

Optimization of the electrostatic interactions between ionized groups and peptide dipoles in proteins

VELIN Z. SPASSOV,^{1,2} RUDOLF LADENSTEIN,¹ AND ANDREJ D. KARSHIKOFF¹

¹Centre for Structural Biochemistry, Karolinska Institutet, NOVUM, S-14157 Huddinge, Stockholm, Sweden

²Institute of Biophysics, Bulgarian Academy of Sciences, 1113 Sofia, Bulgaria

(RECEIVED January 6, 1997; ACCEPTED March 6, 1997)

Abstract

The three-dimensional optimization of the electrostatic interactions between the charged amino acid residues and the peptide partial charges was studied by Monte-Carlo simulations on a set of 127 nonhomologous protein structures with known atomic coordinates. It was shown that this type of interaction is very well optimized for all proteins in the data set, which suggests that they are a valuable driving force, at least for the native side-chain conformations. Similar to the optimization of the charge–charge interactions (Spasov VZ, Karshikoff AD, Ladenstein R, 1995, *Protein Sci* 4:1516–1527), the optimization effect was found more pronounced for enzymes than for proteins without enzymatic function. The asymmetry in the interactions of acidic and basic groups with the peptide dipoles was analyzed and a hypothesis was proposed that the properties of peptide dipoles are a factor contributing to the natural selection of the basic amino acids in the chemical composition of proteins.

Keywords: electrostatic interactions; Monte-Carlo simulations; peptide dipoles; protein stability

The role of the peptide groups in functional properties of proteins has long been in the scope of many studies. Two features of these groups seem to be of dominant importance. The first one is the ability of the peptide amide nitrogens and carbonyl oxygens to form hydrogen bonds and thus to constitute the secondary structure elements of proteins. The second one is their dipolar nature. Taken together, these two properties result in an organized, nonrandom, distribution of the orientation of the peptide dipoles and superposition of their electric field over large portions of the protein molecule. On this basis, the concept of the α -helix macro dipole has been introduced (Wada, 1976). Hol et al. (1978) assumed that the field of an α -helix macro dipole is equal to the field of a half positive unit charge at the amino end and a half negative unit charge at the carboxyl end. This idea has been applied in analysis of the electrostatic interactions in a number of studies (Friend & Gurd, 1979; Hol et al., 1981; Warwicker & Watson, 1982; Rogers & Steinberg, 1984). The α -helix macro dipole representation seems, however, to be a rather rough approximation. It has been shown, for example, that the electrostatic effect of an α -helix is not associated with the macro dipole, but rather with the few dipoles confined to the end turns of the helix (Åqvist et al., 1991). Moreover, it has been shown (Åqvist et al., 1991) that models based on this approximation might lead to incorrect conclusions due to neglecting the large dielectric effect of the protein/solvent environment.

The alignment of the peptide dipoles in the secondary structure elements was found to be significant for different properties of proteins, such as counterion (Hol et al., 1978) and nucleotide binding (Muegge et al., 1996), enzymatic activity (Warwicker & Watson, 1982; Karshikov et al., 1993), structural stability of proteins (Hol et al., 1981; Hol, 1985), but not always as a stabilizing factor (Gilson & Honig, 1989).

The fundamental role of the peptide dipoles in protein electrostatics becomes evident in their interactions with the side-chain charges. It has been found that the ionizable groups are situated in regions where favorable interactions with the peptide dipoles occur (Olson & Spasov, 1987). This finding is consistent with the observation that the destabilizing effect of the charge dehydration is compensated by the electrostatic contribution of the peptide dipoles (Bashford & Karplus, 1990; Oberoi et al., 1996).

The spatial arrangement of the peptide dipoles and side-chain charges is a result of the delicate balance of different forces stabilizing native protein structure, including electrostatic interactions. In other words, as far as the orientations and the positions of the peptide dipoles and the positions of the side-chain charges are a result of forces that are not only electrostatic, the magnitude of dipole–charge interactions will implicitly contain the contribution of all other interactions. In a series of papers (Spasov & Atanasov, 1994; Spasov et al., 1994, 1995), we have introduced the concept of optimization of a given type of interactions. It represents the divergence of the value of the energy of selected interactions in the native structure from that calculated for some reference state. The reference state can be defined as a structure where the interactions of interest are set to zero. The divergence, or

Reprint requests to: A. Karshikoff, Centre for Structural Biochemistry, Karolinska Institutet, NOVUM, S-14157 Huddinge, Stockholm, Sweden; e-mail: aka@csb.ki.se.

a parameter related to it, the optimization parameter, gives a measure of the contribution of the selected type of interactions to protein properties, including structural stability, in the context of all other interactions. We have applied this approach to the analysis of charge–charge interactions (Spassov & Atanasov, 1994; Spassov et al., 1994). It has been shown that the optimization parameter mirrors some general rules governing stabilization of native protein structure. It has also been shown that the optimization of charge–charge and protein–solvent interactions are related to the thermotolerance of proteins (Spassov et al., 1995).

In this paper, we have extended our approach to describe non-covalent interactions by an analysis of the optimization of the interactions between side-chain charges and peptide dipoles (charge–backbone interactions). The analysis has been performed on a set of 127 nonhomologous proteins with known three-dimensional structures. We have found that the interactions between peptide dipoles and side-chain charges are very well optimized and can be considered an important factor for stabilization, at least, of the native conformation of the side chains in proteins. We have examined the interactions between the side-chain charges and peptide dipoles from a more or less nontraditional view point. We analyzed the nonequivalence in the interactions of positive and negative charges with the peptide dipoles. On this basis, we have proposed the hypothesis that the properties of the peptide dipoles are a factor contributing to the natural selection of the basic amino acids in the chemical composition of proteins.

