Eur Biophys J (1989) 17:287–297

European Biophysics Journal

Electrostatic interactions in proteins: Calculations of the electrostatic term of free energy and the electrostatic potential field

A. D. Karshikov [2]*, R. Engh [1], W. Bode [1], and B. P. Atanasov [2]

[1] Max-Planck Institut für Biochemie, D-8033 Martinsried, Federal Republic of Germany
[2] Institute of Organic Chemistry, Bulgarian Academy of Sciences, BG-1040 Sofia, Bulgaria

Received July 19, 1989/Accepted in revised form August 30, 1989

Abstract. The pH-dependence of the electrostatic energy of interactions between titratable groups is calculated for some well studied globular proteins: basic pancreatic trypsin inhibitor, sperm whale myoglobin and tuna cytochrome *c*. The calculations are carried out using a semi-empirical approach in terms of the macroscopic model based on the Kirkwood-Tanford theory. The results are discussed in the light of their physicochemical and biological properties. It was found that the pH-dependence of the electrostatic energy correlates with the III–IV transition of cytochrome *c*. The electrostatic field of the cysteine proteinase inhibitor, cystatin, was calculated in two ways. In the first one, the electrostatic field created by the pH dependent charges of the ionizable groups and peptide dipoles was calculated using the approach proposed. In the second one, the finite-difference method was used. The results obtained by the two methods are in overall agreement. The calculated field was discussed in terms of the binding of cystatin to papain.

Key words: Electrostatic interactions, pH-dependence, energy proteins

Introduction

Electrostatic interactions play a central role in a large variety of protein functions such as enzymatic catalysis (Warshel 1978; Sprang et al. 1987), ligand binding of hemoglobins (Perutz 1978), and protein-protein (Salemme 1976) and protein-DNA (Warwicker et al. 1987) interactions. Their role is clearly demonstrated by the pH-dependent properties of the proteins. The description of the pH-dependent phenomena requires knowledge about the ionization behaviour of the titratable groups of the protein, or equivalently, their pK values. The pK values are, in general, different from those of the corresponding titratable side chains free aminoacids in water solution. One of the reasons for this is that the protein environment causes a change in the intrinsic pK (pK_{int}). The value of pK_{int} is usually defined as the pK of a given group in a protein when all other titratable groups are in their neutral form. According to this definition, the change of pK_{int} is related to the energy of formation of a charge in a protein environment (its self energy). The pK values are also modulated by the electrostatic influence of other titratable groups in the protein. There are two main ways to approach the problem: the macroscopic one, based on the continuum description of the system protein – surrounding solvent and the microscopic one, treating the problem at the atomic level.

The oldest macroscopic theory of the electrostatic interactions in proteins is that of Kirkwood and Tanford (Kirkwood 1934; Tanford and Kirkwood 1957). In the Kirkwood-Tanford theory, the protein molecule is represented as a spherical dielectric cavity. The charges carried by the titratable groups are defined as point charges situated on its surface. This model has been extended with an iterative procedure accounting for the influence of the charge multipole on the dissociation of the titratable groups (Tanford and Roxby 1972). Later, it was modified by characterizing the ionizable groups according to their solvent accessibilities to describe their different dielectric environments (Shire et al. 1974; March et al. 1982; Matthew and Richards 1983; Garcia-Moreno et al. 1985). The modified Tanford-Kirkwood model was used for calculating the influence of the charge-charge interactions on the pK values of the titratable groups and the electrostatic stability of sperm whale myoglobin (Garcia-Moreno et al. 1985; Friend and Gurd 1979a, b), the titration curve of cytochrome *c* (Matthew et al. 1978), the influence of electrostatic interactions on the conformational stability (March et al. 1982), and the processes of proton exchange in bovine pancreatic trypsin

* To whom offprint requests should be sent

inhibitor (BPTI) (Matthew and Richards 1983). The results show good agreement with experimental data. In all these calculations, the self energy term is excluded, as the intrinsic pK values are defined outside the model.

The availability of three-dimensional atomic structures of proteins necessitated the development of the microscopic description of protein properties. The most prominent microscopic model of the electrostatic interactions in proteins is that of Warshel and Russell (Warshel and Russell 1984; Russell and Warshel 1985). In this model, the dielectric constant of the system is assumed to be that of free space. The dielectric properties of the protein and the surrounding solvent are described by means of a full set of partial atomic charges, atomic polarizability (permanent and induced dipoles), and the mobility of the solvent molecules in terms of Langevin approximation (Langevin dipoles). The model was applied for the calculation of pK_{int} of the acidic side chains of BPTI (Russell and Warshel 1985). A deviation from the expected values of about 3 pK units was obtained. The authors have pointed out that this discrepancy could be due to the fact that the protein configuration is kept fixed during calculations. They suggest that the creation of large sets of protein conformations, including molecular dynamic techniques in the model, should improve the results.

In the microscopic model discussed above, the charge-charge interactions between titratable groups are omitted. However, there are data showing that the charge-charge interactions can also play a role in the ionization equilibrium of the titratable groups. Sternberg et al. (1987) have investigated the effect on the pK value of the histidine of the active site of subtilisin arising from site-directed mutagenesis of charged groups. A deviation of up to 1 pK unit was found on substitution of different charged groups with uncharged groups or with groups with opposite charge. The catalytic activity of subtilisin is directly related to the ionization of the histidine in the active site. Russell et al. (1987) determined the pK value of this residue by measuring enzyme activity. They also detected a dependence of the pK on the change of the protein charge constellation. The authors assume that there are not significant changes in enzyme structure due to the mutation of the surface charged groups. This suggests that no significant changes in self energy, and thus in pK_{int}, of the histidine are expected. This suggestion is supported by the theoretical evaluations of the change of pK of this residue made by Gilson and Honig (1988).

In this paper we have analyzed theoretically the charge-charge interactions and their role in some pH-dependent properties of proteins. The calculations are based on a semi-empirical approach for determining the ionization behaviour of the titratable groups; the values of pK_{int} are taken from the experimental values for model compounds. The influence of the electrostatic interactions between charged groups on the values of pK of the individual groups are calculated on the basis of the macroscopic modified Kirkwood-Tanford model. A rapidly converging iterative procedure is described which allows calculations at pH values where a great number of groups are ionized simultaneously. The protein structure is kept fixed at its crystalline state; no relaxation of the structure is made and possible conformational changes due to changes of pH are not considered. The approach is tested for BPTI and is applied to calculations of the electrostatic energy of sperm whale metmyoglobin and tuna cytochrome *c*. Good agreement between calculated and experimental results is obtained. In particular, the calculations predict the occurrence of a pH-dependent transition in ferric cytochrome *c* corresponding to the III–IV transition, but not in ferrous cytochrome *c*, where the III–IV transition is not observed. The electrostatic potential field of egg white cystatin is calculated by means of Coulomb's law with a simple dielectric representation. The charges of the ionized groups used in the electrostatic field calculations are obtained by the modified Kirkwood-Tanford model. Calculations are also carried out using the finite-difference relaxation method proposed by Klapper et al. (1986) and Gilson and Honig (1988). Both methods are in general agreement.

