Linking computation and experiments to study the role of charge-charge interactions in protein folding and stability

To cite this article: George I Makhatadze 2017 Phys. Biol. 14 013002

View the article online for updates and enhancements.

PERSPECTIVE

Linking computation and experiments to study the role of charge–charge interactions in protein folding and stability

George I Makhatadze
Center for Biotechnology and Interdisciplinary Studies, and Department of Biological Sciences, Department of Chemistry and Chemical Biology, Rensselaer Polytechnic Institute, 110 8th Street, Troy, NY 12180 USA
E-mail: makhag@rpi.edu

Keywords: protein folding, charge–charge interactions, protein stability

Abstract
Over the past two decades there has been an increase in appreciation for the role of surface charge–charge interactions in protein folding and stability. The perception shifted from the belief that charge–charge interactions are not important for protein folding and stability to the near quantitative understanding of how these interactions shape the folding energy landscape. This led to the ability of computational approaches to rationally redesign surface charge–charge interactions to modulate thermodynamic properties of proteins. Here we summarize our progress in understanding the role of charge–charge interactions for protein stability using examples drawn from my own laboratory and touch upon unanswered questions.

1. Introduction
The energy landscape theory provides a conceptual physico-chemical framework for understanding protein folding. This theory is based on the principle of minimal frustration that ‘...quantifies the dominance of interactions stabilizing the specific native structure over other interactions that would favor nonnative, topologically distinct traps’ [1]. A consequence of this is that the folding energy landscape of naturally occurring proteins is funnel-shaped [1–23]. The shape of this funnel depends on two main factors that can introduce frustration and roughness: topology and the extent of non-native interactions. Topological frustration can occur when native interactions are formed too early and need to be undone leading to backtracking and/or cracking [24–28]. Weak non-native interactions can have a complex effect on the folding landscape [29]: a small amount of weak non-native interactions can assist folding, while larger amounts can create internal friction that will slow folding [30–34]. Strong non-native interactions can lead to formation of intermediate states that increase the roughness of the folding funnel [4, 22, 35, 36]. Understanding in more detail how different types of interactions affect the energy landscape presents one of the current challenges of the protein folding field [1, 2].

The role of interactions between ionizable residues in shaping the protein folding landscape is of particular interest [37]. Ionizable residues are readily present in protein sequences, constituting on average 25% of the total amino acid composition [38]. Ionizable residues are largely located on the exterior of globular proteins where they have favorable interactions with the solvent water. Only a small fraction of ionizable residues can be found in the protein interior. However, in the vast majority of these cases the buried ionizable residues form hydrogen bonds and/or salt bridges with other polar or ionizable residues [39]. Interactions between charges can be both attractive (between unlike charges) and repulsive (between like charges). The strength of the charge–charge interaction depends on the distance as 1/r, which makes them, at short distances, comparable in strength with Lennard–Jones (LJ) interactions. However, the charge–charge interactions are also long-range and they do not decay as quickly as LJ with the increase in distance [40–44]. All this makes charge–charge interactions an important contributor not only to protein stability, but also to the overall behavior of polypeptide sequences as heteropolymers [45, 46].

Here we summarize the progress in understanding the role of charge–charge interactions for protein stability and touch upon unanswered questions.
2. Effect of charge–charge interactions on protein stability

An important step to the solution of the protein folding problem is the detailed understanding of the balance of forces that stabilize the native protein structure [47–52]. It is generally believed that proteins are stabilized by the hydrophobic effect, hydrogen bonding, and packing (van der Waals) interactions. The large favorable contribution of these forces is offset by a large unfavorable entropy change on protein folding. The resulting stability of the protein, as measured by the Gibbs energy change \( \Delta G^\circ \), is only on the order of few tens of kilojoules. One type of interaction, the electrostatic interactions between ionizable residues, has not received wide recognition as an important factor for protein stability. It was believed that surface residues are equally well exposed to the solvent in both the native and unfolded states and thus, should not contribute to protein stability.