Results and discussion

Optimization of charge–dipole interactions

Our main assumption is that the side-chain charges would adopt a random distribution if they were not relevant to any protein property. The deviation of the side-chain charge constellation of the native structure from a random charge distribution must reflect the contribution of the electrostatic interactions to the structural stability and other functional properties of proteins. As a measure of this deviation we have introduced an optimization parameter, P_{opt} (see Method of calculation for definition).

Similar to the optimization parameter of charge–charge interactions introduced in our previous work (Spassov et al., 1994), P_{opt} is related to the difference between the charge–backbone interaction energy calculated for the native protein structure and the mean value calculated for a sufficiently large number of randomly distributed side-chain charges (see Equation 4). The larger the difference, the larger the negative value of the optimization parameter. Figure 1 represents the distribution of $\Delta G_{cd,rd}$ for the protein thermitase. The energy difference determining the optimization, $\Delta\Delta G_{cb} = (\Delta G_{cb,nv} - \langle \Delta G_{cd,rd} \rangle)$, shown in this figure is typical for a well-optimized protein.

The reference state is of some interest. One can expect that randomly distributed charges in the field of dipoles that are *not systematically oriented* will give $\langle \Delta G_{cd,rd} \rangle \approx 0$. Indeed, the maximum of $f(\Delta G_{cd,rd})$ for thermitase (Fig. 1) is very near zero. If the peptide dipoles are not systematically oriented, this result must be independent of the charge composition (the number of the positive and negative side-chain charges) because the probabilities that a given charge is situated in positive or negative field are equal. However, the maxima of $f(\Delta G_{cd,rd})$, calculated separately for the positive and negative charges, are far from zero (Fig. 1). For all structures presented in the data set (Table 1), the charge–backbone

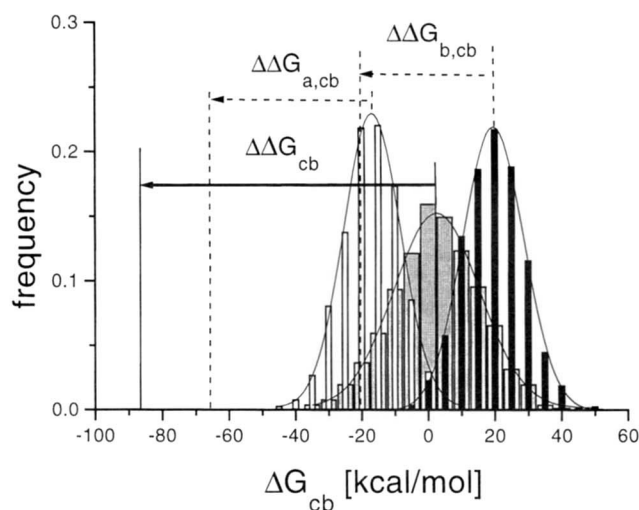


Fig. 1. Distribution, $f(\Delta G_{cd,rd})$, of charge–backbone interactions in thermitase: grey bars, random distributions of all side-chain charges; hollow bars, $f_a(\Delta G_{cd,rd})$, random distributions of the negative charges (acidic side chains); black bars, $f_b(\Delta G_{cd,rd})$, random distributions of the positive charges (basic side chains); continuous lines, Gaussian fit. The charge–backbone interaction energy, ΔG_{cb} , for the native charge distribution is marked by a vertical solid line. $\Delta\Delta G_{cb} = \Delta G_{cb} - \langle \Delta G_{cd,rd} \rangle$ gives the energy difference between the native state and the reference state and, expressed in σ , represents the optimization parameter, P_{opt} (see Equation 4). $\Delta\Delta G_{cb,a}$ and $\Delta\Delta G_{cb,b}$ have the same meaning for the negative and positive side-chain charges, respectively. Note that the net charge–backbone interaction energy is essentially larger for the negative charges (the left-hand dashed vertical line) than for the positive charges. Values of $\Delta\Delta G_{cb,a}$ and $\Delta\Delta G_{cb,b}$ are, however, similar, hence, the optimization of the charge–backbone interactions for both acidic and basic side chains is similar.

interaction energy for the negative charges is always attractive, $\langle \Delta G_{cd,rd} \rangle_a < 0$, whereas for the positive charges, it is always repulsive, $\langle \Delta G_{cd,rd} \rangle_b > 0$. This indicates that $\langle \Delta G_{cd,rd} \rangle$ depends on the charge composition, hence the charge–backbone interactions are influenced by a *systematic orientation* of the peptide dipoles. This issue will be discussed later in this paper.

Inspection of the degree of the optimization calculated separately for the basic and acidic groups shows that it is significant for almost all structures, independent of the group type. Thus, in the native structure, the distributions of the charges of both acidic and basic groups can be essentially shifted from the random distribution. According to our interpretation of the degree of optimization, it means that the charge–dipole interactions for both acidic and basic groups contribute to protein properties.