Method of calculations

Intrinsic pK

The determination of the pK of the individual titratable groups in proteins has two problems: finding their intrinsic pK values and evaluating the influence of the protein charge multipole. The macroscopic Tanford-Kirkwood model gives a strong dependence of self energy on the radial position of a given titratable group; pK_{int} may increase by up to 25 pK units when the titratable group is moved from water to the center of the protein (Warshel et al. 1984). On the other hand, the microscopic calculation of pK_{int} produces a deviation of about 5 kcal/mol (more than 3.5 pK units) in the fixed-atom approximation (Russell and Warshel 1985). The deviation increases in the case of basic groups (Warshel and Russell 1984). This accuracy is not sufficient for analysis of the pH-dependence of the charge-charge interactions since the definition area of pK_{int} may entirely cover the titration region of a given group.

The intrinsic proton affinity of the different titratable groups in a protein can be simulated by the use of experimental data for appropriate model com-

Table 1. Values[a] of pK_{int} of the ionizable groups and coordinate determination of the point charges

Group	pK_{int}	Coordinates[b]
asp	4.0	(OD1 + OD2)/2
glu	4.5	(OE1 + OE2)/2
his[c]	6.3/10.8	(NZ1 + NE2)/2
lys	10.4	NZ
tyr	10.0	OH
arg	12.0	NH
cys	9.0	SH

[a] The values of pK_{int} are taken from experimental data selected by Matthew (1985)
[b] Atom name designation is made according to Brookaven Protein Data Bank nomenclature (Bernstein et al. 1977)
[c] The second value is for the deprotonation at alkaline pH

pounds. Such an approach can be supported by the fact that practically all ionizable groups are accessible to the solvent (Matthew 1985), i.e. the polarity of the vicinity of each charged group should not be dramatically different from that of the model compounds in water solution. This approach has already been used for investigations of charge-charge interactions (Shire et al. 1974; Matthew and Richards 1983; Matthew 1985). In the present study, the experimentally obtained pK values for the ionization of different model compounds are used as pK_{int} values for the corresponding types of titratable groups in the protein. This ad hoc approximation satisfies the needs of this investigation since the positions of the equilibrium constant on the pH-scale are needed rather than the exact values of the self energy changes. The set of intrinsic pK values for the different types of titratable groups is gives in Table 1.

Influence of charge-charge interactions on pK

At a given pH, an individual titratable group may be in its charged or neutral form. When the pH is changed, the protein charge constellation is also changed and hence the characteristics depending on the electrostatic interactions. The influence of the charge constellation on the pK values of the individual titratable groups is calculated by the iterative procedure described below. Following Tanford and Roxby (1972), the charges $z_i(pH)$ of each ionizable group i can be calculated from the Henderson-Hasselbach Eq.:

$$lg[\alpha_i^{(k)}/(1-\alpha_i^{(k)})] = pH - pK_i^{(k)}, \quad (1)$$

where α_i is the degree of dissociation of group i and (k) indicates the kth iteration step. The charge values are then $z_i(pH) = -\alpha_i$ for acidic and $z_i(pH) = 1-\alpha_i$ for basic groups. The influence of the protein charge distribution on the pK_i is taken into account through:

$$pK_i'^{(k)} = pK_{int,i} + \Delta pK_i^{(k)}. \quad (2)$$

The electrostatic correction term in (2) is:

$$\Delta pK_i^{(k)} = -\frac{1}{2.3\ kT} \sum_{j \neq i} z_j^{(k-1)}(pH)\, e_j\, \sigma_{ij}\, W_{ij}, \quad (3)$$

where W_{ij} is the function of electrostatic interactions between the charges i and j and σ_{ij} is a solvent accessibility correction factor described below. The values of pK_i are finally obtained by averaging the $pK_i'^{(k)}$ with the $pK_i^{(k-1)}$ for convergence reasons (see below):

$$pK_i^{(k)} = (pK_i^{(k-1)} + pK_i'^{(k)})/2. \quad (4)$$

The iterative procedure starts with a set of initial values of $pK_i^{(0)} = pK_{int,i}$ and $\Delta pK_i^{(1)} = 0$. In the first step, the charge set $z_i^{(1)}$ is calculated at $pK_i^{(1)} = pK_{int,i}$ in (1). The charge values z_i are then substituted in (3) to calculate the correction term $\Delta pK_i^{(2)}$, and using (2) and (4), the new set of $pK_i^{(2)}$ is obtained. These values are replaced again in (1). The iterations are repeated until the following conditions are satisfied:

$$\Delta_i = \text{abs}(pK_i^{(k)} - pK_i^{(k-1)}) < \varepsilon_1 \quad (5)$$

for every i, and

$$2.3 \sum \Delta_i\, \alpha_i(1-\alpha_i) < \varepsilon_2. \quad (6)$$

The iterative procedure used here differs from that proposed by Tanford and Roxby (1972) in (2) and (4), as well as in the convergence criterion (6). These differences appear to be essential in the cases of simultaneous ionization of a large number of groups. Thus, for instance, more than 60–70 iterations are needed for calculations of the charge set of cytochrome c in the interval pH 10.5–pH 11.5, using Tanford-Roxby's procedure. With the procedure proposed here, the number of iterations is reduced to 3–4. A comparison of the two procedures is given in Fig. 1. The error of the calculations of the net charge of the protein can exceed 2 protonic units if only condition (5) is considered. Using conditions (5) and (6) in parallel the error of the calculations of the net charge is reduced to 0.1 protonic units when the values of ε_1 and ε_2 are chosen as 0.1. Values below 0.1 have negligible effects on the final results.

