SALT BRIDGES AND CONFORMATIONAL FLEXIBILITY: EFFECT ON PROTEIN STABILITY

A. Karshikoff 1 and I. Jelesarov 2
Karolinska Institute, Department of Biosciences and Nutrition, Stockholm, Sweden
University of Zurich, Department of Biochemistry, Switzerland
Correspondence to: Andrey Karshikoff
E-mail: aka@csb.ki.se

ABSTRACT
Salt bridges are believed to have an important role in stabilisation of native protein structure. The assessment of their contribution to the electrostatic term of the free energy is a yet unsolved task. One can point out a number of reasons: beginning with conceptual issues, such as the complex nature of the interplay between the different types of non-covalent interactions and going to details, such as the interactions of the participating groups with their environments. Here we focus on the interplay between electrostatic interactions the functional groups forming salt bridges are involved in and the conformational flexibility of the protein molecule. We show that the connection between these two factors appears to be one of the keys for a better understanding of the forces determining stability of proteins.

Keywords: proteins, modelling, stability, electrostatic interactions, salt bridges

Introduction
The tree-dimensional structure of the native proteins is result of a very delicate balance between different types of non-covalent interactions, uniquely determined by the sequence of the polypeptide chain. The physical principles non-covalent interactions obey are well known and ubiquitous in proteins. This, however, does not simplify the task of predicting their role in protein stability and functional properties. The complex interplay of non-covalent interactions causes sometimes surprising behaviour of proteins. In the context of the considerations presented in this paper an instructive example are the charge reversal mutations, which are expected to improve binding of charged substrates. The failure to increase binding of, say, a negatively charged substrate by mutating a negative group in the binding side to a positive one have been clearly explained on the basis of theoretical calculations [10]. This example also illustrates that rational design aiming at improving or changing properties must be preceded by a careful analysis of the mutual dependence of non-covalent interactions. This point is especially important when electrostatic interactions are in the limelight. Two reasons can be given. First, electrostatic interactions are long-range interactions, a feature distinguishing them from all other types of non-covalent interactions. Hence, charges can “send signals” to distant parts of the protein molecule. Second, electrostatic interactions take place in an inhomogeneous dielectric medium, since the dielectric permittivity of the protein globule is non-isotropic and differs by much from the dielectric permittivity of water. This fact turns out to be the most serious difficulty when electrostatic interactions are to be predicted.

In the following we consider one aspect of the connection between electrostatic interactions and protein functional properties, namely the role of salt bridges in protein stability.

Salt Bridges
In proteins, when two oppositely charged functional groups are close to each other, one speaks about the existence of a salt bridge. For the purposes of our considerations the term functional group is defined as a group of atoms, which change its charge upon protonation (or deprotonation). Hence, these groups are also ionisable groups. The distance at which two such groups form a salt bridge is usually taken to be r = 4 Å, a value deduced by Barlow and Thornton [4] on the basis of the analysis of 38 protein structures. This criterion was found to be most appropriate for the purpose of statistical analyses of ion pairs in proteins. Even using a limited data base, these authors pointed out that the highest peak of separation distribution is at r = 3 Å. This value of r is close to the average distance between the proton donor and the proton acceptor atoms in a hydrogen bond. It is clear that the physical properties of an ion pair connected by a hydrogen bond differ from those of a pair separated by, say, 4 Å. Hydrogen bonds have directionality, a feature that is often crucial for enzymatic activity, ligand binding, protein-protein interactions, etc. To emphasize the difference we define salt bridge as a couple of functional groups connected by hydrogen bond(s). As a distance criterion we chose 3.2 Å, approximately the upper limit of the proton donor-acceptor distance in a hydrogen bond. On the other hand, functional groups which are in close proximity, but do not satisfy the above distance criterion we classify as ion pairs. The chosen distance criterion neither revises the conclusions based on other distance criteria nor introduces a new concept. It merely reflects the nature of the physical phenomena that are subject of this investigation. This can be seen in Table 1, where the hydrogen bond lengths of the most common hydrogen
bonds found in proteins are compared. As seen, the hydrogen bond length of salt bridges (charged functional groups) does not differ significantly from hydrogen bonds where one or both partners are neutral.