The first comprehensive study of the importance of surface charge–charge interactions for protein stability was demonstrated on the ubiquitin molecule [53]. The study used Tanford–Kirkwood methodology to identify the contribution of individual charged residues to the overall stability of the protein [54, 55]. These residues were then substituted to neutral or oppositely charged residues and their stabilities were measured experimentally [53]. Comparison of experimentally measured changes in the stability of the variants relative to the wild type ubiquitin supported the results of the calculations. In particular, charge-neutralization of the amino acid residues that were predicted to contribute unfavorably to protein stability indeed led to an increase in the stability of corresponding ubiquitin variants. Furthermore, substitution to an opposite charge led to even higher increase in stability of the variants than charge-neutralization, again in agreement with the computational predictions. Importantly, substitutions to oppositely charged or charge-neutral amino acids at positions predicted to have a favorable contribution to the protein stability produced proteins that were less stable than the wild type. Finally, substitutions at surface positions that were predicted not to contribute to the protein stability due to the charge–charge interaction were found to have stability similar to the wild type ubiquitin. The effects of single site substitutions on protein stability were subsequently tested on over 15 different proteins that varied in size, secondary structure content and overall topology [56–67]. In all cases it was found that computational modeling can qualitatively predict the effects of substitutions at surface charge positions on the proteins stability.

Success in predicting single-site substitutions allowed further extension of the computational model to the analysis of the contribution of salt bridges to protein stability. Salt bridges, which can be ionic or neutral are defined as pairs of potentially negatively (OD1 and OD2 for Asp and OE1 and OE2 for Glu) and positively charged (NZ for Lys, NH1, and NH2 for Arg) atoms within 4 Å [68]. The ubiquitin molecule contains a surface salt bridge between residues K11 and E34. The strength of this salt bridge and also of the inverted E11/K34 salt bridge was measured experimentally using the double mutant cycle [69]. It was found that the net strength of these salt bridges is the same 3.8 kJ mol\(^{-1}\), however the protein containing the K11/E34 salt bridge is 2.2 kJ mol\(^{-1}\) more stable than the protein containing the inverse E11/K34 salt bridge. Computational modeling suggests that the overall charge–charge interactions are more favorable for the K11/E34 pair than for E11/K34 again supporting the notion that the overall charge–charge interactions are important for modulating global protein stability [69]. Such context-dependence of the distribution of charge–charge interactions can occur at different separation in sequence. Even two charged residues that are adjacent in sequence can have different contributions to the global protein stability depending on the nature of the charged groups (same or opposite) and the overall electrostatic environment where such sequences are found [64, 65].

Further support for the importance of charge–charge interactions in modulating global protein stability came from the analysis of the cold shock family of proteins [70–72]. Cold shock proteins from thermophilic organisms are more stable than the corresponding mesophilic homologs. Computational modeling of the charge–charge interactions suggested that thermophilic proteins have better optimized interaction relative to their mesophilic counterparts [56, 61, 69, 73–75]. To test this experimentally, a hybrid protein was created that included the sequence of the mesophilic protein with surface charge distribution of that of the thermophilic protein. Experimental measurements of stability showed that the hybrid protein is \( >20 ^\circ C \) more stable that the mesophilic one supporting the notion that a large part of the thermophilic adaptation in this protein family came from the optimization of charge–charge interactions on the protein surface [69, 75]. Importantly, both proteins retain the ability to bind single stranded nucleic acids, which is their natural function as RNA-chaperones [76].

The importance of charge–charge interactions for stability is not limited to globular proteins [77–82]. The sequence of the fibrillar protein collagen is characterized by the YGX repeat, where X is frequently proline and Y is hydroxyproline (three letter abbreviation Hyp, one letter O). Hyp residues are the result of post-translational modification by the enzyme prolyl hydroxylase found exclusively in eukaryotes. On average 20% of the collagen sequence contains OGP triplets. Relative Hyp content correlates with collagen stability. Importantly, 40% of YGX triplets contain at least one ionizable residue, and 20% of residues in both X and Y positions contain ionizable side chains.
[94, 100–105]. Computational modeling showed that KGE, EGK, KGD, DGK containing sequences have very different stability [82]. These differences are largely associated with a different degree of direct or water-mediated salt bridges between oppositely charged residues. Since prokaryotes do not have prolyl hydroxylase activity, it appears that the stability of bacterial collagens can be exclusively modulated by charge–charge interactions.