Calculation of the electrostatic interaction function

The electrostatic interaction function is calculated according to Kirkwood-Tanford theory (Tanford and Kirkwood 1957; Tanford and Roxby 1972) as modified by Shire et al. (1974). One of the problems of the calculation of electrostatic interactions in proteins is the description of the dielectric shape. The finite-difference relaxation method (Warwicker and Watson 1982; Klapper et al. 1986) takes into account the actual shape of the protein. However, the iterative evaluation of the self-consistent ionization behaviour of the

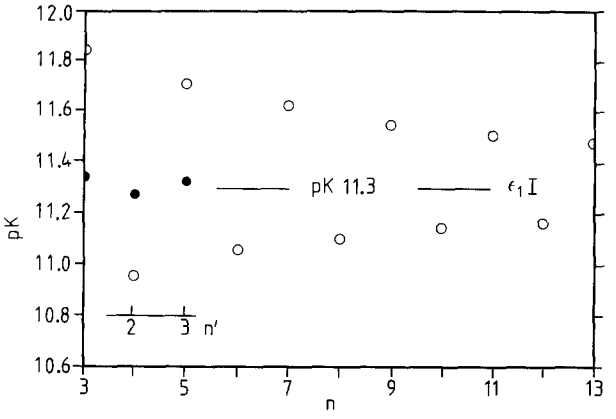

Fig. 1. Comparison of the iterative procedures: Tanford and Roxby (1972) – "o" the pK value at iteration number n; in this work – "•" the pK value at iteration number n'. The example is taken for a lysyl residue of Cytochrome c. Our procedure converges here after 3 iterations, whereas that of Tanford and Roxby (1972) needs more than 70 iterations

titratable groups coupled with this method would need excessive computing time. On the other hand, Gilson and Honig (1988) demonstrated that the change of pK due to charge-charge interactions is well predicted by both the finite-difference relaxation method and the Tanford-Kirkwood model. Thus, the classical Tanford-Kirkwood model seems to be attractive because of its computational economy.

Another problem of the macroscopic treatment of the problem is the evaluation of the dielectric constant of the protein microphase. There are a variety of investigations suggesting values of the dielectric constant between 2 and 50 (Warshel et al. 1984; Gilson and Honig 1986; Nakamura et al. 1988). The theoretical evaluations made by Gilson and Honig (1986) give a value between 2.5 and 4. The high values of the dielectric constant are often used when the calculations are based on Coulomb relation. These effective values are extracted from effects such as the change of pK due to electrostatic interactions (Sternberg et al. 1987; Russell et al. 1987), or reduction of the energy of charge-charge interactions in respect to vacuum (Warshel et al. 1984). In fact, the high effective dielectric constant is a product of models using low permittivity of the protein (Matthew and Gurd 1986; Sternberg et al. 1987). Of course, one can use the Coulomb relation with an effective dielectric constant fitted as average value ($\varepsilon_{\text{eff}} \approx 40$) or as a function of the distance (Warshel et al. 1984; Russell and Warshel 1985). However, there are data which are not in full agreement with such an approximation. Thus for instance, the ε_{eff} evaluated by Sternberg et al. (1987), Russell et al. (1987), and Bashford et al. (1988) cannot be described by the formula for $\varepsilon(r)$ proposed by Warshel et al. (1984); there are values of ε_{eff} above the upper limit of this formula. The large deviation of the values of ε_{eff} suggests that an average value of the dielectric constant also cannot reproduce all experimental data. Availability of dependence of ε_{eff} on the ionic strength clearly shows that it does not reflect properties inherent only to the protein medium. This effect is disunited by the low dielectric models which allows a more detailed analysis.

According to the above considerations, the following simplifications are assumed for the present calculations: *1)* the dielectric constant inside the protein is considerably smaller than that of the surrounding solvent, i.e., $\delta = \dfrac{\varepsilon_p}{\varepsilon_s} \ll 1$ allowing neglect of terms δ^2 and higher order terms, *2)* all the charges are at depth d from the surface of the protein, *3)* the ionic strength is taken into account using the Debye-Hueckel approximation. Thus, for pair-wise interactions

$$W_{ij} = \frac{K_1}{R}\left(\frac{R}{\varepsilon_p r_{ij}} - \frac{1-2\delta}{\varepsilon_p(1+r_{ij}^2/R^2-2\varrho+\varrho^2)^{1/2}} - \frac{1}{\varepsilon_s \varrho} \cdot \ln\frac{(1+r_{ij}^2/R^2-2\varrho+\varrho^2)^{1/2}-1+\varrho+r_{ij}^2/(2R^2\varrho)}{r_{ij}^2/(2R^2\varrho)}\right) - \frac{K_2}{a} C \tag{7}$$

where the constants $K_1 = 331.9$ and $K_2 = 1.4$ are chosen so that W_{ij} is in kcal/mol, R is the protein radius, r_{ij} is the distance between charges i and j in angstroms, $\varrho = (R-d)^2/R^2$, a is the ion-exclusion radius, $C = \sqrt{I}/(1+0.33\,a\sqrt{I})$ is the Debye-Hueckel term and I is the ionic strength. The calculated interaction is modified as proposed by Shire et al. (1974), correcting for the local dielectric environment of the charge group:

$$W_{ij}\,\sigma_{ij} = W_{ij}\,(1-(SA_i + SA_j)/2). \tag{8}$$

The solvent accessibilities SA_i are calculated by the method of Lee and Richards (1971) as A_i/A_t, where A_i is the accessible area of the side chain of residue i in the protein and A_t is the accessible area of the corresponding side chain type in extended conformation. Thus, if two groups are fully accessible to the solvent, $SA_i = SA_j = 1$, they could have an accessible area equal to those of free residues. In this case, the ionization of these groups could be determined only by their intrinsic pK defined for free groups. Conversely, the ionization behaviour of a pair of groups with $SA_i = SA_j = 0$ is determined according (7). All other possible states between these two extreme cases are considered by the linear relationship (8).

Calculation of the electrostatic potential field.

The electrostatic potential field is calculated using Coulomb's law with a simple model for the dielectric con-

stant. The field is calculated at grid points which are characterized only by the values of dielectric constant: inside the protein a value of ε_p is assigned at each grid point, and outside the protein a value of ε_s is assigned. The dielectric environment of the charged atoms is accounted for by using their solvent accessibilities. The potential at grid point k is calculated by

$$\Psi_k = K_1 \sum_i \frac{q_i}{((S_i + G_k)(\varepsilon_s - \varepsilon_p)/2 + \varepsilon_p) r_{ik}}, \quad (9)$$

where q_i is the value of the ith charge obtained through the Kirkwood-Tanford model, S_i ($0 \leq S_i \leq 1$) is the solvent accessibility of the ionizable atom or atomic groups, $G_k = 1$ if the kth grid point is outside the protein and $G_k = 0$ if the kth gridpoint is inside the protein, and r_{jk} is the distance between charge i and grid point k in angstroms. The electrostatic potential is also calculated using the method of finite differences proposed by Gilson and Honig (1988) for comparison.