In Figure 2, the optimization parameters, calculated for the set of 127 nonhomologous protein structures (Table 1), are plotted versus the molecular weight expressed by the number of non-hydrogen atoms, N_{at} . A functional relationship between the optimization parameter and the molecular weight cannot be defined. However, the few structures that show no optimization ($P_{opt} > -1$) belong to proteins with low molecular weight. Also, it is seen that all proteins with a molecular weight above 28 kDa ($N_{at} \approx 2,000$) are very well optimized ($P_{opt} < -3$). In our previous work (Spassov et al., 1994), we have shown that an optimization parameter less than -3 corresponds to a very low probability that a charge distribution in the native structure is the result of a random process. A larger dispersion of P_{opt} is observed for proteins below 28 kDa, but the average value of P_{opt} for these structures is still about -3 .

Table 1. PDB entry codes of the protein structures used in this work

| | | | | | | | | | |
|------|------|------|------|------|------|------|------|------|------|
| 6RLX | 1CRN | 9INS | 3MT2 | 2OVO | 5RXN | 1AAP | 1ROP | 1CDT | 1PI2 |
| 1NXB | 1R69 | 1SN3 | 2CTX | 2HIP | 1HOE | 1CC5 | 1UBQ | 351C | 2FXB |
| 1TPK | 2GN5 | 3B5C | 3FXC | 1PCY | 1SAR | 2MCM | 8RNT | 2CDV | 4CPV |
| 2TRX | 6B25 | 1FKB | 4FD1 | 1YCC | 1CCR | 1C2R | 1RNB | 1CY3 | 1MSB |
| 1TGI | 1PAZ | 2CCY | 7RSA | 1GMF | 1BP2 | 1BBH | 2MHR | 2AZA | 3CHY |
| 1LZT | 3FGF | 1LZ1 | 1ECA | 1IF1 | 1THB | 3FXN | 1MBA | 2SOD | 1SDH |
| 2SNS | 1END | 4CLN | 1LPE | 2LH4 | 1MBD | 2I1B | 2RN2 | 2SGA | 2LZM |
| 5P21 | 2FCR | 1CD4 | 2ALP | 1BBP | 1GKY | 3SDP | 1GCR | 1COL | 8DFR |
| 3ADK | 1SGT | 3GAP | 9PAP | 1HNE | 2CNA | 1GST | 1LTE | 5TIM | 1THM |
| 1RVE | 3BLM | 2CBA | 1AAI | 2TSC | 1HSA | 6ABP | 1RHD | 2GBP | 1FNR |
| 4APE | 3APR | 8ATC | 3TLN | 5CPA | 1IPD | 2LBP | 1GOX | 1ALD | 8ADH |
| 1NSB | 6XIA | 1PHH | 3PGK | 7AAT | 2CPP | 4ENL | 1CSC | 1PII | 1NPX |
| 3GRS | 1GLY | 1PGD | 1COX | 8CAT | 1LFI | 6ACN | | | |

It follows that charge-backbone interactions are well optimized for most of the structures, which suggests that these interactions are a notable driving force, at least for the native side-chain conformations. In Spassov et al. (1995), we have proposed that a molecular mass of about 28 kDa is critical for the formation of a structural domain with a well-defined hydrophobic core. This value coincides very well with the relation between P_{opt} and the molecular mass observed in this study. The increased P_{opt} for proteins with well-developed hydrophobic core probably reflects the enhanced efficiency of the electrostatic interactions due to enlargement of the low permittivity medium.

In our previous work (Spassov et al., 1994), we have analyzed the degree of the optimization of charge-charge interactions in

different structural and functional types of proteins: proteins with or without disulphide links; secondary structural classes, which include proteins of α -helical, β -strand, or mixed $\alpha\beta$ type; and enzymes and proteins without enzymatic function. We have repeated this analysis for the optimization of charge-dipole interactions. The results are listed in Table 2.

The optimization of charge-backbone interactions calculated for the enzymes in the data set is shifted by more than 1σ unit in comparison with that calculated for proteins without enzymatic activity. We have also observed such a tendency for the optimization of the charge-charge interactions (Spassov et al., 1994). It is well known that the function of the enzymes as catalysts requires a protein molecule to provide and stabilize a unique environment appropriate for catalysis of a given chemical reaction. A contribution to this uniqueness comes obviously from the specific arrangement of the charges, especially in the cases of charged substrates or active sites involving titratable groups. The increased degree of optimization of both charge-charge (Spassov et al., 1994) and charge-backbone (this work) interactions, being an indication that the charge constellation in an enzyme molecule deviates from a random distribution more than in other proteins with respect to both mutual distances between charged groups and their position in the field of the peptide dipoles, reflects not only the stabilizing role of the side-chain charges, but also their "functionally driven" arrangement. A very similar conclusion has been made earlier by Hwang and Warshel (1988). On the basis of an analysis of ion pair reversal, these authors have stated that "the environment designed by nature to stabilize a given ion pair is not usually optimized for the inverted pair."

