reliable pK_a values were summarized in (Figure 6.6A). The reason for the non-HH behavior of the titration curves (i.e. the physical meaning of the minor transitions) could also be estimated according to the crystal structure. However, since the values and physical meaning of these pK_a (fitted from small transitions) were not reliable as discussed, they would not be further analyzed and discussed.

6.2.6 Identification of coupled Asp and Glu by reciprocal relationship of microscopic pK_a values and global fitting of titrational events (GloFTE)

Careful analysis of microscopic pK_a values revealed that the major and minor transitions of Glu-50 and Glu-62 had reciprocal relationship (Figure 6.8). For both residues, the major transitions accounted for about 70 % of total chemical shift change, and the major/minor pK_a values for Glu-50 and Glu-62 were 5.5/2.5 and 2.2/5.2 respectively for 298 K, 5.5/2.4 and 2.1/5.1 respectively for 333 K. The high reciprocal correlation indicated Glu-50 and Glu-62 were strongly interacting. The interaction of Glu-50 and Glu-62 was validated by the crystal structures. It was found that the side chains of both Glu-50 and Glu-62 were well defined and the separation distance was 6.7 Å. This short separation distance between Glu-50 and Glu-62 allowed these two residues experienced repulsive forces from each other.

The interaction between Glu-50 and Glu-62 was also shown in the result of global fitting of titrational events (GloFTE), of which the fitted titration curves can describe the non-HH pH dependency of chemical shifts due to electrostatic coupling of two charges. If a pair of charges which have comparable pK_a values has strong electrostatic interaction among them, it will be expected that using GloFTE should result in reasonable good fitting for both curves. It was found that the normalized

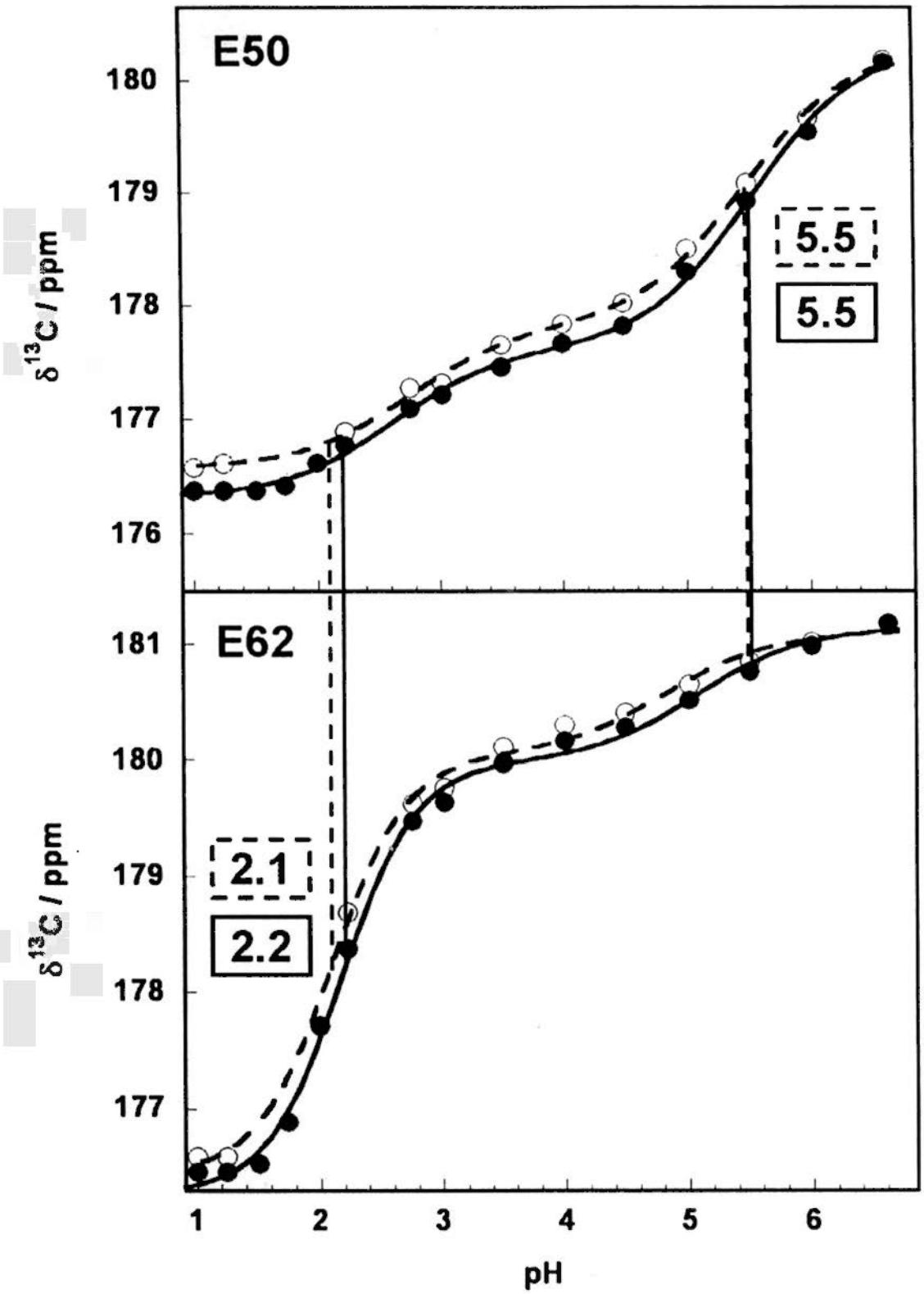


Figure 6.8 The reciprocal relationship of major and minor transitions. The strong reciprocal relationship between the major and minor transitions of Glu-50 and Glu-62 at 298 K (filled circle, continuous lines) and at 333 K (open circle, dashed lines). The fitted major microscopic pK_a values of each residue at 298 K (in solid box) and at 333 K (in dashed box) were also shown.

chemical shifts of Glu-50 and Glu-62 could be well described by GloFTE at both 298 K and 333 K (Figure 6.9). The fitting result showed Glu-50 has a very strong interaction with Glu-62 of 12.6 kJ mol^{-1} at 298 K and 11.4 kJ mol^{-1} at 333 K. The pK_a values of Glu-50 and Glu-62 at both 298 K and 333 K obtained from GloFTE, would be used for further analysis (Table 6.1).

6.2.7 pK_a values of Asp and Glu are predicted in native and denatured states of Arg-to-Lys variant

Using the modeled crystal structure of Arg-to-Lys variant, pK_a values of Asp and Glu were predicted by three different online servers, namely PROPKA[58,59], H++[117,118], and KARLSBERG+[119,120]. The reliable microscopic pK_a values obtained from NMR experiments were compared with the predicted values (Figure 6.10). It was found that H++ and KARLSBERG+ had reasonably good predictions on pK_a values, having correlation with experimental pK_a values at 298 K / 333 K of 0.80 / 0.80 and 0.79 / 0.78 respectively, although the H++ prediction has scaling error, with the slope of linear best fit line equal to 0.59 (1.00 for KARLSBERG+). On the other hand, PROPKA returned less satisfactory results. The predicted pK_a values by PROPKA have correlation with experimental pK_a values at 298 K / 333 K of 0.63 / 0.62. The small slope of linear best fit line (0.33) indicated PROPKA prediction on Arg-to-Lys also has serious scaling problem.

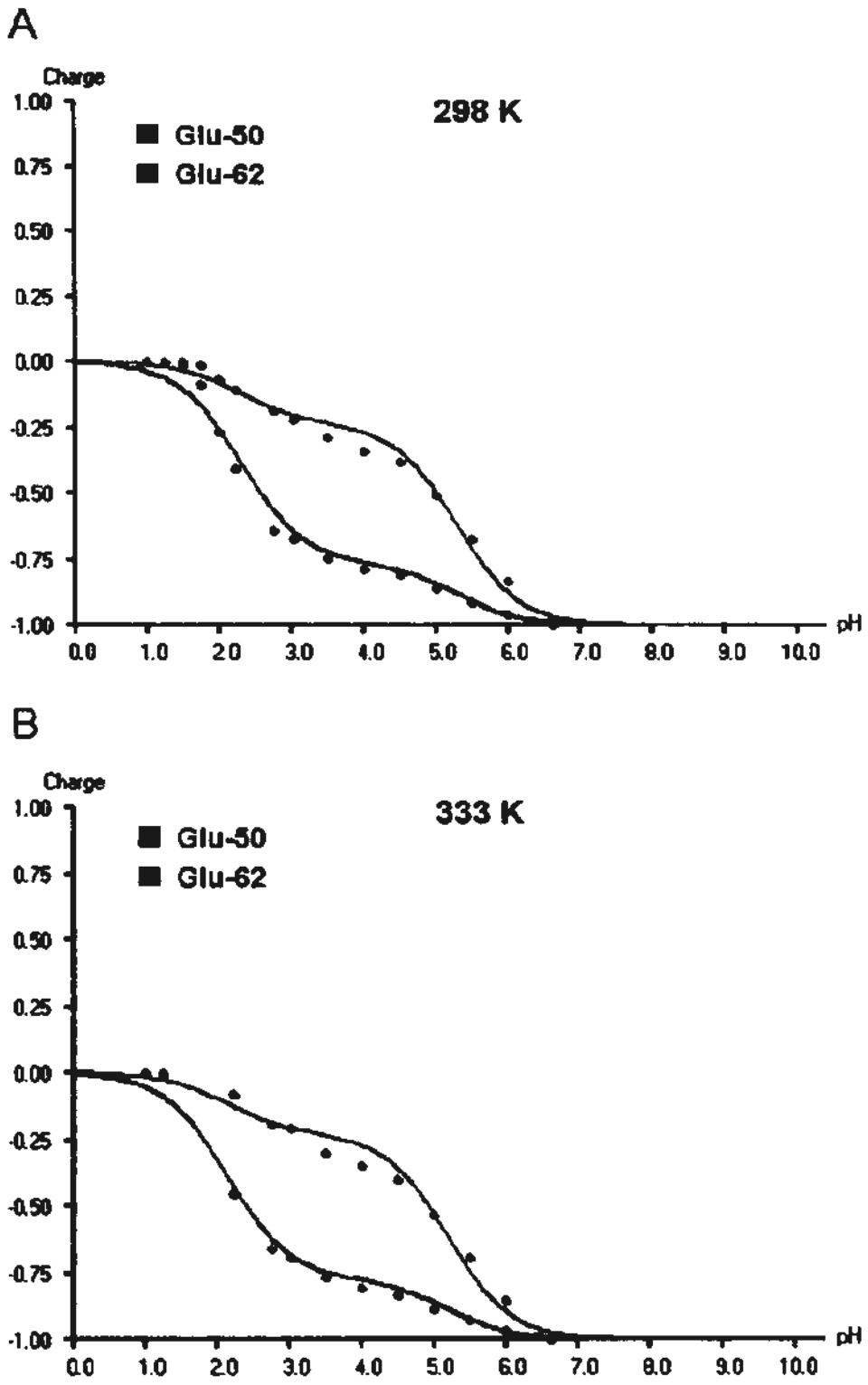


Figure 6.9. Globle fitting of titrational events (GloFTE). Globle fitting of titrational events (GloFTE) of Glu-50 (red) and Glu-62 (green) as an isolated system at (A) 298 K and (B) 333 K. The ^{13}C carboxyl chemical shifts for GloFTE were normalized as charge between 0 and -1 by pKaTool[113]. Be noted the reciprocal relationship between major and minor transitions of Glu-50 and Glu-62 was also captured in microscopic fitting.