Modelling the α-helix and peptide dipole moments

It has been pointed out in a number of papers (Hol et al. 198; Hol et al. 1981; Warwicker and Watson 1982) that the electrostatic field of α-helix dipoles, which arise from alignment of backbone dipoles, plays a significant role in the electrostatic interactions in proteins. In the calculations of pK_i, the α-helix dipoles are simulated by placing one half an electronic charge at each end of the α-helix. Thus, according to Hol et al. (1978), effective charges of $+e/2$ and $-e/2$ are attached to the nitrogen atom at the amino terminus and to the carbon atom at the C-terminus, respectively, of α-helices having more than 6 peptide bonds. Our previous calculations have shown that this approximation is sufficient to obtain agreement between calculated and experimental pK values. However, it appears insufficient in some particular cases (discussed below) if a detailed picture of the electrostatic potential field is needed. Then, all the peptide dipoles are taken into account in the calculations of the electrostatic potential field. They are simulated by placing half a positive and negative charge e at coordinates $\boldsymbol{r}^{+e}$ and $\boldsymbol{r}^{-e}$ chosen in such a manner that dipole moment $\mu = 1.2\ 10^{-29}$ Cm (Hol et al. 1978) is obtained:

$$\mathbf{r}^{-e} = \frac{2}{3}\boldsymbol{r}_O + \frac{1}{3}\boldsymbol{r}_N \quad \text{and}$$

$$\boldsymbol{r}^{+e} = \boldsymbol{r}^{-e} + \frac{r_C - r_O}{r_{CO}} \frac{\mu}{0.5}, \quad (10)$$

where $\boldsymbol{r}_O$, $\boldsymbol{r}_N$, and $\boldsymbol{r}_C$ are the coordinates of the oxygen, nitrogen and carbon atoms respectively and r_{CO} is the carbon-oxygen distance. In this approximation the charges remain in the plane of the peptide group and the dipole moment created is parallel to the C–O bond.

Application to proteins

The only input data needed are the atomic coordinates of the protein molecule and a standard list of intrinsic pKs. The set of pK_{int} used here and the manner of determination of the charge coordinates are given in Table 1. The calculations of the solvent accessibilities are made with a counteratom radius equal to those of one water molecule. The dielectric constants chosen are $\varepsilon_p = 2$ and $\varepsilon_s = 80$. Test calculations are carried out for the protein BPTI with coordinates taken from the Brookhaven Protein Data Bank (Bernstein et al. 1977), file 2PTC. The coordinates of sperm whale metmyoglobin obtained by Takano (1977) and of tuna cytochrome *c* obtained by Takano and Dickerson (1981a; 1981b), all available from the Brookhaven Protein Data Bank (Bernstein et al. 1977), were used.

FORTRAN software packages have been developed for calculations of the electrostatic interactions. The finite-difference calculations have been carried out using the appropriate programs of the package DELPHI (Klapper et al. 1986; Gilson and Honig 1988). These programs have been adapted for our purposes. The electrostatic potential output data are converted through the PROTEIN crystallographic package developed in the Max-Planck Institute of Biochemistry in Martinsried (FRG) and displayed on Evans and Sutherland picture systems using FRODO (Jones 1978) graphic facilities.

Results and discussions

Calculations of pK values

The values of pK are calculated on a semi-empirical basis: pK_{int} are taken from experimental data for model compounds and the influence of charge-charge interaction is calculated on the basis of the modified Kirkwood-Tanford macroscopic model. The main goal of these calculations is the adequate description of the ionization state of the protein at a given pH. Test calculations were carried out for BPTI. This protein was chosen because experimental data are available for all ionizable groups (March et al. 1982). In Table 2, the pK values calculated here are compared with the experimental and calculated values obtained by March et al. (1982). There are small differences between the two sets of calculated values as the pK values calculated here are to some extent in better agreement with the experimental results. These differences can originate from the different crystallographic struc-

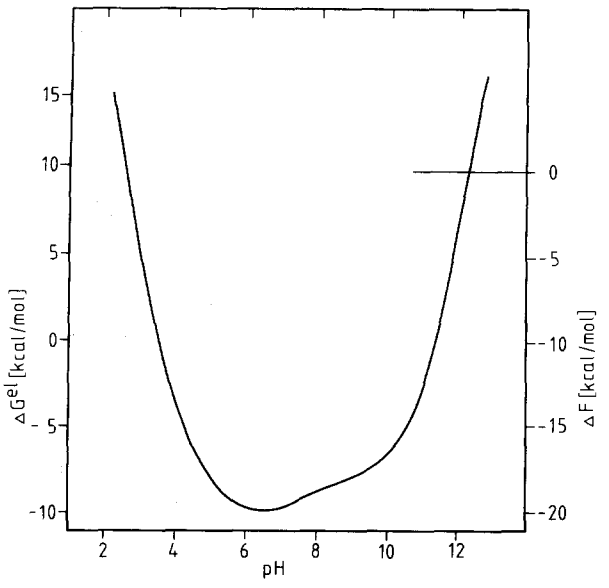

Fig. 2. pH-Dependence of ΔG^{el} compared with the free energy of alkaline denaturation (right) of sperm whale met-hydroxy myoglobin. ΔF is evaluated by shifting the origin along the ordinate so that it crosses ΔG^{el} at the point of denaturation, pH 12, where the electrostatic destabilization interactions are equal to all other interactions which stabilize the native structure. The value of 20 kcal/mol (the right scale) then corresponds to the free energy at alkaline denaturation

Table 2. Values of pK of the titratable groups of BPTI

Group	pK_{calc} (deviation)[a]	pK_{exp}	pK_{calc} (deviation)[a]
N-term	7.94 (0.00)	7.94[b]	7.56[b] (0.38)
Glu 7	4.07 (0.18)	3.89	4.23 (0.23)
Glu 49	4.49 (0.49)	4.00	4.43 (0.43)
Asp 3	3.85 (0.28)	3.57	3.93 (0.36)
Asp 50	3.26 (0.08)	3.18	2.89 (−0.31)
Tyr 10	9.54 (0.08)	9.46	9.94 (0.94)
Tyr 21	10.16 (0.22)	9.94	10.05 (0.11)
Tyr 23	10.96 (−0.04)	11.00	10.25 (−0.75)
Tyr 35	9.20 (−1.4)	10.60	10.22 (−0.38)
Lys 15	10.41 (0.24)	10.43	10.43 (0.28)
Lys 26	10.45 (0.24)	10.10	10.43 (0.33)
Lys 41	11.08 (0.42)	10.60	10.70 (0.10)
Lys 46	10.20 (0.33)	9.87	10.35 (0.48)
C-term	3.06 (0.01)	3.05	3.49 (0.44)