The values of $\langle P_{opt} \rangle$ calculated for the proteins classified regarding the other two criteria (presence of disulfide links and folding types) do not show a statistically significant difference from the average obtained for the whole data set (see Table 2), i.e., the optimization of the charge-backbone interactions is insensitive to these criteria. It has been found earlier (Spassov et al., 1994) that the lack of disulfide bridges is compensated by a higher optimization of charge-charge interactions. The absence of such a correlation for $\langle P_{opt} \rangle$ suggests that charge-backbone interactions do not appear as a factor compensating the deficiency of stabilizing covalent crosslinks and vice versa. A reason for this is probably the difference in the environment where charge-charge and charge-backbone interactions occur. Very often, salt bridges are formed on the protein surface by side chains belonging to different structural domains, thus contributing to the stabilization of the folded struc-

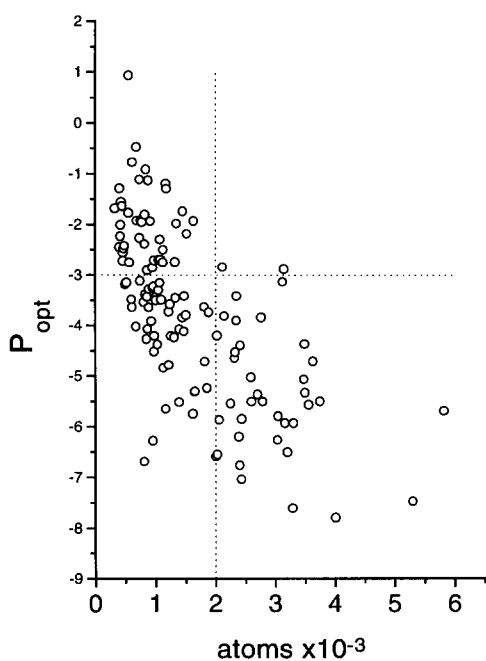


Fig. 2. Plot of optimization parameters, P_{opt} , calculated for 127 non-homologous structures (Table 1) versus number of non-hydrogen atoms in proteins. The vertical dashed line separates proteins with well-developed hydrophobic core (right-hand side) according to the criterion discussed in (Spassov et al., 1995). Values of P_{opt} below the horizontal dashed line are considered as well optimized structures (Spassov et al., 1994).

Table 2. Mean values of the optimization parameter, P_{opt} , estimated for proteins of different structural and functional classes represented in the data set

| | For the data set | Function ^a | | Disulfide links ^b | | Folding type ^c | |
|---------------------------|------------------|-----------------------|------|------------------------------|------|---------------------------|---------|
| | | E | N | SS | N | α | β |
| $\langle P_{opt} \rangle$ | -3.8 | -4.7 | -3.5 | -3.9 | -3.7 | -3.5 | -3.7 |
| Standard deviation | 1.7 | 1.3 | 1.2 | 1.7 | 1.7 | 1.5 | 1.5 |

^aFunctional types: enzymes (E) and proteins without enzymatic function (N). $\langle P_{opt} \rangle$ was calculated for proteins with N_{at} between 1,000 and 3,000 (60 structures) because the representation of the two types of proteins in this interval is equivalent. Below this range, only few enzymes are represented, whereas only few proteins without enzymatic function have $N_{at} > 3,000$. In this way, a possible influence of the molecular weight dependence on $\langle P_{opt} \rangle$ is reduced.

^bProtein with (SS) and without (N) disulfide links.

^cFolding type was defined according to Spassov et al. (1994). Proteins of mixed $\alpha\beta$ type are not uniformly represented in the data set with respect to the molecular mass, therefore they were not considered here.

ture. The interactions between side-chain charges and the peptide dipoles essentially can be attractive (see below, Table 3), however, they can be neutralized by the dehydration effect due to burial of the charges (the charge–backbone interactions are significant on relatively short distances). Therefore, these interactions seem to be inefficient for the stabilization of the folded structure in the way the charge–charge interactions do.

The insensitivity of the optimization parameter to the folding type is more or less a surprising result. As discussed above, the dependence of $\langle \Delta G_{cd,md} \rangle$ on the charge composition shows that the charge–backbone interactions are influenced by a systematic orientation of the peptide dipoles. Given that the optimization parameter is a measure of these interactions, one can conclude that the systematic orientation of the peptide dipoles due to the secondary structure elements in proteins does not cause this effect: the mean values of the optimization parameters for the proteins of α and β dominant folding type are statistically equivalent. It follows that the efficiency of interaction between the ionized groups and the peptide dipoles seems to be insensitive to the formation of main-chain macro dipoles. Our results contradict the conclusion made by Gandini et al. (1996), who have suggested that side-chain–main-chain interactions depend on the type of secondary structure involved. One possible reason for this discrepancy might be the difference in the criteria used in the two approaches. We analyzed the positions of the side chains in the electrostatic field of the peptide dipoles, whereas Gandini et al. (1996) have investigated the hydrogen bonding of ionizable groups with the backbone atoms.

Table 3. Average electrostatic interaction energy, ΔG_{cb} , between titratable side chains and peptide dipoles in kcal/mol/res^a

| Side chain | Asp | Glu | His | Lys | Arg |
|-----------------------------------------|--------|--------|--------|--------|--------|
| Residues in the set | 1,552 | 1,500 | 553 | 1,698 | 1,084 |
| Energy | -3.12 | -1.54 | 0.61 | -0.24 | -0.12 |
| Standard deviation | (3.86) | (2.73) | (1.55) | (1.72) | (2.28) |
| Residues in attractive interactions (%) | 91 | 83 | 33 | 51 | 47 |

^aThe average is taken over the common pool of titratable groups in the data set.