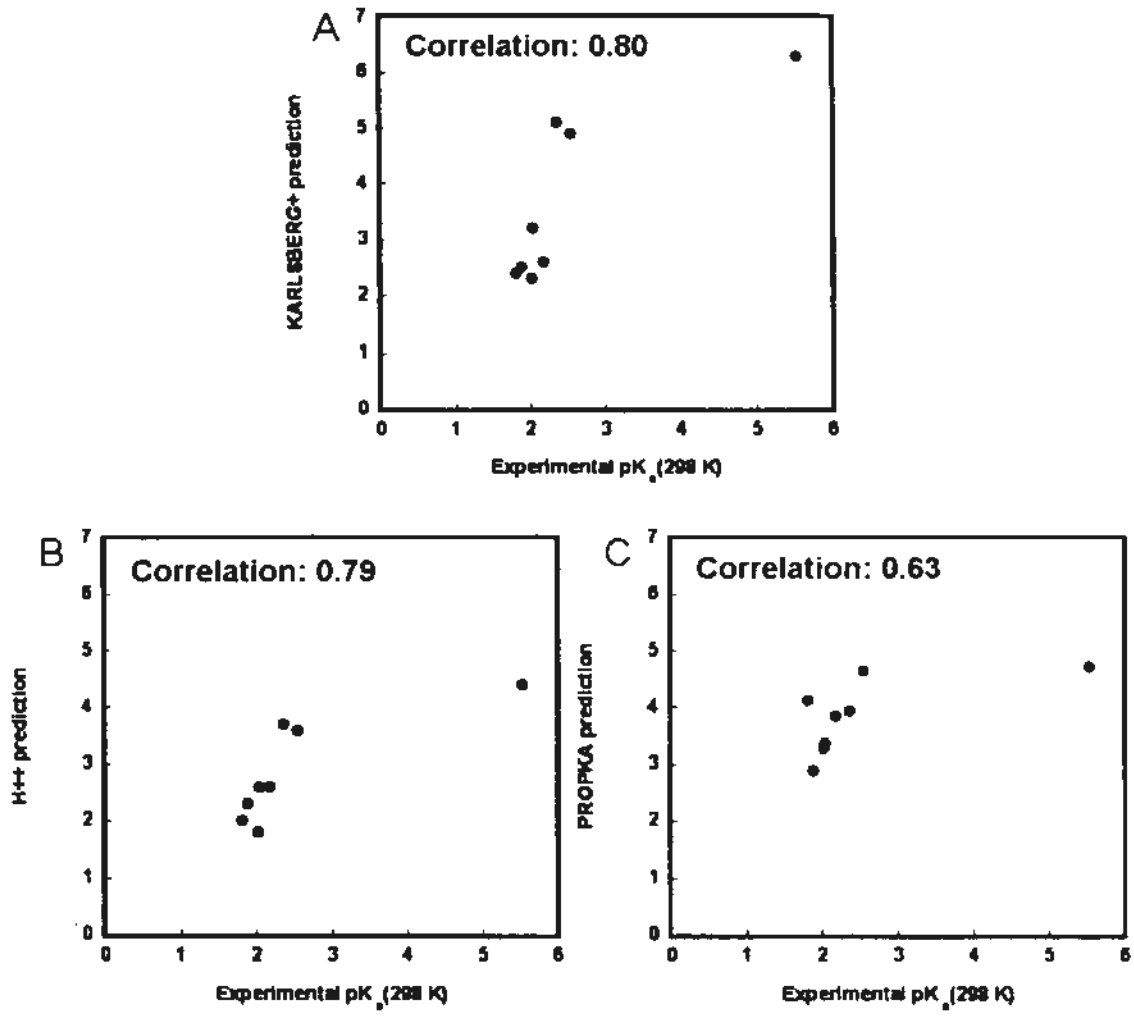


Figure 6.10. Correlation of experimental and predicted pK_a. Correlations of experimental pK_a obtained at 298 K and predicted pK_a predicted by online servers of (A) KARLSBERG+, (B) H++, and (C) PROPKA. The predictions were based on modeled quintuple Arg-to-Lys variant crystal structures that included the C-terminal flexible tail.

Since no residual structures are expected in denatured state of Arg-to-Lys variant (see Chapter 4), a polyproline II based denatured state model of Arg-to-Lys variant was built. pK_a values of all Asp and Glu were predicted by H++ online server, which gave good prediction when predicting pK_a values in native state (KARLSBERG+ prediction broke down since the submitted molecule was too large).

6.2.8 Random coil pK_a values of Asp and Glu in model peptides obtained by pH titration

To estimate the random coil pK_a values of Asp and Glu, side chain pK_a values in 5-residue model peptides Ac-GG(D/E)GG-NH_2 at 298 K and 333 K were obtained from classical pH titration (Figure 6.11). Poly-glycine-based model peptides have been widely used to prevent residual structure of peptide in random coil chemical shift studies[159-161]. The titratable residues were placed in the middle of the model peptides to prevent any terminal effect on its chemical shifts, which may in turn affect the estimation of the pK_a values of Asp and Glu, because the pH dependency of terminal effect is unknown. In this study, N- and C- terminal of the model peptides were protected by additional acetate and amide groups respectively, which mimic the real situation of being in the middle of polypeptide, and at the same time remove the two terminal charges which would certainly influence the estimation of target pK_a values. The removal of terminal charges provides a bonus advantage that it makes the

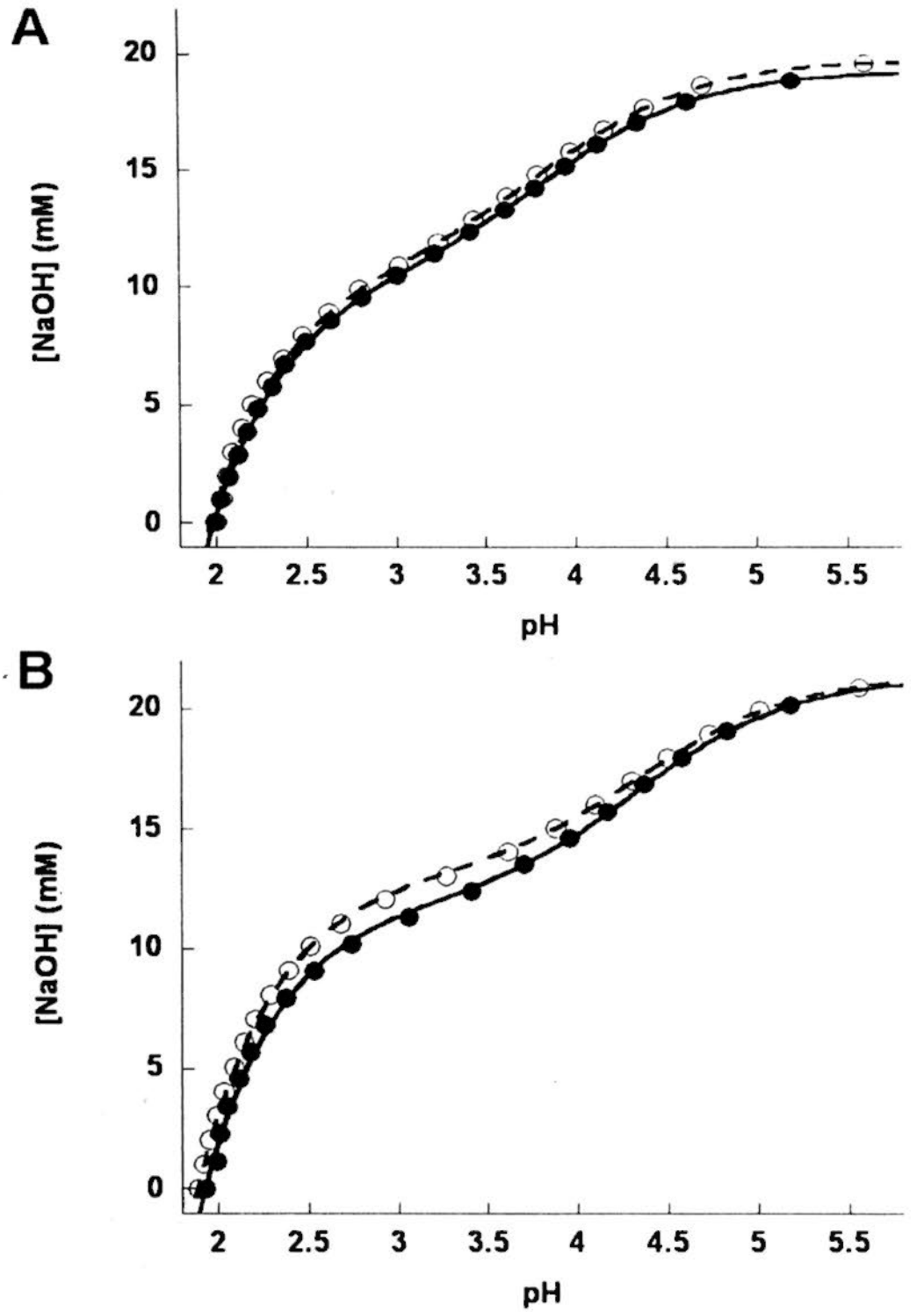


Figure 6.11. pH titration of model peptide. pH titration of terminal-protected 5-residue peptides of (A) Ac-GGDGG-HN_2 and (B) Ac-GGEGG-NH_2 at 298 K (filled circle, continuous line) and 333 K (open circle, dashed line). N-terminal and C-terminal of the peptides were protected by acetyl gropu and amide groups respectively. Accurately about 10 mM of peptides were titrated with commercially available standard 1.0 M NaOH.

side chain of Asp or Glu in the model peptide be the only titratable group. Although it has been reported recently that simple 2D NMR experiment is already good enough for estimating the pK_a values in disordered peptide fragment[207], peptide with only one titratable group enables us to find the target pK_a value accurately by a procedure-wise even more simple classical pH titration. The fitted pK_a values of Asp and Glu in the model peptide at 298 K were found to be 3.9 ± 0.0 and 4.3 ± 0.0 respectively, and to be 3.9 ± 0.0 and 4.4 ± 0.0 respectively at 333 K.

6.2.9 Surface charge contribution to protein stability of Arg-to-Lys variant.

The contribution to protein stability of a titrable group can be calculated if its pK_a values in native state and in denatured state are known. The fact that wild-type *T. celer L30e* has no observable residual structure (see Chapter 4) and virtually identical properties of Arg-to-Lys variant to wild-type *T. celer L30e* (see Chapter 5) evidently infers that Arg-to-Lys variant also has no observable residual structures in its denatured state. Based on this assumption, the free energy contributions of residues with reliable pK_a values at 298 K and 333 K were calculated using the random coil pK_a values obtained from model peptides as the pK_a values in denatured state (Figure 6.12).

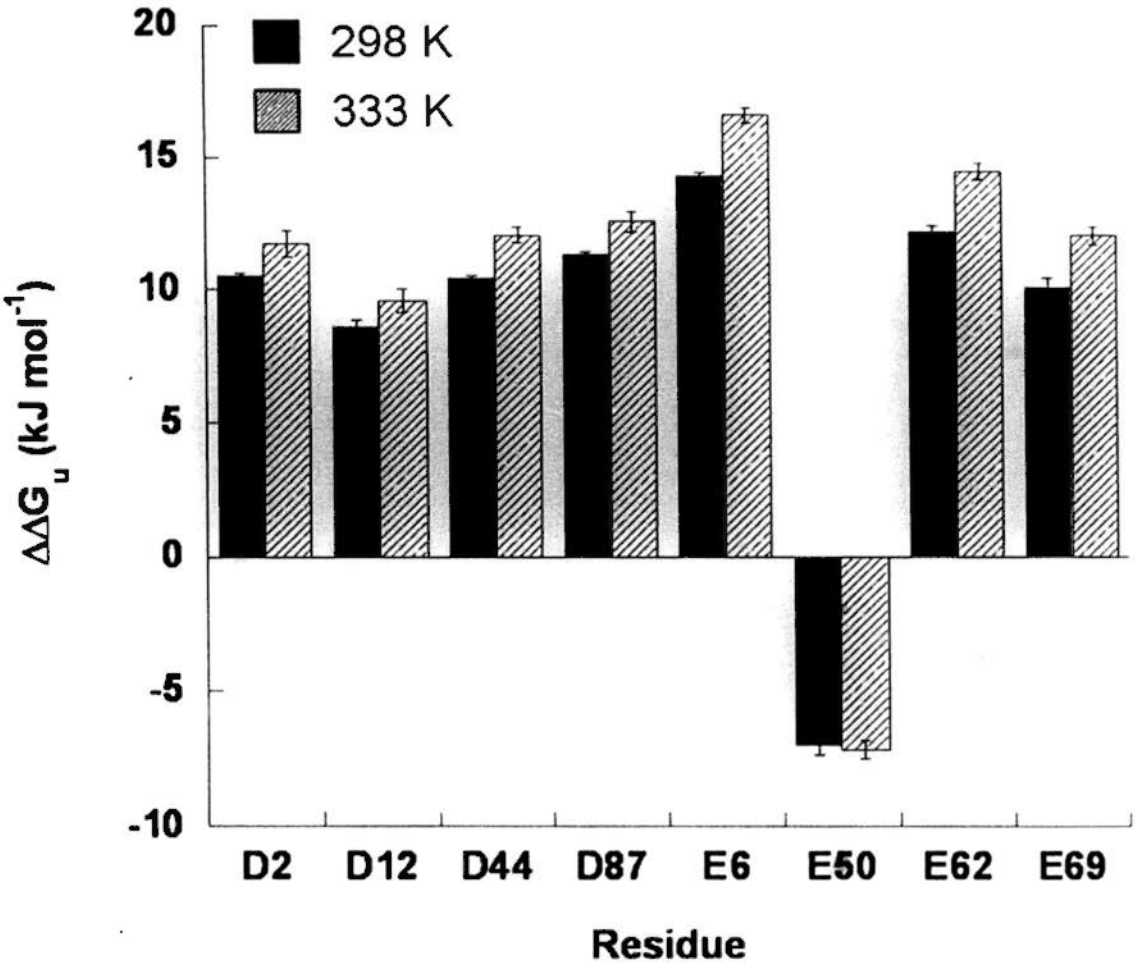


Figure 6.12. The electrostatic contribution of acidic residues. The electrostatic contribution to protein stability of acidic residues which have reliable pK_a values at 298 K (filled) and at 333 K (shaded). Assumptions that the quintuple Arg-to-Lys variant has no residual structures, and the pK_a values of Asp and Glu obtained from pH titration of peptides were equal to that of denatured state, were made in the calculation of the electrostatic contribution. For Glu-50 and Glu-62, the pK_a values were obtained by GloFTE, while the other pK_a values were obtained by microscopic pK_a fits.