[a] Deviation of the calculated pK from experimental values
[b] Data in this column taken from March et al. (1982)

tures taken as input data and from the simplifications made for the present calculations. A difference between experimental and theoretical values of about 0.3 pK units is observed for the acidic groups. This overestimation probably comes from the empiricism in determining the pK_{int}. Greater differences are obtained for tyrosyl and lysyl residues. It is argued by March et al. (1982) that the molecule undergoes conformational change at alkaline pH. Thus, these differences can be produced by the fixed-atoms approximation used in the present calculations. Nevertheless, this discrepancy is much smaller than the pH range of ionization of given titratable group. It should be pointed out, once again, that what is essential here is the value of pK, but not the energy associated with the ionization a of given group. On the basis of the results discussed above, it can be concluded that the approach can predict sufficiently well the pK values of the titratable groups.

pH-Dependence of the electrostatic energy

The electrostatic energy is one term of the free energy, that is, for the model employed,

$$\Delta G^{el} = 1/2 \sum_{\substack{i,j \\ i \neq j}} q_i q_j W_{ij}(1-(SA_i + SA_j)/2), \qquad (11)$$

where all assumptions described above are also assumed here. As defined, (11) describes the work needed to move a charge from infinity to its place in the presence of all other charges at given pH and ionic strength. It should be pointed out that ΔG^{el} is the energy of interaction only between charged groups in protein. The energy of formation of the individual charges is implicitly included in the empirical values of pK_{int}. Thus, the above expression does not include this energy.

With these considerations we have calculated the pH-dependence of the electrostatic energy of sperm whale myoglobin. The minimal value of ΔG^{el}, −10 kcal/mol, is obtained at pH 6.5 (Fig. 2). The position of the minimum coincides with that obtained by Garcia-Moreno et al. (1985), but its depth is considerably smaller. It has been shown that the midpoint of alkaline denaturation of hydroxymetmyoglobin is at pH 12.2–12.3 (Acampora and Hermans 1967; Hermans and Acampora 1967). At this pH the electrostatic repulsive forces are equal to all other pH-independent interactions which stabilize the native structure. Shifting the origin along the ordinate so that it crosses ΔG^{el} at pH 12.2, the free energy difference reaches 20.0 kcal/mol. This is in good agreement with the experimental results of the free energy change for alkaline denaturation of sperm whale myoglobin: 16–20 kcal/mol (Hermans et al. 1969; Atanasov et al. 1972). This result also confirms the generally shared view that half of the stabilization energy of native proteins is due to the electrostatic interactions. It is seen here that the pH-dependence of the stabilization energy of the protein is mainly due to the electrostatic interactions between titratable groups.

The pH-dependence of ΔG^{el} calculated for ferrous and ferric tuna cytochrome *c* is shown on Fig. 3. The electrostatic energy of ferrous cytochrome *c* shows sta-

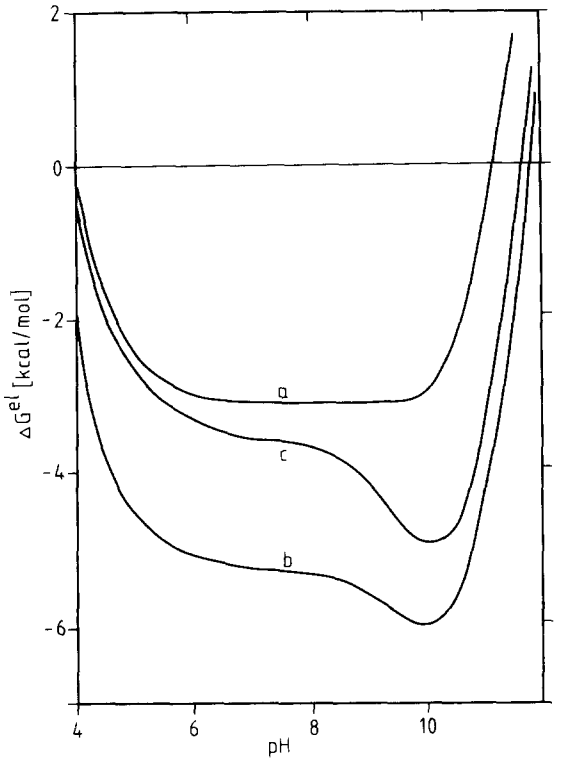

Fig. 3. pH-Dependence of ΔG^{el} for tuna cytochrome *c*: a – ferro; b – ferri, crysttallographic form "*I*"; c – ferri, crystallographic form "*O*". A sigmoidal shape is seen for pH dependence of ΔG^{el} for both "*I*" and "*O*" ferri forms with inflexion point at pH 9.1–9.2, corresponding to the pK of alkaline isomerization

bilization of the native structure due to ionizable groups and is characterized by a wide plateau between pH 5.5 and pH 8.5. The crystals of ferric cytochrome *c* have two molecules per asymmetric unit, designated as "*I*" (inner) and "*O*" (outer) crystallographic forms (Takano and Dickerson 1981 b). The minimum values of ΔG^{el} are between 4 and 6 kcal/mol, which is in good agreement with the measurements of the electrostatic stabilization energy (5.6 kcal/mol) (Scheiter et al. 1979). The pH dependence of the electrostatic energy for both ferric forms shows a pronounced sigmoidal form with a minimum at pH 10. The midpoint of the sigmoidal curvature of ΔG^{el} for both ferric forms is at pH 9.1–9.2, which corresponds to the pK of the III–IV transition typical of these proteins (Davis et al. 1974; Kihara et al. 1976). The change of ΔG^{el} amounts to 0.75 kcal/mol for form "*I*" and 1.4 kcal/mol for form "*O*". These values are about half the experimentally obtained change of the free energy at the III–IV transition (2.1–2.8 kcal/mol) (Davis et al. 1974). No transition is experimentally observed for ferrous cytochrome *c*. Accordingly, no pH-dependence of ΔG^{el} is obtained in the corresponding pH-interval for that form. This result suggests that the electrostatic interactions between ionizable groups probably play a dominant role in the III–IV transition in ferric cytochrome *c*. It is also notable that groups with pK corresponding to the inflexion point of ΔG^{el} are not detected, which suggests that the electrostatic interactions between the charges of the multipole as a whole are responsible for the pH-dependence observed. It was supposed that III–IV transition is connected with a conformational change of the heme crevice regulated by electrostatic interactions between groups situated in its vicinity (Osheroff et al. 1980). According to our results, it can be assumed that the change of the energy of charge-charge interactions in the whole charge multipole may be a key factor of the electrostatic regulation of this conformational change.

Electrostatic potential field of cystatin

The determination of the electrostatic field of a protein is a useful tool which allows the investigation of protein properties from another perspective. It can, for example, provide information about protein-protein recognition and thus processes of complexation and inhibition. For this reason and to compare two different methods of calculation of the electrostatic potential field, we have calculated the electrostatic potential for the cysteine proteinase inhibitor cystatin; the results are discussed in this section.

