Model for calculation of electrostatic interactions in unfolded proteins

P. J. Kundrotas and A. Karshikoff
Department of Biosciences at Novum Research Park, Karolinska Institutet, SE–141 57 Huddinge, Stockholm, Sweden
(Received 16 May 2001; published 11 December 2001)

An approach for the calculation of electrostatic interactions and titration properties of unfolded polypeptide chains (denatured proteins) is proposed. It is based on a simple representation of the denatured proteins as a state with titratable sites distributed on the surface of a sphere, radius of which is assumed to be equal to the radius of gyration, R_g, of an unfolded molecule. Distances between the charges, d, obey constraints arising from the protein sequence. Criteria for evaluation of the parameters R_g and d were obtained from computer simulations on a polypeptide consisting of 20 identical amino acids (polylysine). The model was applied for calculation of titration curves of denatured barnase and staphylococcal nuclease. It was demonstrated that the approach proposed gives considerably better agreement with the experimental data than the commonly used null approximation. It was also found that titration properties of denatured proteins are slightly, but distinguishably influenced by the amino-acid sequence of the protein.

DOI: 10.1103/PhysRevE.65.011901 PACS number(s): 87.10.+e, 82.60.Lf

I. INTRODUCTION

The three-dimensional structure of protein results from a delicate interplay of a number of interactions, mainly noncovalent in character. Among them electrostatic interactions are of special interest. Their role becomes manifested in any *p*H-dependent properties of proteins, including structural stability. It is obvious, for instance, that acid denaturation of proteins is driven mainly by the reduction of electrostatic stabilization. Also, there is plenty of evidence that electrostatic interactions are a key factor for the thermal tolerance of proteins from hyperthermophilic organisms. These proteins preserve their native structure at temperatures close to, and some of them even at higher than that of boiling water. For all these, an essential increase of salt bridge number is observed (for review, see, e.g., Ref. [1]).

Electrostatic interactions cannot be measured directly. Therefore, their theoretical prediction is of key importance for the correct understanding of protein stability. Usually, theoretical efforts are focused on prediction of measurable quantities that directly depend on electrostatic interactions. These are, for instance, the protonation/deprotonation equilibrium constants of the titratable groups in proteins measured by NMR. Efforts towards theoretical prediction of these constants (or their equivalent, the *p*K values) are mainly concentrated on native proteins and studies aimed at analysis of electrostatic properties of unfolded proteins are scarce.

Most often, the denatured state is modeled by means of the null approximation (NA), where the electrostatic interactions are set to zero and the titratable groups are characterized by *p*K values of amino acids with the alpha amino and carboxyl groups blocked by nontitratable residues. The *p*K values of such model compounds can be obtained both experimentally and from quantum chemical calculations. This makes NA a convenient reference state for *p*K calculations of native proteins (see, for review Ref. [2]). NA becomes, however, nonapplicable for prediction of other quantities, such as the electrostatic term of unfolding energy, because the assumption for zero electrostatic interactions does not hold [3]. The fact that electrostatic interactions in denatured state cannot be neglected has been clearly demonstrated by Fersht and co-workers in a series of studies [4–6]. These authors have demonstrated that the *p*K values of the acidic groups in barnase are on average by 0.4 *p*H units lower than those of NA.

Sheafer *et al.* [7] used an extended conformation for the calculation of electrostatic interactions in denatured state. In this model, the titratable groups are characterized by maximum solvent accessibility, which corresponds to the commonly accepted assumption that titratable groups are fully hydrated in denatured state. A similar model has been proposed by Warwicker [8]. Denatured protein is again represented by an extended conformation, but locations of the ionizable sites with respect to molecule backbone are adjusted by additional parameters. This model has successfully been applied for calculation of the *p*H-induced denaturation of a synthetic leucine zipper [9]. Yang and Honig [10] have employed a hybrid approach to analyze the influence of *p*H and of the ionic strength on stability of sperm whale apomyoglobin. The unfolded state of this molecule has also been modeled by NA whereas intermediate states of the protein have been described using a combination of the *p*K values for NA and for the native state.