All hydrogen bonds listed in Table 1 belong to the class of the moderate hydrogen bonds. The energy of formation of moderate hydrogen bonds is between $-15$ and $-4$ kcal/mol. Hence, when considering a salt bridge as a hydrogen bond, one should formally expect a significant stabilising contribution. Surprisingly, the range of statements about the contribution of salt bridges to protein stability spans from strong and moderate to negligible [19, 25], and even destabilizing [18]. Quantitatively, the stabilising contribution of individual salt bridges has been evaluated to be between $-0.5$ kcal/mol [17] and $-3$ to $-5$ kcal/mol [1]. It is interesting to know which are the factors determining the large variation in the energetic contribution of salt bridges. The answer of this question has not only an academic value. Such knowledge is essential for the purposes of protein modelling and protein re-design by rational mutagenesis.

A few main factors modulating the energetics of salt bridges can be mentioned. The first one is the desolvation of the charged functional groups participating in a salt bridge. In the less polar medium of the protein molecule the energy of charging increases. Therefore, the ionisable groups tend to adopt their neutral forms. In extreme cases the uncharged form of ionisable group which is completely buried in the protein (that is, the group is completely desolvated) can be favoured by much in comparison to the completely solvated state [24]. Hence, the desolvation of functional groups reduces their ability to participate in salt bridges. This explains why salt bridges are usually located on the surface of the protein molecule. The charge-charge interactions are the second important factor. Indeed, the close proximity of oppositely charged groups results in a favourable interaction energy, which opposes the unfavourable desolvation effect. The sum of these two factors we refer to as electrostatic interactions term. The third factor is the flexibility of the protein molecule. This inherent property of proteins may facilitate a dynamic process of formation and disruption of salt bridges. In the following we will show that electrostatic interactions are strongly influenced by the conformational flexibility. In fact, the interplay between electrostatics and conformational flexibility determines the electrostatic term of the unfolding free energy of proteins.

### Electrostatic Term of the Unfolding Free Energy

By definition the unfolding free energy, $\Delta G^u$, is given by

$$\Delta G^u = \Delta G^f - \Delta G^r,$$

where $\Delta G^u$ and $\Delta G^r$ denote the free energy of the unfolded and of the folded state, respectively. These free energies are

<table>
<thead>
<tr>
<th>ACCEPTOR</th>
<th>DONOR</th>
<th>$R_{(A\cdots H)}$</th>
<th>ACCEPTOR</th>
<th>DONOR</th>
<th>$R_{(A\cdots H)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>H–N</td>
<td>H–N</td>
<td>1.96 – 2.00</td>
<td>O</td>
<td>C</td>
<td>1.93</td>
</tr>
<tr>
<td>C=O</td>
<td>H+N</td>
<td>1.99</td>
<td>O</td>
<td>C</td>
<td>1.89</td>
</tr>
<tr>
<td>C=O</td>
<td>H+R</td>
<td>1.87 – 1.94</td>
<td>C</td>
<td>O</td>
<td>1.84</td>
</tr>
<tr>
<td>C=O</td>
<td>H+R</td>
<td>1.79 – 1.97</td>
<td>C</td>
<td>O</td>
<td>1.80</td>
</tr>
<tr>
<td>C=O</td>
<td>HO</td>
<td>1.64 – 1.70</td>
<td>C</td>
<td>O</td>
<td>1.40</td>
</tr>
</tbody>
</table>