Asymmetry of charge–backbone interactions in proteins

It was noted that the interactions of the acidic and basic groups with the backbone charges are not equivalent and that this effect is not due to the arrangement of the peptide dipoles in secondary structure elements. From the observation that $\langle \Delta G_{cd,md} \rangle_a$ is less than $\langle \Delta G_{cd,md} \rangle_b$ for all proteins in the data set, illustrated in Figure 1 for thermolysin, it becomes evident that the peptide dipoles create a positive electrostatic field over the side-chain atoms. Indeed, the electrostatic potential averaged over the common pool of atoms in the data set with respect to their side-chain positions is positive for all carbon and oxygen atoms (Fig. 3). The nitrogens at δ and ϵ positions, including the ionizable atoms of histidines, are also in a positive field. As seen from Table 3, the average charge–backbone interaction energy for histidines is positive, because only one third of them are in attractive contacts. This result is in accord

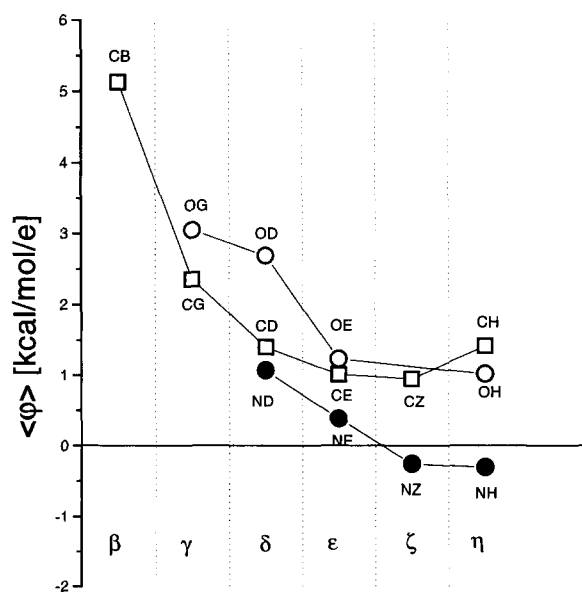


Fig. 3. Average electrostatic potential of the peptide dipoles on the side-chain atoms: squares, carbon atoms; open circles, oxygen atoms; filled circles, nitrogen atoms. Atom positions in the side chains are designated according to the PDB nomenclature (Bernstein et al., 1977).

with the fact that histidines are often uncharged at physiological pH (Plesniak et al., 1996) and are treated as buried (nontitratable) groups. An example for this is myoglobin, where few histidines become titratable only after denaturation (Friend & Gurd, 1979). The average potential is slightly negative only for the nitrogens from the guanidino and amino groups of the arginines and lysines. Graphically, the electrostatic potential produced by the backbone charges is represented in Figure 4. In both α -helix and β -sheet structures, the side chains are immersed in positive potential. These data demonstrate nothing else but the fact, not often discussed probably due to its triviality, that the peptide dipole is oriented in a way that the positive potential encloses the side chain. Thus, the observed asymmetry in the charge–backbone interactions follows directly from the nature of the peptide bond.

Implications of charge–backbone interactions on stability and chemical composition of proteins

Oberoi et al. (1996) have noticed that most of the large charge–backbone interactions (1) are attractive, (2) are paired with a large

positive Born term, and (3) correspond to anionic groups. Taking into consideration the asymmetry of charge–backbone interaction, discussed above, this observation can be generalized for the acidic groups in proteins. Charge–backbone interaction energy for the acidic groups is essentially negative, and almost all of them are in attractive interactions with the peptide dipoles (Table 3; Fig. 3). It follows that a general tendency of compensation of the effect of dehydration of the acidic groups exists that results in a stabilization of their charged form.

The behavior of the basic residues is, however, different. The charge–backbone interaction energy for the basic groups is still favorable, but negligible, and only about half of them are involved in attractive interactions (Table 3). Due to their long side chains, the basic residues (lysines and arginines) can adopt conformations that reduce the unfavorable interactions with the peptide dipoles and increase the hydration of the ionizable atoms, thus stabilizing their charged form. Given that the stabilization of the charged form of ionizable groups is a prerequisite for the salt bridge stability, the apparent reduction of the charge–backbone interaction energy of