6.3 Discussion

6.3.1 The electrostatic repulsion is commonly experienced by surface acidic residues in the quintuple Arg-to-Lys variant due to exceptionally high surface charge density

To investigate the titrational properties of charged residues, pH dependencies of side chain carboxyl chemical shifts of all 13 acidic residues were obtained from modified 2D H(CA)CO NMR experiments. 11 out of 13 titration curves exhibited more than one transitions, with Glu-6 and Asp-87 being the exceptions.

NMR is a powerful technique to determine pK_a values of residues in a protein by measuring chemical shifts of the titratable groups at various pH. Measuring chemical shifts has advantages that it provides atomic level resolution so that the titrational behavior for each atom can be traced and analyzed. In addition, chemical shift is very sensitive to the local chemical environment. The high sensitivity of chemical shift enables it to reflect any interactions of the corresponding group. However, being highly sensitive to local chemical environment can also complicate the analysis of chemical shifts. Titration curve of a residue can exhibit more than one transition if it is coupled to its nearby by titrating groups with similar pK_a values[111,208,209]. We cannot simply assume any non-HH titrational behavior is due to electrostatic coupling between charges (titrational event) in analysis of experimental

titration curves obtained from NMR experiment, because other pH-dependent non-titrational events such as protonation/deprotonation of nearby non-interacting charge and structural rearrangement, which affect the chemical shifts but not the protonation state of the target charged residue, may also contribute to the resultant titration curves. The effects (magnitudes and directions) of non-titrational events are difficult to be predicted unless detail structural information is available[199]. As illustrated in Figure 6.7, the understanding of the transitions in titration curves is even complicated by the "uniqueness" problems where the titration curve itself and the fitted pK_a values can be an average function or value, not to mention the intrinsic experimental error of measured chemical shifts. Therefore, before using the fitted pK_a values for further interpretation, it is essential to distinguish whether the non-HH titrational behavior of the titration curve is due to titraional event in order to determine if the non-HH behaved charges are experiencing electrostatic repulsion from nearby charges. For example, for a pair of nearby like charge, their non-HH behavior of their titration curves should be due to electrostatic interaction among them. Their protonation/deprotonation of the interacting charges are said to be coupled by titrational event, and their minor transitions were describing the protonation/deprotonation of their interacting partner. Thus, their major and minor transitions should show a highly reciprocal relationship to each other.

Based on the Boltzmann distribution-derived pH-dependent protonation state populations, titration curve of residues fitted by global fitting of titrational events (GloFTE) method can account for the effects of electrostatic interaction of nearby charges, but effect of any non-titratable events will be ignored in the fitting. Therefore, the fitting error of GloFTE should be small if the residues involved in the fitting were only electrostatically coupled[114]. The good fitting result of GloFTE for fitting Glu-50 and Glu-62 together (Figure 6.9) indicated that there was strong electrostatic repulsion between Glu-50 and Glu-62. The electrostatic repulsion between Glu-50 and Glu-62 was further supported by the microscopic fitting (Figure 6.8). Microscopic pK_a fitting fitted the transitions independently without any bias for matching the pK_a values of different residues. However, the pK_a values fitted from the major and minor transition of Glu-50 and Glu-62 still exhibited a high reciprocal relationship, indicating the protonation and deprotonation were coupled. According to the crystal structure, the separation distance between these two residues was 6.7 Å, which was short enough for experiencing repulsive electrostatic interaction from each other but too long for having other kinds of interaction such as van der Waals interactions. This structural information provided another independent evidence that the protonation and deprotonation of Glu-50 and Glu-62 were coupled due to electrostatic repulsion.

It is worthwhile to note that Glu-50 and Glu-62 were not the only residues which experienced repulsive electrostatic interactions. 11 out of 13 the acidic residues, except Glu-6 and Asp-87, had two observable transitions in microscopic pK_a fittings. One more example for strong electrostatic interaction between like charges pair was the case of Glu-47 and Asp-48. Similar to the case of Glu-50 and Glu-62, Glu-47 and Asp-48 were located near to each other with separation distance of 5.2 Å. Also, they had good fitting result in GLoFTE, and their two pK_a values were reciprocal to each other, although they have no major transition for fitting reliable pK_a values. For the other residues, interacting partners were difficult to assign. One reason is that the minor transition is too small to fit a robust pK_a value for analysis by both GloFTE and microscopic fitting, which can likely explain the cases of Asp-2, Asp-12, Asp-44, and Glu-69 (Figure 6B). However, their minor transition should due to long range electrostatic interaction with other like charges as all of them are highly exposed and are relatively isolated with respect to acidic residues (so that only long range interactions could take place). Another reason is that no titration data of interacting group obtained, which is the case of Glu-100. Being the C-terminal residue, side chain of Glu-100 certainly has electrostatic interaction with backbone C-terminal. However, no titration data of C-terminal could be obtained in the modified 2D H(CA)CO. As Glu-100 is located in a flexible tail of

the protein, the pK_a values of Glu-100 are best reference to those of free Glu amino acid (4.3 for side chain, 2.2 for "backbone" carboxyl group). In this sense, the pK_a values of the side chain and backbone carboxyl groups of Glu-100 at 298 K / 333 K should be 4.8 ± 0.1 / 4.8 ± 0.1 and 2.8 ± 0.1 / 2.7 ± 0.1 respectively (Figure 6B). The cases of Glu-64 and Glu-90 were not straight forward. Obviously they have two transitions in their titration curves (Figure 6B), but no nearby (< 8.0 Å) charges in the crystal structure could explain their non-HH behavior. Long range electrostatic interaction, or even non-titrational event such as structural rearrangement (preferred rotamer or conformation of the side chains) may be the reason for their non-HH titration curves, although further experiments have to be done to prove this hypothesis.

Nevertheless, we can see that the coupling of protonation and deprotonation of surface charges of the Arg-to-Lys variant was very common, where the coupling was likely due to the global electrostatic interaction among them because of the unusual high surface charge density of the protein.

It is well documented thermophilic proteins have much more surface charges when compared with their mesophilic homologues[56,92,93,96-98]. Being a thermophilic protein, Arg-to-Lys variant of *T. celer L30e* is no exception. Out of the 101 residues in the construct, 27 of them are charged residues distributed on the surface of the

protein (Figure 6.1). The charged atoms (side chain O or N) of the charged residues have already covered about one-forth of the total protein surface (1394.3 $Å^2$ out of 5745.8 $Å^2$). The unusually high surface charge density of the quintuple Arg-to-Lys variant evidently suggested that the protonation/deprotonation coupling of the surface acidic charges was due to the electrostatic repulsion among them.

6.3.2 Arg-to-Lys variant has good structural arrangement of surface charges for stabilizing protein

Statistical analysis of comparison between thermophilic and mesohpilic homologues pair revealed thermophilic proteins in general have increased number of surface charges suggesting the extensive surface charge-charge interaction takes an important role in improving protein thermostability[56,92,93,96-98]. However, increased surface charge density would also increase the repulsion between same charges in addition to the favorable electrostatic interactions. As discussed in section 6.3.1, most of the surface acidic residues experienced repulsive electrostatic interactions as indicated by their non-HH titration curves. Yet, the high surface charge density also increased number of salt bridges which were shown to have significant role in stabilizing the protein as discussed in Chapter 3. Therefore, to determine whether the global electrostatic interactions of a charge is stabilizing or destabilizing, structural arrangement of charges should also be taken into account. If the protein has

optimized charge arrangement, most of the charges should be stabilizing event they experience repulsive forces.

pK_a shift approach is the most suitable method to address this problem. While the electrostatic repulsion experienced by a charge can be indicated by it non-HH titration curve, the stabilizing effect due to its global electrostatic interactions can be estimated from its reliable pK_a value fitted from the major transition (provided that the pK_a value in denatured state is known). In our study, 8 out of 13 acidic residues have reliable native pK_a values fitted by microscopic model or GloFTE. As there is no obvious residual structure in the denatured state of wild-type *T. celer L30e* (see Chapter 4), it is not expected that the quintuple Arg-to-Lys variant, which is almost identical to the wild-type protein in structural, thermodynamics, and electrostatic properties (see Chapter 5), has strong residual structures in its denatured state. Therefore, random coil pK_a values obtained from model peptides (3.87 and 4.31 or Asp and Glu respectively at 298 K, 3.87 and 4.36 for Asp and Glu respectively at 333 K) were used as the denatured pK_a values for the acidic residues in the pK_a shift approach calculation. Be noted that this may not be always valid as no observable residual structures (α-helix like structure) is not equivalent to no interaction (especially electrostatic interaction) in the denatured state (see Section 6.3.3).

As an independent validation of the experimentally obtained pK_a values, they

were compared with the computational predicted pK_a values (Figure 6.10). Based on the modeled crystal structure of Arg-to-Lys variant (modeled the flexible C-terminal tail), pK_a values of the 13 acidic charges were predicted in 3 independent online servers, namely KARLSBERG+, H++, and PROPKA. While QM/MM based PROPKA results only acceptably correlated with reliable experimental pK_a values obtained at 298 K with correlation of 0.63, Poisson-Boltzmann based KARLSBERG+ and H++ predictions gave a good correlation of 0.80 and 0.79 respectively. Nevertheless, all of them could correctly predict the overall trend. For example, salt bridged residues Glu-6 and Glu-62 had significantly downshifted pK_a values, and destabilizing Glu-50 had elevated pK_a value. The convergence of experimental and predicted values provided an independent evidence for the accuracy of the overall trends of the experimental obtained pK_a values.

In pK_a shift approach, large downshift of native pK_a value when compared to the denatured state pK_a value indicates the strong stabilizing effect of the charge. It was found that 7 out of the 8 residues which have reliable pK_a values have largely downshifted pK_a values (Table 6.1) when compared to the model denatured pK_a values. The only destabilizing residue was Glu-50, where the reason is obviously the strong repulsive force of nearby Glu-62 as discussed in section 6.3.1. However, Glu-62 was found to have largely downshifted native pK_a value (2.4 obtained from

GloFTE). The stabilizing effect of Glu-62 comes from the strong salt bridge coupling free energy with Lys-46 as discussed in Chapter 3. This result showed that the stabilizing effect of salt bridge outweighed the destabilizing effect of repulsive electrostatic interaction for Glu-62. Other salt-bridge forming residue Glu-6 (with Arg-92) also has a significant downshifted native pK_a value, revealing that salt bridge interactions strongly stabilize the protein. Asp-12 and Asp-44 also have close proximity (< 8 Å) to nearby unlike charges (Lys-8 and Lys-9 for Asp-12, Lys-42 for Asp-44) (Figure 6.1). Therefore their downshifted pK_a values are likely due to the favourable electrostatic interactions with their nearby unlike charges. Not only salt bridge interaction, long range electrostatic interaction also play a role in stabilizing protein as indicated by the downshifted native pK_a value of the well isolated residues of Asp-2, Glu-69, and Asp-87 (Figure 6.1). The downshifted native pK_a values of Asp-2 and Glu-69 should also be contributed by their helix capping effect (Asp-2 caps helix-2, Glu-69 caps helix-4). The fact that most of the residues have downshifted pK_a values indicated most of the surface charges stabilized the protein, even most of them (except Glu-6 and Asp-87) more or less experienced repulsive electrostatic interactions as indicated by their minor transition. This observation leads to an important conclusion that the structural arrangement of the charges in the quintuple Arg-to-Lys variant is optimized for stabilizing the protein, where the

surface charges contribute to protein stability through salt bridges, long range electrostatic interaction, or interaction with permanent dipole of helix so that the repulsive forces among them were outweighed.

The stabilizing surface charge distribution of this the quintuple Arg-to-Lys is consistent with the theoretical calculation on electrostatic contributions to protein stability, which suggested charged groups on protein surfaces (especially for thermophilic proteins) are in most cases optimized for maximizing stabilizing interactions[53]. In fact, optimizing surface charge-charge interaction was demonstrated to be a relatively convenient method for improving protein stability[210]. One good example for improving protein stability by optimizing surface charge-charge interaction was recently demonstrated by Makhatadze G.I. and co-workers[211]. They had optimized surface charges distribution of human acylphophatase (AcPh) and Cdc42 GTPase by an algorithm called TKSA-GA (genetic algorithm (GA) for optimizing charge-charge interaction, with the interaction energies calculated by surface area corrected (SA) Tanford-Kirkwood (TK) formalism), and experimentally showed the melting temperatures of the designed variants of AcPh and Cdc42 have increased by 10 °C. Other examples of improving protein stability by optimizing surface charge distribution have also been reported elsewhere[60,210,212].