Mean hydrogen bond length, $r_{(A\cdots H)}$, in Å, (A indicates the proton acceptor atom). For some hydrogen bonds $r_{(A\cdots H)}$ values are given within intervals illustrating the variability of this parameter. Data taken from [11].
related to some appropriate reference state. Since the free energy is defined up to an additive constant, we can choose as the reference point the state where the free energies of the unfolded and of the folded state are equal. In this way $\Delta G^o$ can also be expressed in the following way

$$\Delta G^o = RT \ln \frac{Z^F}{Z^U}.$$  

$Z^U$ and $Z^F$ are the partition functions of the system in the unfolded and folded states of the protein, respectively. If we know the partition functions, $Z^{U,F}$, the unfolding free energy can be calculated. In order to extract the electrostatic term of $\Delta G^o$ we will assume an unfolding process in which only the electrostatic interactions are relevant. A physical phenomenon closely to this hypothetical process is the pH-dependent unfolding. If one ignores all, but electrostatic interactions, for the closest to this hypothetical process is the pH-dependent unfolding. Since the free energies of the unfolded and of the folded state are equal. In this way $\Delta G^o$ can be calculated. In order to extract the electrostatic term of the system and $\Delta G^o$ we will assume an unfolding process in which only the electrostatic interactions are relevant. A physical phenomenon closely to this hypothetical process is the pH-dependent unfolding. If one ignores all, but electrostatic interactions, for the closest to this hypothetical process is the pH-dependent unfolding. Since the free energies of the unfolded and of the folded state are equal.

The term $\Delta G^o$ describes the microscopic states of the system, $\{x\}$, is then $2^N$ which is an astronomical figure even for a protein of medium size. The methods for calculation of the sum in Eq. (1), as well as of the quantities in Eqs. (2 and 3) are out of the scope of our discussion. A description of the theory and the computational methodologies is given in [13].

Eq. (2) is also valid for the unfolded state. The assumption of a fixed structure is, however, not applicable in this case. Since the unfolded polypeptide chain adopts any arbitrary conformation, the positions of the ionisable groups cannot be defined. In order to take this into account, the vector $x$, describing the microscopic states of the system, should include the variation of the spatial coordinates of the ionisable groups: $x = (x_1, x_2, \ldots, x_N)$. The dipole charges corresponding to the spatial coordinates of the ionisable groups. A solution of Eq. (2) with $x$ defined as above can be found by means of Monte Carlo simulation [13].

The theoretical outline presented above aims at showing the basic assumption made when the electrostatic term of the free energy, $\Delta G^e$, is of interest. These assumptions introduce limitations, which should be kept in mind in the interpretation of experimental data. One of these limitations has already been mentioned: the fixed structure of the native protein molecule. Another important assumption is the additivity of the different terms of the free energy, i.e. the change of the electrostatic term of the free energy of unfolding does not influence the other terms and vice versa.

We shall skip the mathematical rearrangements and present the final expression for $\Delta G^e$ used for calculations:

$$\Delta G^e(pH) = \int_{pH_1}^{pH_2} \left( U^F(pH) - U^F(pH) \right) d\text{pH}.$$  

Comprehensive descriptions of the ideology and deduction of Eq. (4) can be found in the literature [13, 15, 26]. The quantities $U^F$ are the average number of protons bound to the protein molecule in the unfolded and in the folded state, respectively:

$$U^F(pH) = \sum_{i=1}^{N} \theta_i(pH),$$  

where $\theta_i(pH)$ are the degree of protonation of ionisable group $i$ in the two states. As it can be seen, in order to solve Eq. (4) one needs to know the titration curves of the folded and of the unfolded states of the protein. Experimentally, it is not straightforward to measure the titration curves of the two states at identical conditions and within a given pH interval. However, on the basis of the theory presented above it is easy to calculate them because

$$\theta_i(pH) = \frac{\sum_x x_i e^{-W(x)/RT}}{\sum_x e^{-W(x)/RT}},$$  

where $W(x)$ is given by Eq. (2).

The lower limit of the integral in Eq. (4) is the reference pH$_1$ value where the two states of the protein have identical charge.
protonation states: \( \nu^\alpha(pH) - \nu^\alpha(pH) = 0 \). Obviously, \( \Delta G^{\nu\alpha}(pH) = 0 \). In general, different reference states can be chosen. For instance, Yang and Honig [26] have chosen pH equal to 0, whereas Langella et al. [16] defined it as an abstract state, at which all titratable groups in the protein are in their neutral forms.