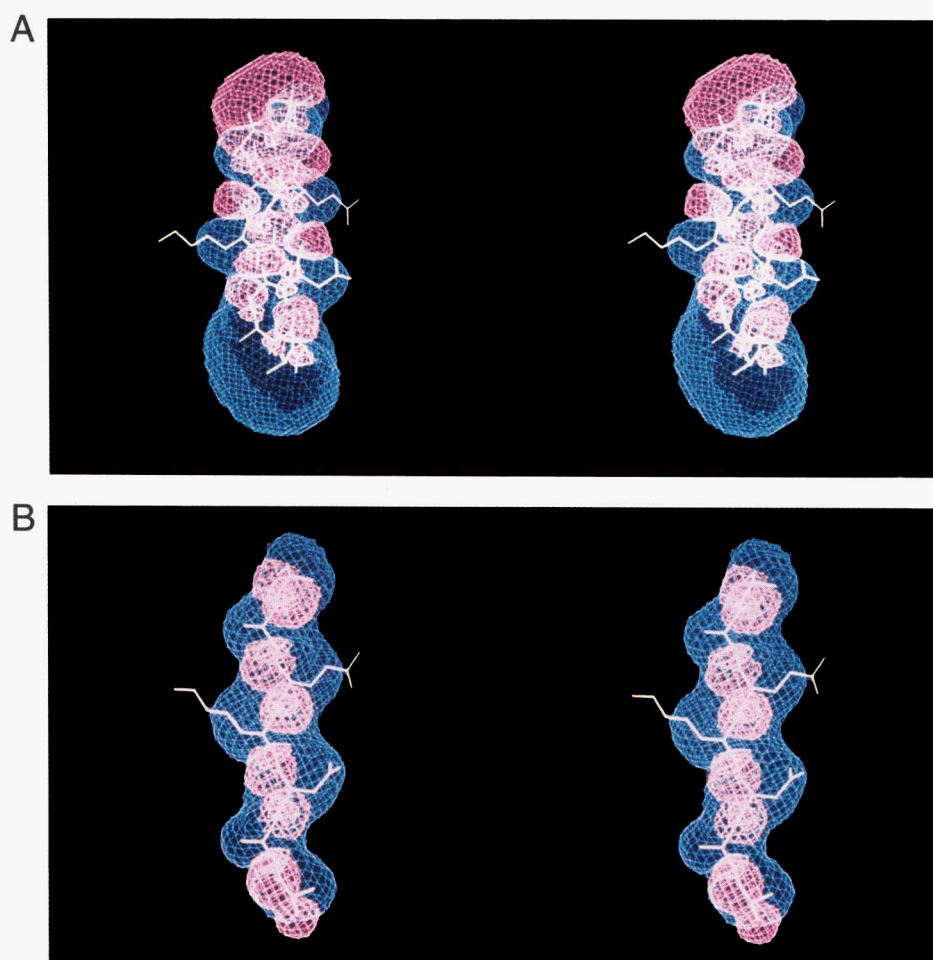


Fig. 4. Stereo view of the electrostatic potential created by peptide dipoles: blue for positive, red for negative potential, contoured at 0.5 and -0.5 kcal/mol/e, respectively. **A:** α -Helix, (Ala)₄-Asp-Ala-Lys-Ala-Glu-(Ala)₄. **B:** β -Strand (Ala)₂-Asp-Lys-Glu-(Ala)₂. The positive potential embraces the space where most conformations of the acidic side chains can be realized for both α -helix and β -strand. The charge of lysine is able to “escape” the positive electrostatic potential due to side-chain length. Although out of the scope of this study, it is important to note the increased electrostatic potential at the two ends of the α -helix, which comes from the terminal dipoles, an effect pointed out earlier by Åqvist et al. (1991). The zero isopotential surface (not shown) is perpendicular to the helix and converges rapidly to a plane, thus an α -helix can still be considered as a macro dipole if long-range interactions have to be assessed, say, for purposes of molecular recognition. Model building was performed by Insight II (Biosym Technologies, Inc., San Diego, California).

the basic groups is a stabilizing factor. It follows that, due to the asymmetry of the charge–backbone interactions, the stabilization of the charged form of the titratable groups (hence of the salt bridges) in native proteins is realized in different ways. For the acidic groups, this is the tendency of compensation of the dehydration effect by the favorable charge–backbone interactions, whereas, for the basic groups, this is the length of the side chain. It turns out that the side-chain length of the basic residues is a factor for protein stability. Let us consider an imaginary protein molecule containing basic groups with side-chain lengths comparable to those of the acidic groups, say, bases with γ - and δ -amino groups. Such short side chains would be immersed in the positive potential of the peptide bond (Fig. 4), thus both charge–backbone interactions and dehydration effect would destabilize the charged form of these groups. This effect would contribute to salt bridge destabilization at alkaline pH, hence the alkaline denaturation would be shifted toward neutral pH. Our speculations were made on the basis of averaged quantities (Fig. 3; Table 3), showing certain tendencies rather than concrete values. Therefore, an assessment of the postulated shift of the alkaline denaturation cannot be done accurately. Nevertheless, taking into account the contribution of the Born term alone, first introduced for p*K* calculations by Warshel et al. (1984), one may expect a reduction of p*K* of such groups by about 2–4 units (Bashford & Karplus, 1990; Karshikoff, 1995; Gandini et al., 1996). The p*K* values of *n*-propyl- and *n*-butylamine, which can serve as models of the shorter side chains, are close to that of lysines, between 10.5 and 10.6, respectively (Evans & Hamann, 1951). It follows, then, that the intrinsic p*K* in proteins of these hypothetical basic groups would be between 6.5 and 8.5, hence proteins would be destabilized at the physiologically relevant region of pH.

From our analysis, it follows that the asymmetry in the electrostatic interactions between side-chain charges and peptide dipoles is inherent to proteins and is compensated by the relatively long side chains of the basic residues. Thus, the electrostatic properties of the peptide bonds appear to be a factor responsible for the natural selection of the type of the basic amino acids in the chemical composition of proteins.

Method of calculation

The contribution of side-chain charges (the charges carried by the titratable groups in their ionized form) to the electrostatic term of free energy of proteins can be represented as

$$\Delta G_{el} = \Delta G_s + \Delta G_{cc} + \Delta G_{cp} + \Delta G_{cb}, \quad (1)$$

where ΔG_s is the free energy change do to transfer of the side-chain charges from water to protein, ΔG_{cc} is the charge–charge interaction energy, ΔG_{cp} is the electrostatic energy contribution of the side-chain charges and side-chain dipoles, and ΔG_{cb} is the interaction energy between the side-chain charges and backbone dipoles. The optimization of the interactions reflected by the last quantity in Equation 1 was in the scope of our study. ΔG_{cb} can be calculated easily by means of Equation 2:

$$\Delta G_{cb} = \sum_i^N q_i \phi_b(r_i), \quad (2)$$

where $\phi_b(r_i)$ is the electrostatic potential created by the peptide dipoles at the position, r_i , of a side-chain charge q_i , and N is the number of the side-chain charges in protein.