Egg white cystatin is a small protein which inhibits papain-like cysteine proteinases, binding reversibly and stoichiometrically. On the basis of the known structures of cystatin and papain we have recently suggested a model for the complex of these proteins (Bode et al. 1988). The electrostatic interaction energy of the model complex was calculated using GROMOS (WF van Gunsteren, HJC Berendsen; BIOMOS, University of Groningen, 9747 AG Groningen, The Netherlands) which calculates the electrostatic interaction energies using Coulomb's law and partial charges assigned to all atoms. The energies were found to be favorable in all cases; all variations of the model, before and after energy minimization, possessed attractive interactions. We then calculated the electrostatic potential field of both proteins to clarify the details of the electrostatic interactions. The electrostatic potential was calculated using two different computing techniques: *1*) based on the Coulomb relation (9) including peptide dipoles and the charges of the ionizable group obtained according to the Kirkwood-Tanford theory (see Methods), and *2*) the finite-differences method (Klapper et al. 1986; Gilson and Honig 1988).

The electrostatic potential field calculated by the first method at pH 7 and ionic strength 0.1 M is illustrated in Fig. 4. The net charge of the protein is −0.25 protonic units. Outside the protein, it has a form approximately similar to that of a dipole, with the positive pole situated at the putative binding site and the

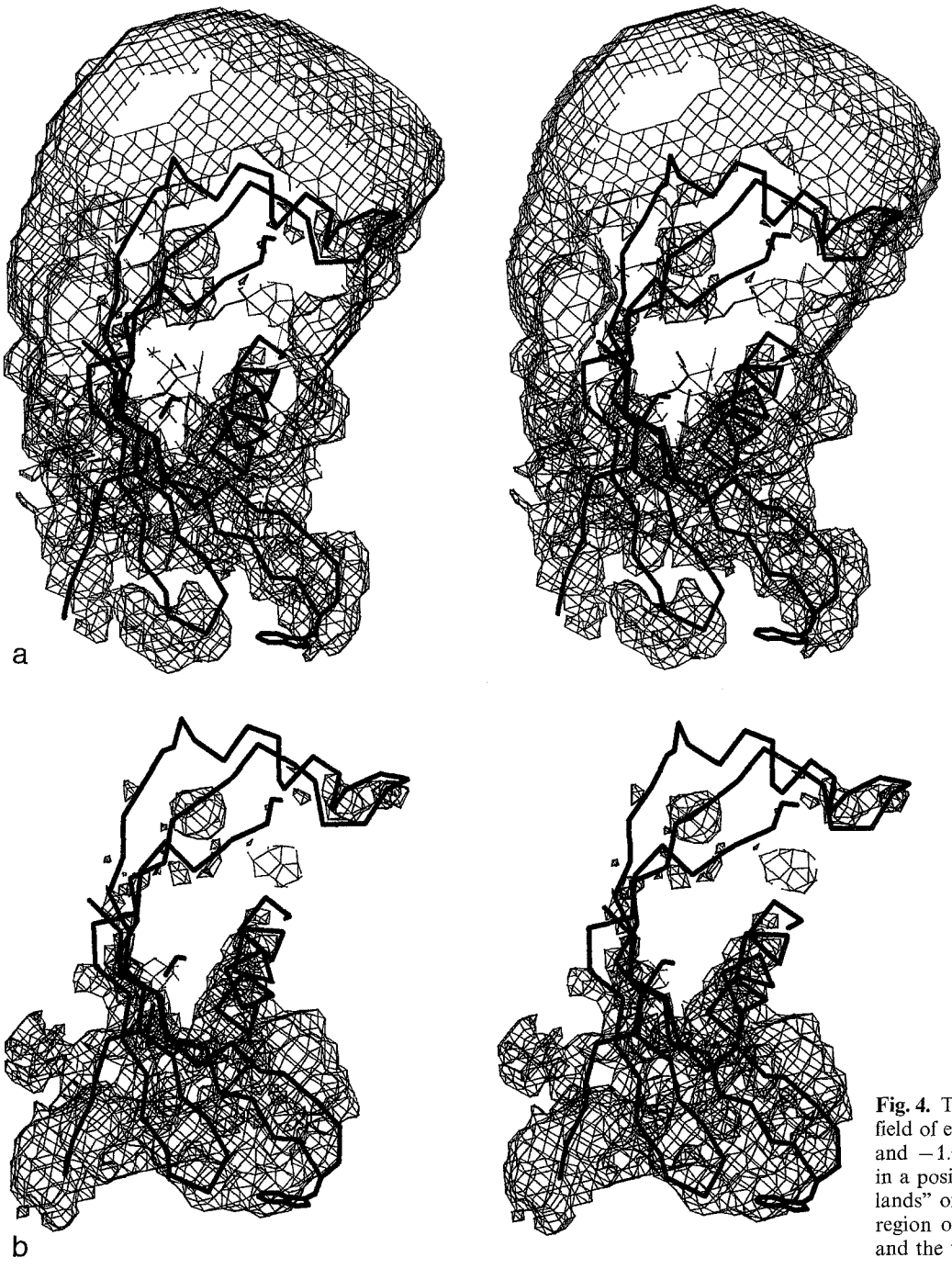

Fig. 4. The electrostatic potential field of egg white cystatin at 1.0 (**a**) and −1.0 (**b**) kcal/mol. The "edge" in a positive potential (**b**), with "islands" of negative potential in the region of the two β hairpin loops and the tryptophan residue (**a**)

negative pole at the region of the two α-helices and the C-terminal group. The positive potential occurs at the space around the N-terminal chain up to Val 14, including also the "edge region" of the protein (the two β-hairpin loops; at the bottom of Fig. 4b). Another positive region of the electrostatic potential field is formed by the cluster of basic groups Lys 31, Arg 68, and Lys 91 (top of Fig. 4b). The influence of the peptide dipoles is visible at the two α-helices where positive regions are seen. In contrast to the relatively smooth and widely distributed negative region, the positive region is perturbed with a number of negative "islands". Two of them are separated in the region of the two β-hairpin loops Gln 53 – Ile 58 and Ile 102 – Gln 197 and are caused by the corresponding carbonyl groups. It is of interest to note that this is the region of cystatin in contact with the binding site of papain (Bode et al. 1988), which is also positively charged.