6.3.3 High charge density in sequence can downshift pK_a values in denatured state

The contributions to protein stability of residues with reliable pK_a values were calculated according to the difference between the reliable pK_a vales and random coil pK_a values obtained from pH titration of model peptides (Figure 6.11). The calculated free energy contributions were compared with that obtained from mutagenesis experiments[54] (Figure 6.13). Two data only have good correlation of 0.7. However, Asp-12 and Asp-44 were found to be the outliners. Regardless of these two residues, the correlation between two sets of data was increased to 0.9.

One reason to explain the low correlation of free energy contribution of Asp-12 and Asp-44 obtained from pK_a shift approach and mutagenesis may come from the accuracy of pK_a values in denatured state. Because of the large scaling factor used in the calculation of free energy contribution from pK_a values (i.e. -RTln(10)), a small error in either native or denatured state pK_a values can be largely scaled up (change in 1 pK_a unit account for 5.7 kJ mol^{-1}). To improve the accuracy of the fitted native pK_a values in native state protein, a small pH interval of 0.25 was applied in pH ranging from pH 1.0 to pH 3.0, where most of the major titrations had appeared, to maximize the resolution of the titration curves. However, the accuracy of native pK_a values in this study should be reliable as the used pK_a values were fitted from large

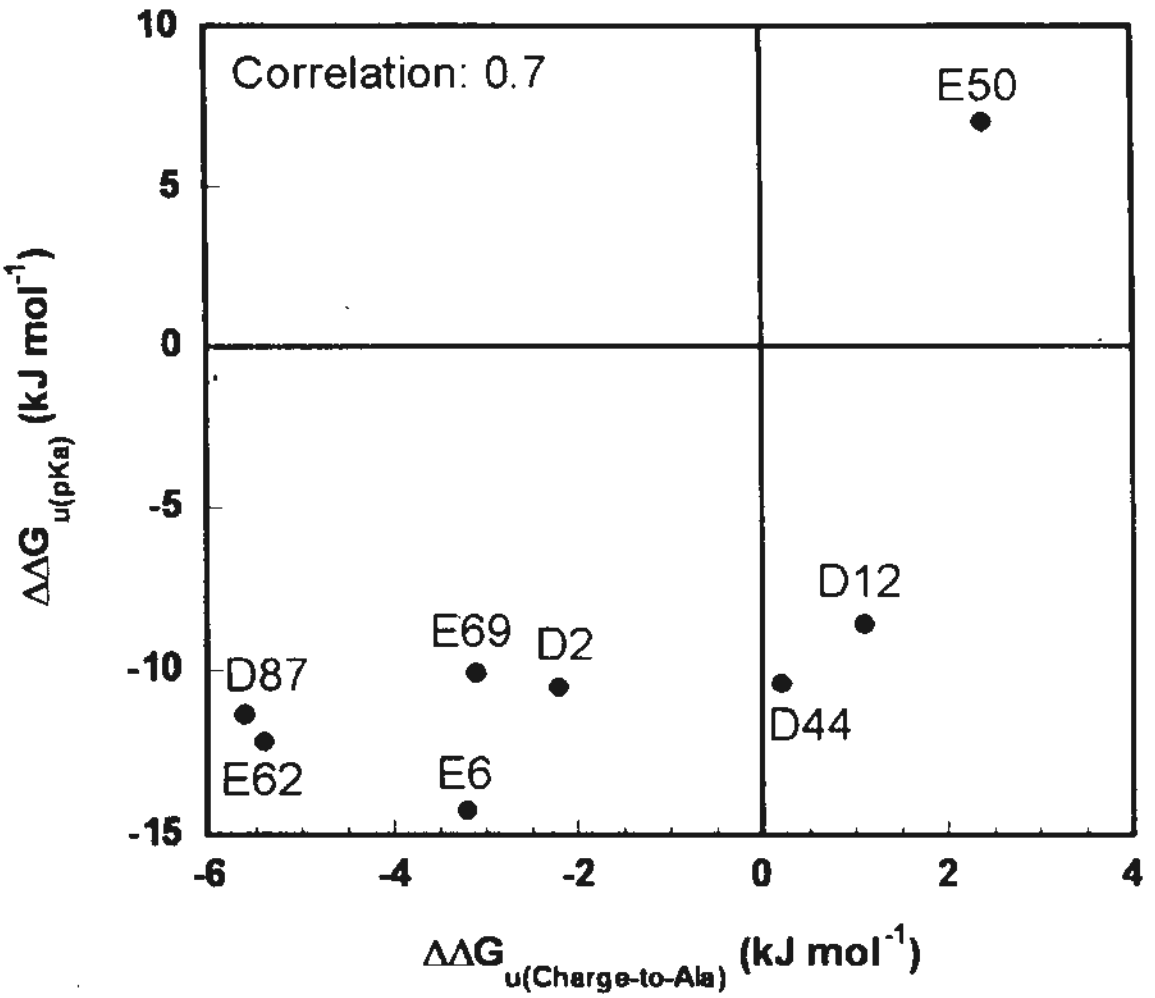


Figure 6.13. Comparison of pK_a shift approach and mutagenesis. Comparison between $\Delta\Delta G_u$ found by single charge-to-Ala mutation and that found by pK_a shift approach. Two set of data correlated well in general (correlation 0.7). Asp-12 and Asp-44 were found not well correlated. Be noted the correlation of these two set data improve greatly from 0.7 to 0.9 if the data point of Asp-12 and Asp-44 were excluded.

major transition with low fitting error. On the other hand, random coil pK_a values derived from standard peptide used as the denatured state pK_a in the calculation may not be confidently reliable. As discussed above, the usage of peptide pK_a values is based on the assumption that no observable residual structure in wild-type *T. celer L30e* and the extremely high similarity between Arg-to-Lys variant and wild-type *T. celer L30e*. However, one major difference between model peptide and denatured protein with residual structures is that the charges in the former is isolated, but more than one charge can appear in the local environment of a residue in the latter. Since the charge density in the protein sequence is very high (in average charged residue appear in every 4 residues), charged residues may have interactions with each other and alter the denatured pK_a values even no residual structures (more precisely, α-helix structures, which secondary shift and NOE can detect) are present[129]. If strong attractive electrostatic interactions of a charged residue were present in the denatured state, the denatured state pK_a value will also be greatly reduced, leading to the probability that the residue is less stabilizing or even destabilizing even it has a downshifted native state pK_a value. In the native protein, pK_a values of Asp-12 and Asp-44 were greatly downshifted by strong favourable interaction with Lys-9 and Lys-42 respectively according to the crystal structure. However, according to the protein sequence, Asp-12 and Asp-44 are the only two residues which are flanked by

two sequentially nearby (within i ± 3) unlike charges (Lys-9 and Lys-15 for Asp-12; Lys-42 and Lys-46 for Asp-44), which may interact with Asp-12 and Asp-44 and downshift their denatured state pK_a values. It may be the reason why Asp-12 and Asp-44 has strongly downshifted native state value, but the mutagenesis experiment showed they were slightly destabilizing. To verify this hypothesis, a polyproline II (PPII) based denatured state model were built. PPII was used as the template of backbone because its conformation was found to be one of the most important conformations in the ensemble of denatured states of proteins[165,167-169]. The pK_a values of all Asp and Glu in this denatured state model were predicted by H++ online server. The result showed the predicted pK_a values of Asp-12 (3.8) and Asp-44 (4.0) were indeed smaller when compared with that of other Asp (4.2 for Asp-2; 4.3 for Asp-48, 4.5 for Asp-87). This predicted result not only supports the hypothesis that explain the inconsistence between pK_a and mutagenesis data, but also reveals the high charge density in protein sequence can lead to destabilizing effect due to the stabilization of the denatured state.

From the result, we can also found that the free energy contributions estimated by pK_a data are much larger than those estimated by mutagenesis data. Since the free energy contributions is calculated from rather unrealistic denatured state pK_a values, interpretation of the magnitude does not seem to have much meaning because of the

assumption that denatured state pK_a is equal to random coil pK_a of model peptide may not hold.

6.3.4 Increasing temperature will linearly scale up the stabilizing or destabilizing effect of charged residues

To investigate the effect of temperature on protein stability, pK_a values of Asp and Glu were estimated at both 298 K and 333 K. For all residues with reliable pK_a value, their differences in the change of pK_a ($\Delta pK_{a(333K-298K)}$) and the change of free energy contribution to protein stability ($\Delta\Delta G_{u(333K-298K)}$) were calculated and summarized in figure 6.14.

Although 5 out of the 8 reliable pK_a values in native state protein have shown decreased values with 35 °C higher in temperature, $\Delta\Delta pK_a$ are so small that the magnitudes do not exceed 0.1 units (Figure 6.14). Taking the fitting error into account, an objective conclusion is that pK_a values of acidic residues in native state protein are temperature independent. The temperature dependency of free energy contribution also depends on the effect of temperature on denatured state pK_a values. Although pK_a values estimated from poly-glycine based model peptides may not truly reflect the pK_a values in denatured state, its temperature dependency of pK_a should be taken as a good reference. pH titration of model peptides at 298 K and 333 K showed the pK_a values of Asp and Glu are essentially not affected by temperature.

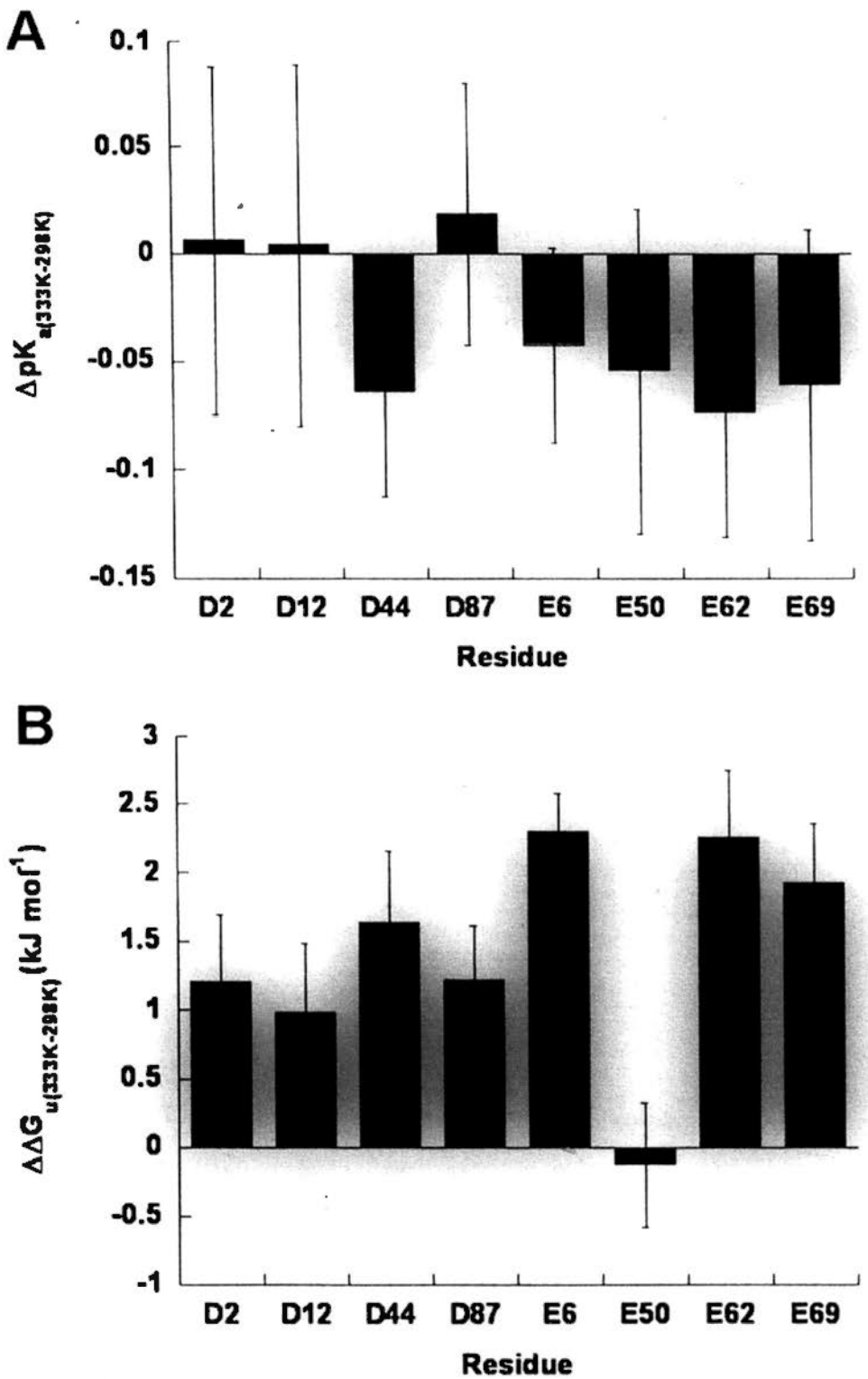


Figure 6.14. Temperature effect on electrostatic contribution. The effect of increasing temperature by 35 K on (A) the difference between ΔpK_a values of a residue at 298 K and 333 K ($\Delta pK_{a(333K-298K)}$), and (B) the difference between electrostatic contribution of a residue at 298 K and 333 K ($\Delta\Delta G_{u(333K-298K)}$). Be noted that $\Delta pK_{a(333K-298K)}$ were so small that one could confidently concluded that ΔpK_a is temperature independent.