Calculation of the electrostatic potential

A continuum model was used for the calculations of the electrostatic potential $\phi_b(r_i)$. The protein was represented as a material with low dielectric constant, $\epsilon(r) = \epsilon_p = 4$, immersed in the high permittivity medium of the solvent, $\epsilon(r) = \epsilon_s = 80$. The dielectric boundary was determined using a probe sphere with radius of 1.4 Å. The low dielectric constant was ascribed to the space covered by the protein atoms or to the space inaccessible to the probe. The atomic radii were taken from Miller et al. (1987). The Poisson equation,

$$\nabla \epsilon(r) \nabla \phi(r) = -4\pi \rho(r), \quad (3)$$

was solved for this system using the finite difference method (Warwicker & Watson, 1982; Klapper et al., 1986; Press et al., 1989; Nicholls & Honig, 1991), where $\rho(r)$ represents the partial charges of the backbone atoms N, H, C, O, and C α , which constitute the peptide dipoles. The coordinates of the main-chain H atoms were obtained by a computer program based on the peptide bond geometry. The charge values were taken from CHARMM parameter set 19 (Brooks et al., 1983). The Poisson equation (zero ionic strength) was chosen because ΔG_{cb} does not depend essentially on the ionic strength (Takahashi et al., 1992). Also, our preliminary calculations, using the Poisson–Boltzmann equation, showed that the influence of the ionic strength on the quantities considered in this work is negligibly small. The finite difference calculations were performed in a $64 \times 64 \times 64$ grid box centered at the geometric center of the protein molecule. The box length was twice the maximal length of the protein, followed by focusing (Klapper et al., 1986) to a box length equal to the maximum length of the protein plus one probe layer (2.8 Å).

Optimization parameter for charge–backbone interactions

The criterion for optimization of the interactions between side-chain charges and peptide dipoles (charge–backbone interactions) is very similar to that developed to study the optimization of charge–charge interactions (Spassov & Atanasov, 1994; Spassov et al., 1994). It is based on comparison of the charge–backbone interaction energy in the native structure, $\Delta G_{cb,nw}$, with that in a reference state. The reference state corresponds to a protein structure (identical to the native one) for which the assumption is made that the electrostatic interactions do not contribute to the structural stability or any other protein property. We assume, then, that the side-chain charges are randomly distributed on the surface of the protein. The reference state is constructed by generating random charge distributions using Monte-Carlo simulation and calculating the charge–backbone interaction energy, $\Delta G_{cd,md}$, for each of them. For a sufficiently large number of random charge distribution, the frequency of occurrence, $f(\Delta G_{cd,md})$, is expected to be Gaussian. Indeed, for 1,000 random charge distributions, this condition was satisfied for all proteins investigated in this study. An illustration of $f(\Delta G_{cd,md})$ is given in Figure 1. The charge–backbone interaction for the reference state, $\langle \Delta G_{cd,md} \rangle$, was defined as the mean value of $\Delta G_{cd,md}$, where the Gaussian distribution has a maximum. Following the concept proposed in our previous study (Spassov et al., 1994), we define the optimization parameter of charge–backbone interactions as:

$$P_{opt} = (\Delta G_{cb,nw} - \langle \Delta G_{cd,md} \rangle) / \sigma, \quad (4)$$

where σ is the standard deviation.

Table 4. Side-chain charged atoms and charge values used in the calculation of $\Delta G_{cb,ntv}$ ^a

| Side chain | Atom | Charge |
|------------|------|--------|
| Asp | OD1 | -0.5 |
| | OD2 | -0.5 |
| Glu | OE1 | -0.5 |
| | OE2 | -0.5 |
| His | NZ1 | 0.25 |
| | NE2 | 0.25 |
| Lys | NZ | 1.0 |
| Arg | NZ | 0.33 |
| | NH1 | 0.33 |
| | NH2 | 0.33 |

^aAtom name designation is made according to PDB nomenclature (Bernstein et al., 1977).

In order to obtain P_{opt} , it is only necessary to calculate the charge-backbone interactions for the native structure and for the each random distribution of charges using Equation 2. For the native structure, the charge values and the atoms where the side-chain charges, q_i , are situated are listed in Table 4. The way of determination of the charges roughly reflects the most probable ionization state of a protein at neutral pH. Peculiarities such as extremely low or high pK values observed for some proteins were not considered. In the case of a random distribution, the charges can occupy all atoms, which are accessible to the solvent and do not belong to the peptide backbone. The atoms for which the distance to at least one peptide atom is independent on the side-chain conformation are excluded from the set of possible charge sites. These are C β , C γ , and O γ . Depending on whether the charges originate from acidic groups, basic groups, or histidines, values of -1, 1, or 0.5 were assigned, respectively. The values of q_i have the magnitude of one electronic charge.

Input data

The calculation of P_{opt} was performed for a set of 127 nonhomologous proteins with high-resolution X-ray structures, all available in Brookhaven Protein Data Bank (PDB) (Table 3). The entries were selected from a representative set of sequence-unbiased proteins proposed by Boberg et al. (1995). Incomplete structures or structures with resolution >2.5 Å were excluded from the set. Minor incompleteness, such as 1–2 unresolved side chains, was accepted. The calculations were performed for monomer structures only. For oligomeric proteins, the first subunit given in PDB was used.