3. Computational design of stable proteins: thermodynamics and kinetics

The ability of computational models to predict the effects of substitutions of the surface charged residues on the stability of several proteins led to the proposal that optimization of the surface charge–charge interactions can be a viable strategy for enhancing protein stability. To experimentally demonstrate this, the design algorithm was applied to seven different proteins: ubiquitin, the activation domain of human procarboxypeptidase A2, the fibronectin type III domain of human tenascin, the N-terminal RNA-binding domain of human U1A protein, Fyn SH3 domain, human acylphosphatase, and CDC42 GTPase [62, 83–85]. These proteins differ in size (from 59 to 198 amino acid residues) and in tertiary fold topology. Importantly, the acylphosphatase (ACP) and CDC42 are enzymes which allow for direct testing of how the design affects the enzymatic activity. An identical computational protocol was applied to these seven different proteins and the resulting designed sequences were subjected to the stability measurements. In all cases the designed proteins that had optimized charge–charge interactions were more stable than the corresponding wild type proteins. Importantly, the designed proteins remain enzymatically active. In the case of CDC42 in addition to unperturbed GTPase activity, the interactions with cellular activator protein CDC42GAP also remained unchanged [85]. The substitutions in ACP and CDC42 were analyzed in terms of evolutionary conservation. In the majority of cases it was found that the computational algorithm selected substitutions from a more frequent at a given position in the sequence to a less frequent one [85, 86]. This suggests that the optimization of charge–charge interactions on the protein surface identifies residue other than a consensus–design approach that uses evolutionary conservation as a means to engineer more stable proteins.

The demonstration of a significant stabilization of proteins with substitutions of the surface charged residues posed a question about the significance of the contribution of the surface charges to the stability and folding of globular proteins. A direct experiment to answer this question requires measurement of the stability of a protein without any charges [87, 88]. This is not easy because a protein without charges is expected to have very low solubility, which makes overexpression in bacteria difficult. This was overcome by expressing and purifying proteins that contained charges and then using specific chemical modifications to eliminate some of the charges. Experiments show that a protein without charges folds and has stability similar to the charged protein [87]. Moreover, the folding rate of a protein without charges is faster than that of the charged protein [88]. This suggests that protein folding is driven primarily by hydrophobic packing and hydrogen bonding, whereas surface charges are important for protein solubility [89], enzymatic activity [90–93] and for protein interactions [94–98].

How is the folding energy landscape modulated by charge–charge interactions? To answer this question the effects of charge–charge interactions on the kinetics of folding were analyzed for 5 protein pairs [99,100]. Each pair included a wild type protein and a designed protein for which the stability was increased by redesigning the charge–charge interactions. It was found experimentally that in all cases the protein with optimized charge–charge interactions folds faster than the corresponding wild type protein while the unfolding rates remain largely unchanged. To understand the detailed mechanism of this increase in folding rates, molecular dynamics simulations using structure-based models (SBM) were used. SBMs are based on the principle of minimal frustration and have been proven to be valuable in understanding the protein folding energy landscape (see e.g. [1–23, 30–36, 101, 102]). The original SBM was supplemented with electrostatic interactions based on the Debye–Hückel formalism. Initial simulations were done using Ca-SBM and later extended using all-atom SBM [24, 86, 99, 100, 103]. Simulations qualitatively recapitulate both thermodynamic and kinetics differences observed experimentally suggesting the validity of the model. This in turn allows detailed analysis of the effect of charge–charge interactions on the folding energy landscape of the studied proteins. The analyses suggest that wild type proteins have frustrated folding landscapes due to energetic and topological frustration. Optimization of the surface charge–charge interactions seems to remove some of that frustration. This leads to an increase of native-like contacts in the transition-state ensembles and provides a less frustrated energy landscape between the unfolded and TS ensembles. Such less frustrated energy landscapes for optimized variants results in faster folding rates as demonstrated by the experimental measurements [99].