The pH dependence of \( \Delta G^\nu \) of the homotrimeric coiled coil protein (Lpp-56) shown in Fig. 1 illustrates the assumptions made above. At the reference pH \(_0\) (-14), \( \Delta G^\nu = \Delta G^\nu = 0 \). Hence, the values of the absolute \( \Delta G^\nu \) scaled on the abscise have no physical meaning. The pH-dependence of \( \Delta G^\nu \) is however expected to be correctly predicted. To verify this, at least two experimental values of \( \Delta G^\nu \) are needed. (Note that we use \( \Delta G^\nu \) firstly, because \( \Delta G^\alpha \) cannot be measured directly, and secondly because we have assumed that at the unfolding process only electrostatic interactions are relevant, i.e. \( \Delta G^\nu = \Delta G^\nu \).)

For Lpp-56 such experimental data are available: \( \Delta G^\nu(pH3\rightarrow pH7) = \Delta G^\nu(pH3) - \Delta G^\nu(pH7) = 14 \text{ kcal/mol} \) [8]. The calculations based on the assumptions given above result in a value of 7 kcal/mol for \( \Delta G^\nu \), which is essentially lower than the experimental observation. The question is now, which of the assumptions leads to this underestimation.

![Fig. 1. The electrostatic free energy of the homotrimeric coiled coil Lpp-56 as a function of pH. The calculations were done using either the MD-generated ensemble of structures (continuous line) or with the X-ray structure (dashed line).](image)

**Conformational Flexibility and Salt Bridge Dynamics**

The Lpp-56 protein represents a suitable object of investigation of electrostatic interactions and salt bridge contribution to structural stability. Lpp-56 is a homotrimer constituted by three \( \alpha \)-helices forming an in-register coiled coil. The molecule is reach of charged groups, the great majority of which is involved in salt bridges. Due to the specifics of the structural organisation of the protein the salt bridges form rings of intersubunit salt bridges, which clamp the superhelix along its length. The crystal structure and, more importantly, the available experimental observations suggest that electrostatic interactions and the large number of salt bridges likely contribute to the stability of the molecule [6, 8].

As pointed out above, one of the assumptions used in theoretical predictions of \( \Delta G^\nu \) is that the folded protein is represented by a single structure. This assumption is valid in rear cases within narrow pH intervals, and severely oversimplifies the physical reality. Proteins are flexible molecules, so that their native structures in solution are better described by ensembles of a large number of conformers. If electrostatic interactions are of interest, the conformational flexibility is often simulated by introducing a high protein dielectric constant [2, 21, 23]. It has been shown, however, that protein dielectric constant cannot be used as a parameter to fit the calculated \( pK \) values to the experimental data [12]. We will show below that protein permittivity cannot be used as parameter at \( \Delta G^\nu \) calculations, either. An alternative approach is to collect conformations using simulation methods such as molecular dynamics [3, 5, 9, 14, 22, 26, 27, 28].

The continuous line in Fig. 1 represents \( \Delta G^\nu(pH) \) calculated on the basis of 7 ns molecular dynamics simulation combined with calculations of \( W(x) \) [7]. The procedure of calculations is simple. During the molecular dynamic simulation the coordinates of the atoms are saved each 2.5 ps (snapshots). For each snapshot calculations of \( W(x) \), \( \theta_i \) and respectively \( \nu^\nu \) are carried out. The final values of \( \nu \) used in Eq. (4) are the average over the values obtained for the individual snapshots.

The value obtained for \( \Delta G^\nu(pH3\rightarrow pH7) \) is 15 kcal/mol (Fig. 1), which is very close to the experimental observation. The good agreement with the experimental data is objective of any computational work, but the results obtained here are at the same time surprising. During the molecular dynamic simulation the side chains, including those of the ionisable groups, may adopt conformations not permitting the formation of salt bridges. Such information lead to a reduction of the average electrostatic energy of interactions and hence, to an overall reduction of the magnitude of \( \Delta G^\nu \). The simulation shows however a different behaviour of the groups involved in salt bridges.