Acknowledgments

This work was supported partially by the European Community project "Biotechnology of Extremophiles" via grant from NUTEK, Sweden and by the Academic Exchange Programme between the Royal Swedish Academy of Science and the Bulgarian Academy of Science.

References

- Åqvist J, Luecke H, Quirocho FA, Warshel A. 1991. Dipoles localized at helix termini of proteins stabilize charges. *Proc Natl Acad Sci USA* 88:2026–2030.
- Bashford D, Karplus M. 1990. pK_a's of ionizable groups in proteins: Atomic detail from a continuum electrostatic model. *Biochemistry* 29:10219–10225.
- Bernstein FC, Koetzle TF, Williams GJB, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. 1977. The Protein Data Bank: A computer-based archival file for macromolecular structures. *J Mol Biol* 112:535–542.
- Boberg J, Salakoski T, Vihinen M. 1995. Representative selection of proteins based on nuclear families. *Protein Eng* 8:501–503.
- Brooks BB, Bruccoleri RE, Olafson BD, States DG, Swaminathan S, Karplus M. 1983. CHARMM: A program for macromolecular energy minimization and dynamics calculations. *J Comp Chem* 4:187–217.
- Evans AG, Hamann SD. 1951. Energy and entropy change of organic reactions. *Trans Faraday Soc* 47:25–49.
- Friend SH, Gurd FRN. 1979. Electrostatic stabilization in myoglobin. pH dependence of summed electrostatic contribution. *Biochemistry* 18:4612–4619.
- Gandini D, Gogioso L, Bolognesi M, Bordo D. 1996. Patterns in ionizable side chain interactions in proteins. *Proteins Struct Funct Genet* 24:439–449.
- Gilson MK, Honig B. 1989. Destabilization of an α -helix-bundle protein by helix dipoles. *Proc Natl Acad Sci USA* 86:1524–1528.
- Hol WGJ. 1985. The variable effective dielectric constant and the importance of the direction of the peptide dipole moment: An investigation of dipole-dipole interactions in $\alpha\alpha$ and $\alpha\beta\alpha$ proteins. *J Mol Struct (Theochem)* 123:27–43.
- Hol WGJ, Halie LM, Sander C. 1981. Dipoles of the α -helix and β -sheet: Their role in protein folding. *Nature* 294:532–536.
- Hol WGJ, van Duijnen PT, Berendsen HJC. 1978. The α -helix dipole and the properties of proteins. *Nature* 273:443–446.
- Hwang JK, Warshel A. 1988. Why ion pair reversal by protein engineering is unlikely to succeed. *Nature* 334:270–272.
- Karshikoff A. 1995. A simple algorithm for calculation of multiple site titration curves. *Protein Eng* 8:243–248.
- Karshikov A, Reinemer P, Huber R, Ladenstein R. 1993. Electrostatic evidence for the activation of the glutathione thiol by Tyrosine 7 in p class GSH transferases. *Eur J Biochem* 215:663–670.
- Miller S, Janin J, Lesk AM, Chothia C. 1987. Interior and surface of monomeric proteins. *J Mol Biol* 196:641–656.
- Muegge J, Schweins T, Langen R, Warshel A. 1996. Electrostatic control of GTP and GDP binding in the oncoprotein p21^{ras}. *Structure* 4:475–489.
- Oberoi H, Trikha J, Yuan X, Allewell NM. 1996. Identification and analysis of long-range electrostatic effects in proteins by computer modeling: Aspartate transcarbamylase. *Proteins Struct Funct Genet* 25:300–314.
- Olson AJ, Spassov V. 1987. Electrostatic environments of charged residues in folded proteins. In: *Scientific Report 1986–87, Research Institute of Scripps Clinic* 13:200.
- Plesniak LA, Connelli GP, Wakarchuk WW, McIntosh LP. 1996. Characterization of a buried neutral histidine residue in *Bacillus circulans* xylanase: NMR assignment, pH titration, and hydrogen exchange. *Protein Sci* 5:2319–2328.
- Rogers NK, Steinberg MJE. 1984. Electrostatic interactions in globular proteins. Different dielectric models applied to the packing of α -helix. *J Mol Biol* 175:527–542.
- Spassov VZ, Atanasov BP. 1994. Spatial optimization of electrostatic interactions between the ionized groups in globular proteins. *Proteins Struct Funct Genet* 19:222–229.
- Spassov VZ, Karshikoff AD, Ladenstein R. 1994. The optimization of the electrostatic interactions in proteins of different functional and folding type. *Protein Sci* 3:1556–1569.
- Spassov VZ, Karshikoff AD, Ladenstein R. 1995. The optimization of protein solvent interactions. Thermostability and the role of hydrophobic and electrostatic interactions. *Protein Sci* 4:1516–1527.
- Takahashi T, Nakamura H, Wada A. 1992. Electrostatic forces in two lysozymes: Calculation and measurement of histidine pK_a values. *Biopolymers* 3(2):897–909.
- Wada A. 1976. The α -helix as an electric macro-dipole. *Adv Biophys* 9:1–63.
- Warshel A, Russell ST, Churg AK. 1984. Macroscopic models for studies of electrostatic interactions in proteins: Limitations and applicability. *Proc Natl Acad Sci USA* 81:4785–4789.
- Warwicker J, Watson NC. 1982. Calculation of the electric field potential in the active site cleft due to α -helix dipoles. *J Mol Biol* 157:671–679.