4. Summary and outlook

The results presented here show major milestones in establishing the importance of surface charge–charge interactions for protein folding and stability. It is now clear that these interactions were not given sufficient attention in the past couple of decades. Ionizable
residues on the protein surface are not there to only provide solubility. They are important in modulating thermodynamic stability of proteins. In addition, interactions between ionizable residues are long-range and thus can alter the conformational ensemble at every step of the folding/unfolding process. More significantly interactions between ionizable residues are nonspecific and can be either attractive or repulsive. Therefore, their potential effects on the folding energy landscape can be highly complex. So far the rational design optimizing interactions between surface charges has been applied to engineer more thermodynamically stable proteins. Can the protein folding/unfolding kinetics be modulated in a rational way? For example, is it possible to design proteins in such way that changes in stability are due to the unfolding rates? Considering that for a given topological fold there is probably a maximum limit of stability, a significant decrease in the unfolding rate might be a useful strategy to design kinetically stable proteins. It is well established that electrostatic interactions modulate the thermodynamics of protein–protein and protein ligand interactions [37, 93, 94, 104, 105]. How do electrostatic interactions modulate the kinetics of these interactions? Again can such interactions be understood in sufficient detail so that the rational modulation of energy landscape, macroscopically manifested in terms of $K_{on}$ and $K_{off}$ will be possible? Solving such complex problems requires a multidisciplinary approach that combines experimental biochemistry and biophysics and extensive computational modeling. Armed with comprehensive theoretical background and having large arsenal of computational and experimental tools these goals are very attainable, realistic and timely.

Acknowledgments

I dedicate this paper to Dr Kamal Shukla who will be remembered for his remarkable contributions to the molecular biophysics community for many years to come. This work was supported by grant MCB-1330249 from the US National Science Foundation.

References

[42] Lee K K, Fitch C A and Garcia-Moreno E B 2002 Distance dependence and salt sensitivity of pairwise, coulombic interactions in a protein Protein Sci. 11 1004–16
[51] Fersht A R 1993 Protein-folding and stability—the pathway of folding of barnase FEBS Lett. 325 5–16
[56] Sanchez-Ruiz J M and Makhatadze G I 2001 To charge or not to charge! Trends Biotechnol. 19
[65] Lodalde V V and Makhatadze G I 2011 Energetics of charge–charge interactions between residues adjacent in sequence Proteins 79 5494–9
protein stability; guidelines for protein engineering J. Mol. Biol. 327 1135–48
[71] Lopez M M, Yutani K and Makhatadze G I 1999 Interactions of the major cold shock protein of Bacillus subtilis CspB with single-stranded DNA templates of different base composition J. Biol. Chem. 274 33601–8
[72] Lopez M M, Yutani K and Makhatadze G I 2001 Interactions of the cold shock protein CspB from Bacillus subtilis with single-stranded DNA—importance of the T base content and position within the template J. Biol. Chem. 276 15311–8
[77] Krylov D, Barch J and Vinson C 1998 Inter-helical interactions in the leucine zipper coiled coil dimer: pH and salt dependence of coupling energy between charged amino acids J. Mol. Biol. 279 959–72
[78] Martí D N, Jeleasov I and Bosshard H R 2000 Interhelical ion pairing in coiled coils solution structure of a heterodimeric leucine zipper and determination of pKa values of Cua side chains Biochemistry 39 12804–18
[87] Loladze V V and Makhatadze G I 2002 Removal of surface charge–charge interactions from ubiquitin leaves the protein folded and very stable Protein Sci. 11 174–7
[90] Huang Y M, Huber G and McCammon J A 2015 Electrostatic steering enhances the rate of cAMP binding to phosphodiesterase: brownian dynamics modeling Protein Sci. 24 1884–9
[94] Lee L P and Tidor B 2001 Barstar is electrostatically optimized for tight binding to barnase Nat. Struct. Biol. 8 73–6