In order to explain this seemingly unexpected behaviour and to give an answer of the question posed at the end of the previous section, we consider two salt bridges linking the helices B and C of Lpp-56. As shown in Fig. 2, a lysine residue from helix C (LysC38) can form alternative salt bridges with two aspartic acids from helix B: LysC38–AspB40 (Fig. 2A) and LysC38–AspB33 (Fig. 2B). The lifetimes of the two salt bridges are compared in Fig. 3. It is important to note that during the first nanosecond none of the salt bridges is formed. The same was observed also for other salt bridges in Lpp-56. Hence, there are salt bridges which are not present in the X-ray structure can be formed during the simulation. (the structure at time zero in Fig. 3). Therefore, the mentioned underestimate value of \( \Delta G^\nu \) calculated with the assumption of a fixed protein structure is likely caused by the longer distances between some functional groups observed in the X-ray structure of Lpp-56. The molecular dynamics simulation “improves” the salt bridge geometry which leads to a more correct prediction of the electrostatic stabilisation of the molecule.
The opposite effect can also be observed. **Fig. 4** shows the time dependence of the proton donor-acceptor distances in the salt bridge LysA15–GluB20 in another protein: the GCN4 leucine zipper. During the first 2.5 nanosecond, including time zero (the X-ray structure), we observe a salt bridge with perfect proton donor-acceptor distance. As illustrated in the Fig., in this time window there is a rotation of the around the Cβ-Cγ bond of the carboxyl group of GluB20. As the consequence the proton acceptor atom is exchanged, yet the salt bridge is still formed. However, after approximately 2.5 ns a structural fluctuation leads to desintegration of the salt bridge for the following 2 ns. The lifetime of this salt bridge is 48% of the total simulation time; hence its contribution to the stability of the protein is lower than the predicted on the basis of the X-ray structure.
undergoes a conformational transition leading to a swap of the salt bridge partner. A quasi stable state is observed in which the interactions of the ε-amino group of the lysine with the two carboxyl groups are mediated by water molecules. Such configurations, known as water mediated salt bridges, are often observed in the crystal structures of proteins. The molecular dynamics simulation carried out in this study suggests that, at least for Lpp-56, they are not long-living formations. After the conformational transition the cross-link between helices B and C is restored.

The considered examples illustrate how important is the conformational flexibility for the adequate prediction of electrostatic interactions in proteins, and of their contribution to protein stability in particular. Regarding the calculation of \( \Delta G^\text{ele} \), as already mentioned, the simplified approach of adjusting the protein dielectric constant has been devised and has become very popular in fact. The presented results demonstrate that a priori parameterizations are not a universal solution. Increasing the protein dielectric constant may have worked for the GCN4 leucine zipper, where electrostatic interactions calculated on the basis of the X-ray structure are overestimated. However, in the case of Lpp-56 the charge-charge constellation in the X-ray structure is such that the strength of electrostatic interactions are underestimated by computation, a problem which couldn’t have been solved by re-parameterization on sound physical grounds.

It is important to realise the limitation of the theory and the computational approach presented above. The molecular dynamics simulation is carried out within an a priori determined net charge of the molecule, which does not change during the simulation. Hence, the ensemble of structures collected by the molecular dynamics simulation corresponds to an a priori selected pH. Therefore, the calculation of electrostatic interactions at any other pH represents an extrapolation assuming the ensemble being representative for this pH. Finally, it should be noted that the calculations are made assuming ergodicity of the system. There are no formal criteria to assess this assumption. In spite of the mentioned limitations, the combination of molecular dynamics and electrostatic calculations is a powerful tool revealing structural properties, including those of the salt bridges, emerging only when the protein molecule is “seen” in motion.

Acknowledgments

Work from I.J.’s own laboratory has been supported in part by the Swiss National Science Foundation.

REFERENCES