The pK_a value differences of Asp and Glu in model peptide for 35 °C temperature increase are 0.0 ± 0.0 and 0.1 ± 0.0 respectively (Figure 6.11).

However, the temperature independency of pK_a values in both native and denatured state does not imply temperature has no effect on the free energy contribution of a charge. Reminded the calculated free energy contribution depends on temperature in addition to the difference of pK_a values in native and denatured state. Even there is no change in pK_a difference, increasing temperature will increase the magnitude of calculated free energy contribution (Figure 6.14). Except for the destabilizing Glu-50, which has up-shifted native pK_a values, the $\Delta\Delta G_{u(333K-298K)}$ was found to be about 1.6 kJ mol^{-1} in average for 35 °C increase in temperature. Although the temperature has the linear scaling effect on free energy contribution, the scaling effect was shown to be small, which improves the protein stability by less than 10 % for 35 °C increase in temperature, regardless the error in calculation.

Summing up, free energy contribution of a charged residue is mainly dependent on the difference of native and denatured pK_a values, which is temperature independent. Increasing temperature can only linearly but slowly scale up the stabilizing or destabilizing effect of a charged residue. Be noted that this linear relationship between electrostatic contribution and temperature is based on the assumption that all assumptions applied in pK_a approach are valid, and has neglected

the effect of solvent on protonation and deprotonation (Born effect).

6.3.5 Concluding remarks

11 out of 13 acidic residues showed non-HH behavior in their titration curves, which indicates the electrostatic interactions among them are very common and complicated. This complicated electrostatic interaction is probably due the unusually high surface charge density. Despite of the extensive unfavourable electrostatic interactions among acidic residues (e.g. 12.6 kJ mol^{-1} between Glu-6 and Glu-50 at 298 K), most of them are found to contribute to protein stability, which is indicated by the large down-shifts of their pK_a values in native protein when compared to that in random coil pK_a values obtained from model peptides. This observation indicated the surface charge distribution of Arg-to-Lys variant, and hence the wild-type *T. celer L30e*, has been optimized for stabilizing, which is consistent with the theoretical calculation that surface charges are in most cases optimized especially in thermophilic protein[53].

Contribution of charges to protein stability can be further complicated by its electrostatic interactions in denatured state. High charge density in sequence may lead to stabilization of some conformation of denatured state, like the cases in Asp-12 and Asp-44, Thus, even Asp-12 and Asp-44 were found to have largely downshifted pK_a values in native proteins, mutagenesis experiment had shown they

are slightly destabilizing[54].

Lastly, we have demonstrated that temperature has no effect on pK_a values, neither in native state nor in denatured state, where the difference between pK_a values in native and denatured state dominants the electrostatic contribution. Increase in temperature could only marginally improve the protein stability, assuming all pK_a shift approach assumptions are valid and protonation/deprotonation is not affected by Born effect.

Chapter 7: Conclusions

Electrostatic interaction has long been proposed to be an important factor for stabilizing protein since the first suggestion by Perutz and Raidt[94,95]. The subsequence statistical analysis of comparison between thermophilic and mesophilic proteins also suggested that the elevated surface charges of thermophilic should be an especially important factor for thermostability of protein, as having more surface electrostatic interactions is one of the common structural features found in thermophilic proteins when compared to their mesophilic homologues[56,92,93,96-98]. Moreover, the experimental evidence supporting the stabilizing role of electrostatic interactions has also been being accumulated[48,50-52]. In this study, a comprehensive study on electrostatic contribution was carried out to investigate how charge-charge interactions, including salt-bridge and structural arrangement of charges, can improve the protein stability, especially for the thermostability. In order to quantitatively investigate the electrostatic contribution to protein stability, two complementary approaches, namely the double mutant cycle approach and pK_a shift approach, were performed.

In the double mutant cycle approach, coupling free energy between two salt bridges (E6/R92 and K46/E62), as well as that of a long range ion pair (E90/R92) were estimated by using circular dichroism to find out the thermodynamic

parameters of the protein model, *Thermococcus celer L30e*, and its charge-to-neutral mutants. It was found that the coupling free energy was temperature independent and was about 3 kJ mol^{-1} per salt bridge. Also, by using a novel analysis of double mutant cycle of ΔC_p, it was also found the interaction of salt bridge plays an important role in reducing ΔC_p. The temperature independency of coupling free energy and the effect of reduced ΔC_p could explain very well the general observation that thermophilic proteins have highly up-shifted and broadened protein stability curves, is due to its elevated strong salt bridge electrostatic interactions when compared with their mesophilic homologs.

To overcome the unwanted crystallization problem of wild-type *T. celer L30e* in low ionic strength neutral pH NMR conditions, which were essential for the pK_a shift approach, a quintuple Arg-to-Lys variant was designed to improve dramatically its crystalline solubility and pH tolerance while conserving its surface charges, structural, thermodynamic, and electrostatic properties. It was also shown that electrostatic interaction plays a critical role in crystallization in low ionic strength conditions, and arginine residue was especially important in crystal packing because of its high ability of forming salt bridges and hydrogen bonds.

In the pK_a shift approach, the native state pK_a values of acidic residues were obtained by fitting the side chain carboxyl ^{13}C chemical shifts to microscopic model

or global fitting of titrational event (GloFTE), whereas the denatured state pK_a values were obtained by conventional pH titration of terminal protected 5-residue glycine-based model peptide. It was found that the surface charge-charge interactions, either attractive between unlike charges, or repulsive between like charges, were strong and complicated because of the high surface charge density of *T. celer L30e*. However, the fact that most of the acidic residues have significantly downshifted native state pK_a values indicated the surface charge distribution of *T. celer L30e* is optimized for stabilizing the protein. In addition, we have also showed that temperature has negligible effect on pK_a values in both native state and denatured state, therefore temperature could only marginally amplify the stabilizing effect in linear manner. It is seemingly the linear correlation of protein stability and temperature found in pK_a shift approach is contradictory to the temperature independency of coupling free energy found in double mutant cycle approach. However, we have to be reminded that the limitation of both approaches. While the double mutant cycle approach may take the non-electrostatic interaction into account, pK_a shift approach will neglect any non-titrational effects of the charged residues that affect protein stability, such as temperature dependent generalized Born (GB) effect. The inclusion and exclusion of non-electrostatic interaction in the electrostatic contribution may be the reason to explain the small inconsistence of the two

approaches. Nevertheless, the linear dependency in pK_a shift approach was small (increase overall protein stability by about 1.6 kJ mol^{-1} with 35 °C increased in temperature), and the fact that temperature has only minimal effect on protein stability was evidently demonstrated.

Denatured state of a protein is in principle of the same importance to protein stability. However, it has been shown that wild-type *T. celer L30e* has no observable residual structure in its guanidine HCl-induced denatured state, indicating that denatured state of *T. celer L30e* should not have large effect on global protein stability, although the high sequential charge density may affect the local electrostatic contribution of some charged residues such as Asp-12 and Asp-44 of *T. celer L30e*.

References

1. Levinthal, C. (1968). Are there pathways for protein folding? *Journal of Chemical Physics* 65, 44-45.
2. Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science* 181, 223-230.
3. Baker, D., Sohl, J. L. & Agard, D. A. (1992). A protein-folding reaction under kinetic control. *Nature* 356, 263-265.
4. Berkenpas, M. B., Lawrence, D. A. & Ginsburg, D. (1995). Molecular evolution of plasminogen activator inhibitor-1 functional stability. *EMBO J* 14, 2969-2977.
5. Dill, K. A. (1990). Dominant forces in protein folding. *Biochemistry* 29, 7133-7155.
6. Kim, P. S. & Baldwin, R. L. (1990). Intermediates in the folding reactions of small proteins. *Annu Rev Biochem* 59, 631-660.
7. Mitraki, A., Fane, B., Haase-Pettingell, C., Sturtevant, J. & King, J. (1991). Global suppression of protein folding defects and inclusion body formation. *Science* 253, 54-58.
8. Thomas, P. J., Qu, B. H. & Pedersen, P. L. (1995). Defective protein folding as a basis of human disease. *Trends Biochem Sci* 20, 456-459.
9. Govindarajan, S. & Goldstein, R. A. (1998). On the thermodynamic hypothesis of protein folding. *Proc Natl Acad Sci U S A* 95, 5545-5549.
10. Ekblad, C. M., Wilkinson, H. R., Schymkowitz, J. W., Rousseau, F., Freund, S. M. & Itzhaki, L. S. (2002). Characterisation of the BRCT domains of the breast cancer susceptibility gene product BRCA1. *J Mol Biol* 320, 431-442.
11. Main, E. R., Jackson, S. E. & Regan, L. (2003). The folding and design of repeat proteins: reaching a consensus. *Curr Opin Struct Biol* 13, 482-489.
12. Mosavi, L. K., Cammett, T. J., Desrosiers, D. C. & Peng, Z. Y. (2004). The ankyrin repeat as molecular architecture for protein recognition. *Protein Sci* 13, 1435-1448.
13. Li, J., Mahajan, A. & Tsai, M. D. (2006). Ankyrin repeat: a unique motif mediating protein-protein interactions. *Biochemistry* 45, 15168-15178.
14. Lowe, A. R. & Itzhaki, L. S. (2007). Rational redesign of the folding pathway of a modular protein. *Proc Natl Acad Sci U S A* 104, 2679-2684.
15. Lowe, A. R. & Itzhaki, L. S. (2007). Biophysical characterisation of the small ankyrin repeat protein myotrophin. *J Mol Biol* 365, 1245-1255.
16. Werbeck, N. D. & Itzhaki, L. S. (2007). Probing a moving target with a plastic unfolding intermediate of an ankyrin-repeat protein. *Proc Natl Acad*

Sci U S A 104, 7863-7868.

17. Barrick, D., Ferreiro, D. U. & Komives, E. A. (2008). Folding landscapes of ankyrin repeat proteins: experiments meet theory. *Curr Opin Struct Biol* 18, 27-34.
18. Werbeck, N. D., Rowling, P. J., Chellamuthu, V. R. & Itzhaki, L. S. (2008). Shifting transition states in the unfolding of a large ankyrin repeat protein. *Proc Natl Acad Sci U S A* 105, 9982-9987.
19. Baldwin, R. L. & Rose, G. D. (1999). Is protein folding hierarchic? II. Folding intermediates and transition states. *Trends Biochem Sci* 24, 77-83.
20. Baldwin, R. L. & Rose, G. D. (1999). Is protein folding hierarchic? I. Local structure and peptide folding. *Trends Biochem Sci* 24, 26-33.
21. Frauenfelder, H., Petsko, G. A. & Tsernoglou, D. (1979). Temperature-dependent X-ray diffraction as a probe of protein structural dynamics. *Nature* 280, 558-563.
22. Karplus, M. & Petsko, G. A. (1990). Molecular dynamics simulations in biology. *Nature* 347, 631-639.
23. Rasmussen, B. F., Stock, A. M., Ringe, D. & Petsko, G. A. (1992). Crystalline ribonuclease A loses function below the dynamical transition at 220 K. *Nature* 357, 423-424.
24. Pace, C. N. (1990). Conformational stability of globular proteins. *Trends Biochem Sci* 15, 14-17.
25. Makhatadze, G. I. & Privalov, P. L. (1995). Energetics of protein structure. *Adv Protein Chem* 47, 307-425.
26. Privalov, P. L. & Gill, S. J. (1988). Stability of protein structure and hydrophobic interaction. *Adv Protein Chem* 39, 191-234.
27. Baldwin, R. L. (1986). Temperature dependence of the hydrophobic interaction in protein folding. *Proc Natl Acad Sci U S A* 83, 8069-8072.
28. Ooi, T. & Oobatake, M. (1988). Effects of hydrated water on protein unfolding. *J Biochem* 103, 114-120.
29. Rocha, R., Leal, S. S., Teixeira, V. H., Regalla, M., Huber, H., Baptista, A. M., Soares, C. M. & Gomes, C. M. (2006). Natural domain design: enhanced thermal stability of a zinc-lacking ferredoxin isoform shows that a hydrophobic core efficiently replaces the structural metal site. *Biochemistry* 45, 10376-10384.
30. Chen, J. & Stites, W. E. (2001). Energetics of side chain packing in staphylococcal nuclease assessed by systematic double mutant cycles. *Biochemistry* 40, 14004-14011.
31. Chen, J. & Stites, W. E. (2001). Packing is a key selection factor in the

evolution of protein hydrophobic cores. *Biochemistry* 40, 15280-15289.