The electrostatic potential field of cystatin was also calculated using the finite-difference algorithm (Klapper et al. 1986; Gilson and Honig 1988). In this case the CPU time amounted to 5.25 h on a VAX/8550

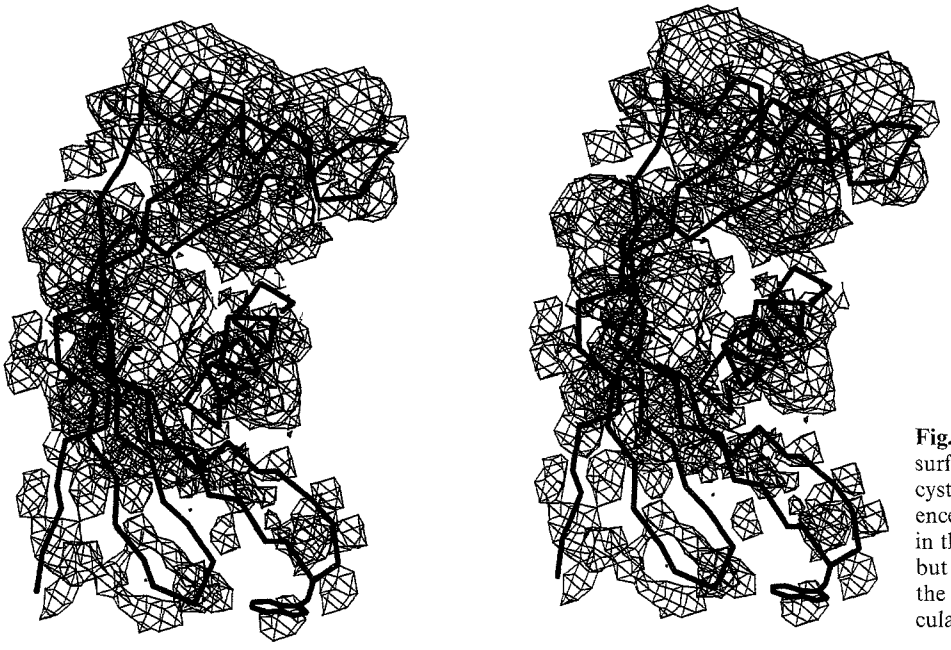

Fig. 5. The electrostatic isopotential surface at −1.0 kcal/mol of egg white cystatin calculated by the finite-difference method. The potential is reduced in the region of the protein main body, but essentially similar in the region of the protein "edge" compared to the calculations basis on (9)

compared with 20 min needed by the calculations carried out with (9). Here, we have included in the calculations all atomic partial charges and the titratable groups are taken in their charged form except for the histidine residue which has 0.5 protonic units. Similar conditions are usually used in calculations of the electrostatic term of energy minimization computations, in molecular dynamics methods, and in a number of analyses of the electrostatic interactions in macromolecules, where the titratable groups are all either neutral or in their charged forms. The net charge of the molecule amounts in this case to 0.6 protonic units using CHARMM parameters available with the DELPHI program package (Gilson and Honig 1988). The negative potential in the region around the main body of the molecule is smaller (Fig. 5). On the other hand it is seen that the two fields in the presumed contact region of the molecule with papain are very similar. The negative "islands" are essentially unchanged, and the indole ring of Trp 104 is now in a region of negative potential. The difference between zero-potential surfaces are also negligible (not greater than one atomic radius) inside the protein. The lack of ionizable groups in the region of the "edge" of the protein suggests that the electrostatic potential is here mainly determined by the partial charges. This result also shows that the peptide dipoles are primarily responsible for the electrostatic potential distribution in this region. Since the potentials differ where there are many titratable groups but are similar in the region of the "edge" of the molecule, one can conclude that the negative potential obtained with (9) is broader due to the total contribution of the charges composing a negative charge multipole, i.e. due to the protein ionization state determined by the value of pH chosen and is not inherent to the model.

One of the principal motivations for calculating the electrostatic potential field of cystatin was to acquire insight and possibly new information regarding the specific mode of interaction between cystatin and cysteine proteinases such as papain. A naive expectation that the electrostatic field may be involved in orienting the inhibitor during its approach to the proteinase was not born out; in fact, the model, proposed on other convincing evidence, indicates that the positive end of cystatin is in contact with the positively charged active site of papain. However, the localized regions of negative potential along putative binding edge of cystatin indicate the reason for the favorable interactions calculated by GROMOS; namely that the carbonyl oxygens, modelled by the peptide dipoles, may form hydrogen bonds with donors in the papain active site. Furthermore, the apparent barrier to the approach of cystatin is relatively small (less than kT) in the physiological pH range. Thus in this case where the interactions are mainly due to the peptide dipoles (pH-independent), the electrostatic field apparently has functional significance only in the contact region.

The calculations made in this work suggest that the use of simple models could also be useful for calculations of electrostatic interactions in proteins. The electrostatic potential field at the contact "edge" of the cystatin with papain is determined mainly by the peptide dipoles of the hairpin β-turns. The difference between the potentials obtained by two methods in the region of the main body of the molecule suggests that

in order to obtain an adequate description of the electrostatic potential the ionization state of the titratable groups should be carefully considered. The calculations of the electrostatic potential field show that the simplifications adopted in this work allow rapid computation with little significant loss in accuracy. Calculations based on (9) can be used for a preliminary analysis of the electrostatic potential field. Applied to myoglobin and cytochrome *c*, this model predicts the experimental results for the energies and pH dependent conformational transitions which suggests that it also enables a rapid analysis of the electrostatic interactions and pH-dependent properties of proteins.

Acknowledgement. This work was partially supported by grant 549, National Council of Sciences, Bulgaria. A. Karshikoff thanks the Alexander-von-Humboldt Stiftung for their support during his stay at the Max-Planck-Institute of Biochemistry, West Germany and thanks also Dr. John Sack (Howard Hughes Med. Inst. Houston, Texas 77030) for his help in adapting FRODO for the purpose of this work. R. Engh is supported by a grant from the United States NSF.

References

Acampora G, Hermans J Jr (1967) Reversible denaturation of sperm whale myoglobin. I. Dependence on temperature, pH, and composition. J Am Chem Soc 89:1543–1546

Atanasov BP, Privalov PL, Chetchinashvili HH (1972) Calorimetric study of the heat denaturation of cyanometmyoglobin. Mol Biol (USSR) 6:33–41

Bashford A, Karplus M, Casten GW (1988) Electrostatic effects of charge perturbations introduced by metal oxidation in proteins. A theoretical analysis. J Mol Biol 203:507–510

Bernstein FC, Koetzle TF, Williams GJB, Meyer EF jr, Brice MD, Rodgers JR, Kennard O, Shimanuochi T, Tasumi M (1977) The protein data bank: a computer based archival file for macromolecular structures. J Mol Biol 112:353–342

Bode W, Engh R, Musil D, Huber R, Karshikov A, Brzin J, Kos J, Turk V (1988) The 2.0 Å X-ray crystal structure of chicken egg white cystatin and its possible mode of interaction with cysteine proteinase. Embo J 7:2593–2599