32. Chen, J. & Stites, W. E. (2001). Higher-order packing interactions in triple and quadruple mutants of staphylococcal nuclease. *Biochemistry* 40, 14012-14019.
33. Holder, J. B., Bennett, A. F., Chen, J., Spencer, D. S., Byrne, M. P. & Stites, W. E. (2001). Energetics of side chain packing in staphylococcal nuclease assessed by exchange of valines, isoleucines, and leucines. *Biochemistry* 40, 13998-14003.
34. Gao, J., Bosco, D. A., Powers, E. T. & Kelly, J. W. (2009). Localized thermodynamic coupling between hydrogen bonding and microenvironment polarity substantially stabilizes proteins. *Nat Struct Mol Biol* 16, 684-690.
35. Wang, M., Wales, T. E. & Fitzgerald, M. C. (2006). Conserved thermodynamic contributions of backbone hydrogen bonds in a protein fold. *Proc Natl Acad Sci U S A* 103, 2600-2604.
36. Kumar, S. & Nussinov, R. (2002). Close-range electrostatic interactions in proteins. *Chembiochem* 3, 604-617.
37. Matthews, B. W. (1993). Structural and genetic analysis of protein stability. *Annu Rev Biochem* 62, 139-160.
38. Fersht, A. R. & Serrano, L. (1993). Principles of protein stability derived from protein engineering experiments. *Current Opinion in Structural Biology* 3, 75-83.
39. Gitlin, I., Gudiksen, K. L. & Whitesides, G. M. (2006). Effects of surface charge on denaturation of bovine carbonic anhydrase. *Chembiochem* 7, 1241-1250.
40. Kumar, S. & Nussinov, R. (1999). Salt bridge stability in monomeric proteins. *J Mol Biol* 293, 1241-1255.
41. Kumar, S. & Nussinov, R. (2000). Fluctuations between stabilizing and destabilizing electrostatic contributions of ion pairs in conformers of the c-Myc-Max leucine zipper. *Proteins* 41, 485-497.
42. Kumar, S. & Nussinov, R. (2001). Fluctuations in ion pairs and their stabilities in proteins. *Proteins* 43, 433-454.
43. Strop, P. & Mayo, S. L. (2000). Contribution of surface salt bridges to protein stability. *Biochemistry* 39, 1251-1255.
44. Hendsch, Z. S., Jonsson, T., Sauer, R. T. & Tidor, B. (1996). Protein stabilization by removal of unsatisfied polar groups: computational approaches and experimental tests. *Biochemistry* 35, 7621-7625.
45. Lumb, K. J. & Kim, P. S. (1995). Measurement of interhelical electrostatic interactions in the GCN4 leucine zipper. *Science* 268, 436-439.

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

47. Marti, D. N., Jelesarov, I. & Bosshard, H. R. (2000). Interhelical ion pairing in coiled coils: solution structure of a heterodimeric leucine zipper and determination of pKa values of Glu side chains. *Biochemistry* 39, 12804-12818.

48. Anderson, D. E., Becktel, W. J. & Dahlquist, F. W. (1990). pH-induced denaturation of proteins: a single salt bridge contributes 3-5 kcal/mol to the free energy of folding of T4 lysozyme. *Biochemistry* 29, 2403-2408.

49. Kumar, S. & Bansal, M. (1998). Dissecting alpha-helices: position-specific analysis of alpha-helices in globular proteins. *Proteins* 31, 460-476.

50. Spek, E. J., Bui, A. H., Lu, M. & Kallenbach, N. R. (1998). Surface salt bridges stabilize the GCN4 leucine zipper. *Protein Sci* 7, 2431-2437.

51. Takano, K., Tsuchimori, K., Yamagata, Y. & Yutani, K. (2000). Contribution of salt bridges near the surface of a protein to the conformational stability. *Biochemistry* 39, 12375-12381.

52. Vetriani, C., Maeder, D. L., Tolliday, N., Yip, K. S., Stillman, T. J., Britton, K. L., Rice, D. W., Klump, H. H. & Robb, F. T. (1998). Protein thermostability above 100 degreesC: a key role for ionic interactions. *Proc Natl Acad Sci U S A* 95, 12300-12305.

53. Xiao, L. & Honig, B. (1999). Electrostatic contributions to the stability of hyperthermophilic proteins. *J Mol Biol* 289, 1435-1444.

54. Lee, C. F., Makhatadze, G. I. & Wong, K. B. (2005). Effects of charge-to-alanine substitutions on the stability of ribosomal protein L30e from Thermococcus celer. *Biochemistry* 44, 16817-16825.

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

56. Fukuchi, S. & Nishikawa, K. (2001). Protein surface amino acid compositions distinctively differ between thermophilic and mesophilic bacteria. *J Mol Biol* 309, 835-843.

57. Hendsch, Z. S. & Tidor, B. (1994). Do salt bridges stabilize proteins? A continuum electrostatic analysis. *Protein Sci* 3, 211-226.

58. Bas, D. C., Rogers, D. M. & Jensen, J. H. (2008). Very fast prediction and rationalization of pKa values for protein-ligand complexes. *Proteins* 73, 765-783.

59. Li, H., Robertson, A. D. & Jensen, J. H. (2005). Very fast empirical prediction and rationalization of protein pKa values. *Proteins* 61, 704-721.

60. Loladze, V. V., Ibarra-Molero, B., Sanchez-Ruiz, J. M. & Makhatadze, G. I. (1999). Engineering a thermostable protein via optimization of charge-charge interactions on the protein surface. *Biochemistry* 38, 16419-16423.

61. Guerois, R., Nielsen, J. E. & Serrano, L. (2002). Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J Mol Biol* 320, 369-387.

62. Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F. & Serrano, L. (2005). The FoldX web server: an online force field. *Nucleic Acids Res* 33, W382-388.

63. Schymkowitz, J. W., Rousseau, F., Martins, I. C., Ferkinghoff-Borg, J., Stricher, F. & Serrano, L. (2005). Prediction of water and metal binding sites and their affinities by using the Fold-X force field. *Proc Natl Acad Sci U S A* 102, 10147-10152.

64. Sippl, M. J. (1995). Knowledge-based potentials for proteins. *Curr Opin Struct Biol* 5, 229-235.

65. Lazaridis, T. & Karplus, M. (2000). Effective energy functions for protein structure prediction. *Curr Opin Struct Biol* 10, 139-145.

66. Funahashi, J., Sugita, Y., Kitao, A. & Yutani, K. (2003). How can free energy component analysis explain the difference in protein stability caused by amino acid substitutions? Effect of three hydrophobic mutations at the 56th residue on the stability of human lysozyme. *Protein Eng* 16, 665-671.

67. Gromiha, M. M., An, J., Kono, H., Oobatake, M., Uedaira, H. & Sarai, A. (1999). ProTherm: Thermodynamic Database for Proteins and Mutants. *Nucleic Acids Res* 27, 286-288.

68. Kumar, M. D., Bava, K. A., Gromiha, M. M., Prabakaran, P., Kitajima, K., Uedaira, H. & Sarai, A. (2006). ProTherm and ProNIT: thermodynamic databases for proteins and protein-nucleic acid interactions. *Nucleic Acids Res* 34, D204-206.

69. Ackers, G. K. & Smith, F. R. (1985). Effects of site-specific amino acid modification on protein interactions and biological function. *Annu Rev Biochem* 54, 597-629.

70. Carter, P. J., Winter, G., Wilkinson, A. J. & Fersht, A. R. (1984). The use of double mutants to detect structural changes in the active site of the tyrosyl-tRNA synthetase (Bacillus stearothermophilus). *Cell* 38, 835-840.

71. Horovitz, A. (1996). Double-mutant cycles: a powerful tool for analyzing protein structure and function. *Fold Des* 1, R121-126.

72. Serrano, L., Horovitz, A., Avron, B., Bycroft, M. & Fersht, A. R. (1990). Estimating the contribution of engineered surface electrostatic interactions to protein stability by using double-mutant cycles. *Biochemistry* 29, 9343-9352.

73. Bosshard, H. R., Marti, D. N. & Jelesarov, I. (2004). Protein stabilization by salt bridges: concepts, experimental approaches and clarification of some misunderstandings. *J Mol Recognit* 17, 1-16.

74. Luisi, D. L., Snow, C. D., Lin, J. J., Hendsch, Z. S., Tidor, B. & Raleigh, D. P. (2003). Surface salt bridges, double-mutant cycles, and protein stability: an experimental and computational analysis of the interaction of the Asp 23 side chain with the N-terminus of the N-terminal domain of the ribosomal protein l9. *Biochemistry* 42, 7050-7060.

75. Makhatadze, G. I., Loladze, V. V., Ermolenko, D. N., Chen, X. & Thomas, S. T. (2003). Contribution of surface salt bridges to protein stability: guidelines for protein engineering. *J Mol Biol* 327, 1135-1148.

76. Forsyth, W. R. & Robertson, A. D. (2000). Insensitivity of perturbed carboxyl pK(a) values in the ovomucoid third domain to charge replacement at a neighboring residue. *Biochemistry* 39, 8067-8072.

77. Huyghues-Despointes, B. M., Thurlkill, R. L., Daily, M. D., Schell, D., Briggs, J. M., Antosiewicz, J. M., Pace, C. N. & Scholtz, J. M. (2003). pK values of histidine residues in ribonuclease Sa: effect of salt and net charge. *J Mol Biol* 325, 1093-1105.

78. Laurents, D. V., Huyghues-Despointes, B. M., Bruix, M., Thurlkill, R. L., Schell, D., Newsom, S., Grimsley, G. R., Shaw, K. L., Trevino, S., Rico, M., Briggs, J. M., Antosiewicz, J. M., Scholtz, J. M. & Pace, C. N. (2003). Charge-charge interactions are key determinants of the pK values of ionizable groups in ribonuclease Sa (pI=3.5) and a basic variant (pI=10.2). *J Mol Biol* 325, 1077-1092.

79. Pace, C. N., Huyghues-Despointes, B. M., Briggs, J. M., Grimsley, G. R. & Scholtz, J. M. (2002). Charge-charge interactions are the primary determinants of the pK values of the ionizable groups in Ribonuclease T1. *Biophys Chem* 101-102, 211-219.

80. Sundd, M. & Robertson, A. D. (2003). Rearrangement of charge-charge interactions in variant ubiquitins as detected by double-mutant cycles and NMR. *J Mol Biol* 332, 927-936.

81. Forsyth, W. R., Antosiewicz, J. M. & Robertson, A. D. (2002). Empirical relationships between protein structure and carboxyl pKa values in proteins. *Proteins* 48, 388-403.

82. Kao, Y. H., Fitch, C. A., Bhattacharya, S., Sarkisian, C. J., Lecomte, J. T. &

Garcia-Moreno, E. B. (2000). Salt effects on ionization equilibria of histidines in myoglobin. *Biophys J* 79, 1637-1654.

83. Lee, K. K., Fitch, C. A., Lecomte, J. T. & Garcia-Moreno, E. B. (2002). Electrostatic effects in highly charged proteins: salt sensitivity of pKa values of histidines in staphylococcal nuclease. *Biochemistry* 41, 5656-5667.
84. Betz, S. F. (1993). Disulfide bonds and the stability of globular proteins. *Protein Sci* 2, 1551-1558.
85. Matsumura, M. & Matthews, B. W. (1991). Stabilization of functional proteins by introduction of multiple disulfide bonds. *Methods Enzymol* 202, 336-356.
86. Pace, C. N., Grimsley, G. R., Thomson, J. A. & Barnett, B. J. (1988). Conformational stability and activity of ribonuclease T1 with zero, one, and two intact disulfide bonds. *J Biol Chem* 263, 11820-11825.
87. Matthews, B. W., Nicholson, H. & Becktel, W. J. (1987). Enhanced protein thermostability from site-directed mutations that decrease the entropy of unfolding. *Proc Natl Acad Sci U S A* 84, 6663-6667.
88. Bowler, B. E. (2007). Thermodynamics of protein denatured states. *Mol Biosyst* 3, 88-99.
89. Pace, C. N., Alston, R. W. & Shaw, K. L. (2000). Charge-charge interactions influence the denatured state ensemble and contribute to protein stability. *Protein Sci* 9, 1395-1398.
90. Shortle, D. (1996). The denatured state (the other half of the folding equation) and its role in protein stability. *FASEB J* 10, 27-34.
91. Wong, K. B., Clarke, J., Bond, C. J., Neira, J. L., Freund, S. M., Fersht, A. R. & Daggett, V. (2000). Towards a complete description of the structural and dynamic properties of the denatured state of barnase and the role of residual structure in folding. *J Mol Biol* 296, 1257-1282.
92. Haney, P. J., Badger, J. H., Buldak, G. L., Reich, C. I., Woese, C. R. & Olsen, G. J. (1999). Thermal adaptation analyzed by comparison of protein sequences from mesophilic and extremely thermophilic Methanococcus species. *Proc Natl Acad Sci U S A* 96, 3578-3583.
93. Kumar, S., Tsai, C. J. & Nussinov, R. (2000). Factors enhancing protein thermostability. *Protein Eng* 13, 179-191.
94. Perutz, M. F. (1978). Electrostatic effects in proteins. *Science* 201, 1187-1191.
95. Perutz, M. F. & Raidt, H. (1975). Stereochemical basis of heat stability in bacterial ferredoxins and in haemoglobin A2. *Nature* 255, 256-259.
96. Querol, E., Perez-Pons, J. A. & Mozo-Villarias, A. (1996). Analysis of protein

conformational characteristics related to thermostability. *Protein Eng* 9, 265-271.

97. Szilagyi, A. & Zavodszky, P. (2000). Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey. *Structure* 8, 493-504.
98. Vogt, G. & Argos, P. (1997). Protein thermal stability: hydrogen bonds or internal packing? *Fold Des* 2, S40-46.
99. Lee, C. F., Allen, M. D., Bycroft, M. & Wong, K. B. (2005). Electrostatic interactions contribute to reduced heat capacity change of unfolding in a thermophilic ribosomal protein l30e. *J Mol Biol* 348, 419-431.
100. Kumar, S., Tsai, C. J. & Nussinov, R. (2001). Thermodynamic differences among homologous thermophilic and mesophilic proteins. *Biochemistry* 40, 14152-14165.
101. Dyballa, N. & Metzger, S. (2009). Fast and sensitive colloidal coomassie G-250 staining for proteins in polyacrylamide gels. *J Vis Exp*.
102. Wang, W. & Malcolm, B. A. (1999). Two-stage PCR protocol allowing introduction of multiple mutations, deletions and insertions using QuikChange Site-Directed Mutagenesis. *Biotechniques* 26, 680-682.
103. Santoro, M. M. & Bolen, D. W. (1988). Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. *Biochemistry* 27, 8063-8068.
104. Collaborative Computational Project, N. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 50, 760-763.
105. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. & Zwart, P. H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66, 213-221.
106. Vagin, A. A., Steiner, R. A., Lebedev, A. A., Potterton, L., McNicholas, S., Long, F. & Murshudov, G. N. (2004). REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr D Biol Crystallogr* 60, 2184-2195.
107. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60, 2126-2132.
108. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66, 486-501.

109. Hooft, R. W., Vriend, G., Sander, C. & Abola, E. E. (1996). Errors in protein structures. *Nature* 381, 272.

110. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6, 277-293.

111. Shrager, R. I., Cohen, J. S., Heller, S. R., Sachs, D. H. & Schechter, A. N. (1972). Mathematical models for interacting groups in nuclear magnetic resonance titration curves. *Biochemistry* 11, 541-547.

112. Blomberg, F., Maurer, W. & Ruterjans, H. (1977). Nuclear magnetic resonance investigation of 15N-labeled histidine in aqueous solution. *J Am Chem Soc* 99, 8149-8159.

113. Nielsen, J. E. (2007). Analysing the pH-dependent properties of proteins using pKa calculations. *J Mol Graph Model* 25, 691-699.

114. Sondergaard, C. R., McIntosh, L. P., Pollastri, G. & Nielsen, J. E. (2008). Determination of electrostatic interaction energies and protonation state populations in enzyme active sites. *J Mol Biol* 376, 269-287.

115. Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M. Y., Pieper, U. & Sali, A. (2006). Comparative protein structure modeling using Modeller. *Curr Protoc Bioinformatics* Chapter 5, Unit 5 6.

116. Sali, A. & Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234, 779-815.

117. Anandakrishnan, R. & Onufriev, A. (2008). Analysis of basic clustering algorithms for numerical estimation of statistical averages in biomolecules. *J Comput Biol* 15, 165-184.

118. Gordon, J. C., Myers, J. B., Folta, T., Shoja, V., Heath, L. S. & Onufriev, A. (2005). H++: a server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic Acids Res* 33, W368-371.

119. Rabenstein, B. & Knapp, E. W. (2001). Calculated pH-dependent population and protonation of carbon-monoxy-myoglobin conformers. *Biophys J* 80, 1141-1150.

120. Kieseritzky, G. & Knapp, E. W. (2008). Optimizing pKa computation in proteins with pH adapted conformations. *Proteins* 71, 1335-1348.

121. Guzman-Casado, M., Parody-Morreale, A., Robic, S., Marqusee, S. & Sanchez-Ruiz, J. M. (2003). Energetic evidence for formation of a pH-dependent hydrophobic cluster in the denatured state of Thermus thermophilus ribonuclease H. *J Mol Biol* 329, 731-743.

122. Murphy, K. P. & Freire, E. (1992). Thermodynamics of structural stability and cooperative folding behavior in proteins. *Adv Protein Chem* 43, 313-361.

123. Myers, J. K., Pace, C. N. & Scholtz, J. M. (1995). Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci* 4, 2138-2148.

124. Robic, S., Berger, J. M. & Marqusee, S. (2002). Contributions of folding cores to the thermostabilities of two ribonucleases H. *Protein Sci* 11, 381-389.

125. Robic, S., Guzman-Casado, M., Sanchez-Ruiz, J. M. & Marqusee, S. (2003). Role of residual structure in the unfolded state of a thermophilic protein. *Proc Natl Acad Sci U S A* 100, 11345-11349.

126. Spolar, R. S., Livingstone, J. R. & Record, M. T., Jr. (1992). Use of liquid hydrocarbon and amide transfer data to estimate contributions to thermodynamic functions of protein folding from the removal of nonpolar and polar surface from water. *Biochemistry* 31, 3947-3955.

127. Hollien, J. & Marqusee, S. (1999). A thermodynamic comparison of mesophilic and thermophilic ribonucleases H. *Biochemistry* 38, 3831-3836.

128. Motono, C., Oshima, T. & Yamagishi, A. (2001). High thermal stability of 3-isopropylmalate dehydrogenase from Thermus thermophilus resulting from low DeltaC(p) of unfolding. *Protein Eng* 14, 961-966.

129. Zhou, H. X. (2002). Toward the physical basis of thermophilic proteins: linking of enriched polar interactions and reduced heat capacity of unfolding. *Biophys J* 83, 3126-3133.

130. Horovitz, A., Serrano, L., Avron, B., Bycroft, M. & Fersht, A. R. (1990). Strength and co-operativity of contributions of surface salt bridges to protein stability. *J Mol Biol* 216, 1031-1044.

131. Sali, D., Bycroft, M. & Fersht, A. R. (1991). Surface electrostatic interactions contribute little of stability of barnase. *J Mol Biol* 220, 779-788.

132. Tissot, A. C., Vuilleumier, S. & Fersht, A. R. (1996). Importance of two buried salt bridges in the stability and folding pathway of barnase. *Biochemistry* 35, 6786-6794.

133. Ibarra-Molero, B., Zitzewitz, J. A. & Matthews, C. R. (2004). Salt-bridges can stabilize but do not accelerate the folding of the homodimeric coiled-coil peptide GCN4-p1. *J Mol Biol* 336, 989-996.

134. Marqusee, S. & Sauer, R. T. (1994). Contributions of a hydrogen bond/salt bridge network to the stability of secondary and tertiary structure in lambda repressor. *Protein Sci* 3, 2217-2225.

135. Lassila, K. S., Datta, D. & Mayo, S. L. (2002). Evaluation of the energetic contribution of an ionic network to beta-sheet stability. *Protein Sci* 11, 688-690.

136. Merkel, J. S., Sturtevant, J. M. & Regan, L. (1999). Sidechain interactions in parallel beta sheets: the energetics of cross-strand pairings. *Structure* 7, 1333-1343.

137. Ge, M., Xia, X. Y. & Pan, X. M. (2008). Salt bridges in the hyperthermophilic protein Ssh10b are resilient to temperature increases. *J Biol Chem* 283, 31690-31696.

138. Sun, D. P., Sauer, U., Nicholson, H. & Matthews, B. W. (1991). Contributions of engineered surface salt bridges to the stability of T4 lysozyme determined by directed mutagenesis. *Biochemistry* 30, 7142-7153.

139. Blasie, C. A. & Berg, J. M. (1997). Electrostatic interactions across a beta-sheet. *Biochemistry* 36, 6218-6222.

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

141. Privalov, P. L. (1990). Cold denaturation of proteins. *Crit Rev Biochem Mol Biol* 25, 281-305.

142. Razvi, A. & Scholtz, J. M. (2006). Lessons in stability from thermophilic proteins. *Protein Sci* 15, 1569-1578.

143. Anil, B., Craig-Schapiro, R. & Raleigh, D. P. (2006). Design of a hyperstable protein by rational consideration of unfolded state interactions. *J Am Chem Soc* 128, 3144-3145.

144. Anil, B., Song, B., Tang, Y. & Raleigh, D. P. (2004). Exploiting the right side of the Ramachandran plot: substitution of glycines by D-alanine can significantly increase protein stability. *J Am Chem Soc* 126, 13194-13195.

145. Stites, W. E., Meeker, A. K. & Shortle, D. (1994). Evidence for strained interactions between side-chains and the polypeptide backbone. *J Mol Biol* 235, 27-32.

146. Gribenko, A. V. & Makhatadze, G. I. (2007). Role of the charge-charge interactions in defining stability and halophilicity of the CspB proteins. *J Mol Biol* 366, 842-856.

147. Wallgren, M., Aden, J., Pylypenko, O., Mikaelsson, T., Johansson, L. B., Rak, A. & Wolf-Watz, M. (2008). Extreme temperature tolerance of a hyperthermophilic protein coupled to residual structure in the unfolded state. *J Mol Biol* 379, 845-858.

148. Fairbrother, W. J., Cavanagh, J., Dyson, H. J., Palmer, A. G., 3rd, Sutrina, S. L., Reizer, J., Saier, M. H., Jr. & Wright, P. E. (1991). Polypeptide backbone resonance assignments and secondary structure of Bacillus subtilis enzyme IIIglc determined by two-dimensional and three-dimensional heteronuclear NMR spectroscopy. *Biochemistry* 30, 6896-6907.

149. Meissner, A. & Sorensen, O. W. (2001). Sequential HNCACB and CBCANH protein NMR pulse sequences. *J Magn Reson* 151, 328-331.

150. Grzesiek, S. & Bax, A. (1993). Amino acid type determination in the sequential assignment procedure of uniformly 13C/15N-enriched proteins. *J Biomol NMR* 3, 185-204.

151. Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A. M. & Clore, G. M. (1989). Overcoming the overlap problem in the assignment of 1H NMR spectra of larger proteins by use of three-dimensional heteronuclear 1H-15N Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1 beta. *Biochemistry* 28, 6150-6156.

152. Montelione, G. T., Lyons, B. A., Emerson, S. D. & Tashiro, M. (1992). An efficient triple resonance experiment using carbon-13 isotropic mixing for determining sequence-specific resonance assignments of isotopically-enriched proteins. *J Am Chem Soc* 114, 10974-10975.

153. Wong, K. B., Freund, S. M. & Fersht, A. R. (1996). Cold denaturation of barstar: 1H, 15N and 13C NMR assignment and characterisation of residual structure. *J Mol Biol* 259, 805-818.

154. Beger, R. D. & Bolton, P. H. (1997). Protein phi and psi dihedral restraints determined from multidimensional hypersurface correlations of backbone chemical shifts and their use in the determination of protein tertiary structures. *J Biomol NMR* 10, 129-142.

155. Cornilescu, G., Delaglio, F. & Bax, A. (1999). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR* 13, 289-302.

156. Santiveri, C. M., Rico, M. & Jimenez, M. A. (2001). 13C(alpha) and 13C(beta) chemical shifts as a tool to delineate beta-hairpin structures in peptides. *J Biomol NMR* 19, 331-345.

157. Wang, Y. & Jardetzky, O. (2002). Investigation of the neighboring residue effects on protein chemical shifts. *J Am Chem Soc* 124, 14075-14084.

158. Wishart, D. S., Sykes, B. D. & Richards, F. M. (1991). Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *J Mol Biol* 222, 311-333.

159. Braun, D., Wider, G. & Wuthrich, K. (1994). Sequence-corrected 15N "Random Coil" chemical shifts. *J Am Chem Soc* 116, 8466-8469.

160. Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S. & Sykes, B. D. (1995). 1H, 13C and 15N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *J Biomol NMR* 5, 67-81.

161. Schwarzinger, S., Kroon, G. J., Foss, T. R., Chung, J., Wright, P. E. & Dyson, H. J. (2001). Sequence-dependent correction of random coil NMR chemical shifts. *J Am Chem Soc* 123, 2970-2978.

162. Ramachandran, G. N. & Sasisekharan, V. (1968). Conformation of polypeptides and proteins. *Adv Protein Chem* 23, 283-438.

163. Billeter, M., Braun, W. & Wuthrich, K. (1982). Sequential resonance assignments in protein 1H nuclear magnetic resonance spectra. Computation of sterically allowed proton-proton distances and statistical analysis of proton-proton distances in single crystal protein conformations. *J Mol Biol* 155, 321-346.

164. Wuthrich, K. (1986). NOE observable 1H-1H distances in proteins. In *NMR of proteins and nucleic acids* 1 edit. (Wuthrich, K., ed.), pp. 117-129. Wiley-Interscience publication, U.S.A.

165. Dukor, R. K. & Keiderling, T. A. (1991). Reassessment of the random coil conformation: vibrational CD study of proline oligopeptides and related polypeptides. *Biopolymers* 31, 1747-1761.

166. Park, S. H., Shalongo, W. & Stellwagen, E. (1997). The role of PII conformations in the calculation of peptide fractional helix content. *Protein Sci* 6, 1694-1700.

167. Tiffany, M. L. & Krimm, S. (1968). New chain conformations of poly(glutamic acid) and polylysine. *Biopolymers* 6, 1379-1382.

168. Tiffany, M. L. & Krimm, S. (1972). Effect of temperature on the circular dichroism spectra of polypeptides in the extended state. *Biopolymers* 11, 2309-2316.

169. Wilson, G., Hecht, L. & Barron, L. D. (1996). Residual structure in unfolded proteins revealed by Raman optical activity. *Biochemistry* 35, 12518-12525.

170. Anderson, A. G. & Hermans, J. (1988). Microfolding: conformational probability map for the alanine dipeptide in water from molecular dynamics simulations. *Proteins* 3, 262-265.

171. Markley, J. L., Meadows, D. H. & Jardetzky, O. (1967). Nuclear magnetic resonance studies of helix-coil transitions in polyamino acids. *J Mol Biol* 27, 25-40.

172. Nakamura, A. & Jardetzky, O. (1967). Systematic analysis of chemical shifts in the nuclea magnetic resonance spectra of Peptide chains, I. Glycine-containing dipeptides. *Proc Natl Acad Sci U S A* 58, 2212-2219.

173. Nakamura, A. & Jardetzky, O. (1968). Systematic analysis of chemical shifts in the nuclear magnetic resonance spectra of peptide chains. II. Oligoglycines. *Biochemistry* 7, 1226-1230.

174. De Simone, A., Cavalli, A., Hsu, S. T., Vranken, W. & Vendruscolo, M. (2009). Accurate random coil chemical shifts from an analysis of loop regions in native states of proteins. *J Am Chem Soc* 131, 16332-16333.

175. Ries-Kautt, M. M. & Ducruix, A. F. (1989). Relative effectiveness of various ions on the solubility and crystal growth of lysozyme. *J Biol Chem* 264, 745-748.

176. Rosenow, M. A., Magee, C. L., Williams, J. C. & Allen, J. P. (2002). The influence of detergents on the solubility of membrane proteins. *Acta Crystallogr D Biol Crystallogr* 58, 2076-2081.

177. Trevino, S. R., Scholtz, J. M. & Pace, C. N. (2007). Amino acid contribution to protein solubility: Asp, Glu, and Ser contribute more favorably than the other hydrophilic amino acids in RNase Sa. *J Mol Biol* 366, 449-460.

178. Hayakawa, T., Ito, T., Wakamatsu, J., Nishimura, T. & Hattori, A. (2010). Myosin filament depolymerizes in a low ionic strength solution containing L-histidine. *Meat Sci* 84, 742-746.

179. Muriaux, D., Girard, P. M., Bonnet-Mathoniere, B. & Paoletti, J. (1995). Dimerization of HIV-1Lai RNA at low ionic strength. An autocomplementary sequence in the 5' leader region is evidenced by an antisense oligonucleotide. *J Biol Chem* 270, 8209-8216.

180. Zhao, Y., Yang, X., Li, X., Bu, Y., Deng, P., Zhang, C., Feng, J., Xie, Y., Zhu, S., Yuan, H., Yu, M. & Liao, F. (2009). Reversible inactivation of an intracellular uricase from Bacillus fastidiosus via dissociation of homotetramer into homodimers in solutions of low ionic strength. *Biosci Biotechnol Biochem* 73, 2141-2144.

181. Harris, L. J., Skaletsky, E. & McPherson, A. (1995). Crystallization of intact monoclonal antibodies. *Proteins* 23, 285-289.

182. Sanishvili, R., Volz, K. W., Westbrook, E. M. & Margoliash, E. (1995). The low ionic strength crystal structure of horse cytochrome c at 2.1 A resolution and comparison with its high ionic strength counterpart. *Structure* 3, 707-716.

183. Walter, M. H., Westbrook, E. M., Tykodi, S., Uhm, A. M. & Margoliash, E. (1990). Crystallization of tuna ferricytochrome c at low ionic strength. *J Biol Chem* 265, 4177-4180.

184. Wong, K. B., Wang, W. K., Proctor, M. R., Bycroft, M. & Chen, Y. W. (2001). Crystallization and preliminary crystallographic studies of a ribosomal protein L30e from the hyperthermophilic archaeon Thermococcus celer. *Acta Crystallogr D Biol Crystallogr* 57, 865-866.

185. Pjura, P. E., Lenhoff, A. M., Leonard, S. A. & Gittis, A. G. (2000). Protein crystallization by design: chymotrypsinogen without precipitants. *J Mol Biol*

300, 235-239.

186. Takahashi, T. (1997). Significant role of electrostatic interactions for stabilization of protein assemblies. *Adv Biophys* 34, 41-54.

187. Takahashi, T., Endo, S. & Nagayama, K. (1993). Stabilization of protein crystals by electrostatic interactions as revealed by a numerical approach. *J Mol Biol* 234, 421-432.

188. Longenecker, K. L., Garrard, S. M., Sheffield, P. J. & Derewenda, Z. S. (2001). Protein crystallization by rational mutagenesis of surface residues: Lys to Ala mutations promote crystallization of RhoGDI. *Acta Crystallogr D Biol Crystallogr* 57, 679-688.

189. Mateja, A., Devedjiev, Y., Krowarsch, D., Longenecker, K., Dauter, Z., Otlewski, J. & Derewenda, Z. S. (2002). The impact of Glu-->Ala and Glu-->Asp mutations on the crystallization properties of RhoGDI: the structure of RhoGDI at 1.3 A resolution. *Acta Crystallogr D Biol Crystallogr* 58, 1983-1991.

190. Chen, Y. W., Bycroft, M. & Wong, K. B. (2003). Crystal structure of ribosomal protein L30e from the extreme thermophile Thermococcus celer: thermal stability and RNA binding. *Biochemistry* 42, 2857-2865.

191. Wong, K. B., Lee, C. F., Chan, S. H., Leung, T. Y., Chen, Y. W. & Bycroft, M. (2003). Solution structure and thermal stability of ribosomal protein L30e from hyperthermophilic archaeon Thermococcus celer. *Protein Sci* 12, 1483-1495.

192. Dasgupta, S., Iyer, G. H., Bryant, S. H., Lawrence, C. E. & Bell, J. A. (1997). Extent and nature of contacts between protein molecules in crystal lattices and between subunits of protein oligomers. *Proteins* 28, 494-514.

193. Czepas, J., Devedjiev, Y., Krowarsch, D., Derewenda, U., Otlewski, J. & Derewenda, Z. S. (2004). The impact of Lys-->Arg surface mutations on the crystallization of the globular domain of RhoGDI. *Acta Crystallogr D Biol Crystallogr* 60, 275-280.

194. Doig, A. J. & Sternberg, M. J. (1995). Side-chain conformational entropy in protein folding. *Protein Sci* 4, 2247-2251.

195. Howe, P. W. (2004). A straight-forward method of optimising protein solubility for NMR. *J Biomol NMR* 30, 283-286.

196. Lange, C., Luque, I., Hervas, M., Ruiz-Sanz, J., Mateo, P. L. & De la Rosa, M. A. (2005). Role of the surface charges D72 and K8 in the function and structural stability of the cytochrome c from Nostoc sp. PCC 7119. *FEBS J* 272, 3317-3327.

197. Chivers, P. T., Prehoda, K. E., Volkman, B. F., Kim, B. M., Markley, J. L. &

Raines, R. T. (1997). Microscopic pKa values of Escherichia coli thioredoxin. *Biochemistry* 36, 14985-14991.

198. Garcia-Moreno, E. B. & Fitch, C. A. (2004). Structural interpretation of pH and salt-dependent processes in proteins with computational methods. *Methods Enzymol* 380, 20-51.

199. McIntosh, L. P., Hand, G., Johnson, P. E., Joshi, M. D., Korner, M., Plesniak, L. A., Ziser, L., Wakarchuk, W. W. & Withers, S. G. (1996). The pKa of the general acid/base carboxyl group of a glycosidase cycles during catalysis: a 13C-NMR study of bacillus circulans xylanase. *Biochemistry* 35, 9958-9966.

200. Onufriev, A., Case, D. A. & Ullmann, G. M. (2001). A novel view of pH titration in biomolecules. *Biochemistry* 40, 3413-3419.

201. Kay, L. E., Xu, G. Y., Singer, A. U., Muhandiram, D. R. & Forman, J. D. (1993). A gradient-enhanced HCCH-TOCSY experiment for recording side-chain H-1 and C-13 correlations in H2O samples of protein. *J Magn Reson Ser B* 101, 333-337.

202. Kay, L. E., Ikura, M., Tschudin, R. & Bax, A. (1990). Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *J Magn Reson* 89, 496-514.

203. Oda, Y., Yamazaki, T., Nagayama, K., Kanaya, S., Kuroda, Y. & Nakamura, H. (1994). Individual ionization constants of all the carboxyl groups in ribonuclease HI from Escherichia coli determined by NMR. *Biochemistry* 33, 5275-5284.

204. Shindo, H. & Cohen, J. S. (1976). Observation of individual carboxyl groups in hen egg-white lysozyme by use of high field 13C-nuclear magnetic resonance. *Proc Natl Acad Sci U S A* 73, 1979-1983.

205. Spitzner, N., Lohr, F., Pfeiffer, S., Koumanov, A., Karshikoff, A. & Ruterjans, H. (2001). Ionization properties of titratable groups in ribonuclease T1. I. pKa values in the native state determined by two-dimensional heteronuclear NMR spectroscopy. *Eur Biophys J* 30, 186-197.

206. Yamazaki, T., Nicholson, L. K., Torchia, D. A., Stahl, S. J., Kaufman, J. D., Wingfield, P. T., Domaille, P. J. & Campbell-Burk, S. (1994). Secondary structure and signal assignments of human-immunodeficiency-virus-1 protease complexed to a novel, structure-based inhibitor. *Eur J Biochem* 219, 707-712.

207. Pujato, M., Navarro, A., Versace, R., Mancusso, R., Ghose, R. & Tasayco, M. L. (2006). The pH-dependence of amide chemical shift of Asp/Glu reflects its pKa in intrinsically disordered proteins with only local interactions. *Biochim Biophys Acta* 1764, 1227-1233.

208. Edsall, J. T., Martin, R. B. & Hollingworth, B. R. (1958). Ionization of Individual Groups in Dibasic Acids, with Application to the Amino and Hydroxyl Groups of Tyrosine. *Proc Natl Acad Sci U S A* 44, 505-518.

209. Klingen, A. R., Bombarda, E. & Ullmann, G. M. (2006). Theoretical investigation of the behavior of titratable groups in proteins. *Photochem Photobiol Sci* 5, 588-596.

210. Sanchez-Ruiz, J. M. & Makhatadze, G. I. (2001). To charge or not to charge? *Trends Biotechnol* 19, 132-135.

211. Gribenko, A. V., Patel, M. M., Liu, J., McCallum, S. A., Wang, C. & Makhatadze, G. I. (2009). Rational stabilization of enzymes by computational redesign of surface charge-charge interactions. *Proc Natl Acad Sci U S A* 106, 2601-2606.

212. Schweiker, K. L., Zarrine-Afsar, A., Davidson, A. R. & Makhatadze, G. I. (2007). Computational design of the Fyn SH3 domain with increased stability through optimization of surface charge charge interactions. *Protein Sci* 16, 2694-2702.