Davis LA,Schejter A, Hess GP (1974) Alkaline isomerization of oxidized cytochrome c. J Biol Chem 249:2642–2632

Friend SH, Gurd FRN (1979 a) Electrostatic interaction in myoglobin. pH dependence of summed electrostatic contribution. Biochemistry 18:4612–4619

Friend SH, Gurd FRN (1979 b) Electrostatic interaction in myoglobin. Interactive free energies between individual sites. Biochemistry 18:4620–4630

Garcia-Moreno BE, Chen LX, March KL, Gurd RS, Gurd FRN (1985) Electrostatic interactions in sperm whale myoglobin. Site specificity, roles in structural elements and external electrostatic potential distributions. J Biol Chem 260:14070–14082

Gilson MK, Honig BH (1988) Energetics of charge-charge interaction in proteins. Proteins: structure, function, Genet 3:32–52

Hermans J jr., Acampora G (1967) Reversible denaturation of sperm whale myoglobin. II. Thermodynamic analysis. J Am Chem Soc 89:1547–1552

Hermans J, Puett D, Acampora G (1969) On the conformation of denaturated proteins. Biochemistry 8:22–29

Hol WGJ, van Duijningen PT, Berendsen HJC (1978) The α-helix dipole and the properties of proteins. Nature 273:443–446

Hol WGJ, Halie LM, Sander C (1981) Dipoles of α-helix and β-sheet: their role in protein folding. Nature 294:532–536

Jones A (1978) A graphic model building and refinement system for macromolecules. J Appl Crystallogr 11:268–272

Kihara H, Saigo S, Nakatani H, Hiromi K, Ikeda-Saito M, Iizuka T (1976) Kinetic study of isomerization of ferricytochrome c at alkaline pH. Biochim Biophys Acta 430:225–243

Kirkwood JG (1934) Theory of solutions of molecules containing widely separated charges with spatial application to zwitterions. J Chem Phys 2:351–361

Klapper I, Hagstrom R, Fine R, Sharp K, Honig B (1986) Focusing of electric field in the active site of Cu-Zn superoxide dismutase: Effect of ionic strenght and amino-acid modification. Proteins: structure, function, Genet 1:47–59

Lee B, Richards FM (1971) The interaction of protein structures. Estimation of static accessibility. J Mol Biol 55:379–400

March KL, Maskalik DG, England RD, Friend SH, Gurd FRN (1982) Analysis of electrostatic interactions and their relationship to conformation and stability of bovine pancreatic trypsin inhibitor. Biochemistry 21:5241–5251

Matthew JB (1985) Electrostatic effects in proteins. Ann Rev Biophys Biophys Chem 14:387–417

Matthew JB, Gurd FRN (1986) Calculation of electrostatic interactions in proteins. Methods Enzymol 130:413–453

Matthew JB, Richards FM (1983) The pH dependence of hydrogen exchange in proteins. J Biol Chem 258:3039–3044

Matthew JB, Friend SH, Botelho LH, Lehman LD, Hanania GIH, Gurd FRN (1978) Discrete charge calculations of potentiometric titration for globular proteins: sperm whale myoglobin, hemoglobin alpha chain, cytochrome *c*. Biochem Biophys Res Commun 18:416–421

Nakamura H, Sakamoto T, Wada A (1988) A theoretical study of the dielectric constant of proteins. Protein Eng 2:177–183

Osheroff N, Borden D, Koppenol WH, Margoliash E (1980) Electrostatic interactions in cytochrome *c*. The role of interactions between residues 13 and 90 and residues 79 and 47 in stabilizing the heme crevice structure. J Biol Chem 255: 1689–1697

Perutz MP (1978) Electrostatic effects in proteins. Science 201:1187–1191

Russell ST, Warshel A (1985) Calculations of electrostatic energies in proteins. The energetics of ionized groups in bovine pancreatic trypsin inhibitor. J Mol Biol 185:389–404

Russell AJ, Thomas PG, Fersht AR (1987) Electrostatic effects on modification of charged groups in the active site cleft of subtilising by protein engineering. J Mol Biol 193:803–813

Salemme FR (1976) A hypothetical structure for an intermolecular electrostatic complex of cytochrome *c* and b_5. J Mol Biol 102:563–568

Schejter A, Zuckerman M, Aviram I (1979) The contribution of electrostatic fators to the stabilization of the conformation of cytochrome *c*. J Biol Chem 254:7042–7046

Shire SJ, Hanania GIH, Gurd FRN (1974) Electrostatic effect in myoglobin. Hydrogen ion equilibria in sperm whale ferrimyoglobin. Biochemistry 13:2967–2974

Sprang S, Standing T, Fletterick RJ, Stroud RM, Finer-Moore J, Xuong N-H, Hamlin R, Rutter WJ, Craik CS (1987) The three-dimensional structure of Asn^{102} mutant of trypsin: Role of Asp^{102} in serine protease catalysis. Science 237:905–909

Sternberg MJE, Hayes FRN, Russell AJ, Thomas PG, Fersht AR (1987) Prediction of electrostatic effects of engineering of protein charges. Nature 330:86–88

Takano T (1977) Structure of myoglobin refined at 2.0 A resolution I. Crystallographic refinement of metmyoglobin from sperm whale. J Mol Biol 110:537–568

Takano T, Dickerson RE (1981 a) Conformation change of cytochrome *c*. I. Ferrocytochrome *c* structure refined at 1.5 A resolution. J Mol Biol 153:79–94

Takano T, Dickerson RE (1981 b) Conformation change of cytochrome *c*. I. Ferricytochrome *c* refinement at 1.8 A and comparison with the ferrocytochrome structure. J Mol Biol 153:95–115

Tanford Ch, Kirkwood JG (1957) Theory of protein titration curves. I. General equation for impenetrable spheres. J Am Chem Soc 79:5333–5339

Tanford Ch, Roxby R (1972) Interpretation of protein titration curves. Application to lysozyme. Biochemistry 11:2192–2198

Warshel A (1978) Energetics of enzyme catalysis. Proc Natl Acad Sci USA 75:5250–5254

Warshel A, Russell ST (1984) Calculations of electrostatic interactions in biological systems and in solutions. Rev Biophys 17:283–422

Warshel A, Russell ST, Churg AK (1984) Macroscopic models for electrostatic interactions in proteins: limitations and applicability. Proc Natl Acad Sci USA 81:4785–4789

Warwicker J, Watson HC (1982) Calculation of the electric potential in the size cleft due to α-helix dipoles. J Mol Biol 157:671–679

Warwicker J, Engelman BP, Steitz TA (1987) Electrostatic calculations and model building suggest that DNA bound to CAP is sharply bent. Proteins: structure, function, Genet 2:283–289