In spite of the good agreement with the experimental data, all these models are designed to solve specific tasks. A more general approach for modeling of the denatured state has been proposed by Elcock [11]. He has shown that employing a simple molecular mechanics protocol that uses the native state as a starting point, an essential improvement in prediction of *p*H effect on protein stability can be achieved. The key point of the model is the artificial ''swelling'' of the protein molecule by increasing of the noncovalent interatomic distances. He has found that the best agreement with the experimental data is obtained with a distance of 6 Å for all pairs of interacting atoms.

A disadvantage of all above models is that only one possible spatial distribution of protein charges is considered, which may not be representative for an unfolded protein. In the present paper we propose a simple, but physically more

realistic model of the denatured state of proteins. It is based on the assumption that titratable sites adopt a quasirandom distribution on the surface of a dielectric sphere with a radius equal to the radius of gyration, R_g, of denatured protein. A restricting condition is applied to charge-charge distances, which are determined by the protein sequence. The rest of the paper is organized as follows. In Sec. II we introduce a model of the denatured state and describe computer simulations on an artificial polypeptide chain, simulations that provide a simple way for evaluation of R_g and charge-charge distance criterion, *d*. Section III provides computational details for electrostatic energy calculations and the Monte Carlo (MC) method used. Section IV is devoted to the application of the model for calculation of titration curves of two denatured proteins, barnase and staphylococcal nuclease, for which experimental data are available. The results are discussed in the context of the quality of the model. Finally, in Sec. V some conclusions are drawn.

II. MODEL

A. Basic concepts

In general, there are two approaches for the calculation of electrostatic interactions in proteins. The most rigorous method, elaborated by Warshel *et al.* [12,13], considers a protein molecule on the atomic level. An alternative approach is the continuum dielectric model. In this approach the protein molecule is represented as a continuum material with dielectric constant ε_p immersed in the medium of the solvent with $\varepsilon_s > \varepsilon_p$. The shape of the ''dielectric cavity,'' as well as loci of ionisable (titratable) sites are determined by the three-dimensional structure of the protein molecule. The Poisson-Boltzmann equation is solved for this system numerically [14–17]. This model is attractive with its conceptual simplicity and the limited number of parameters. Moreover, it is proven that in general this model correctly describes electrostatic interactions in proteins. These features of the continuum dielectric model motivated us to employ it for electrostatic studies in denatured proteins as well.

Our model of the denatured state is based on two fundamental assumptions. The first one is related to the fact that a denatured protein can be considered as an average over all possible conformations of a flexible chain, which results in a sphere inside which most of the protein atoms reside. The radius of this sphere is assumed to be equal to the radius of gyration, R_g, of a protein in its denatured state. The dielectric constant ε_p inside the sphere is a general parameter in the current calculations and may have any value between ε_p^N of the native protein and ε_s.

The second basic assumption is that the titratable sites of a denatured protein in equilibrium are located on the surface of the sphere. This assumption is justified by the fact, that because of differences in the self-energies an unconfined charge is expelled from protein interior (low dielectric media) towards solvent (high dielectric media) [13]. Since charges belong to protein moiety and due to flexibility of a polypeptide chain, it is plausible to assume that titratable sites of a denatured protein in equilibrium tend to be located on the dielectric boundary.

B. Radius of gyration

The radius of gyration, R_g, of a molecule is determined as (Ref. [18], p. 151)

$$R_g = \sqrt{\frac{1}{\sum_k m_k} \left\langle \sum_k m_k r_k^2 \right\rangle}, \tag{1}$$

where the summation is performed over all molecule atoms, whereas the average is performed over all molecule conformations, r_k is the distance between kth atom and the center of the mass for a given conformation of the flexible chain, and m_k is the mass of kth atom. In general, due to changes in conformational population, R_g of protein molecules is not a constant. It depends on *p*H, ionic strength, temperature, etc. For simplicity, however, R_g in the present paper is considered as a constant.

Direct calculation of R_g from the Eq. (1) is a computationally cumbersome task demanding explicit consideration of all protein atoms, covalent bonds, and angles for a large number of conformations. As far as the model treats the protein moiety as a continuum medium, it is highly desirable to find an expression for R_g, which depends on as few parameters as possible. For this purpose we carried out computer simulations on a short polypeptide chain, constructed from $L_r = 20$ lysines (polylysine). All side chains were kept in their extended conformations, whereas the backbone was free to rotate around N—C_α and C_α—C bonds (φ and ψ angles, respectively). The lysine side chain was chosen for the calculations because positions of titratable sites for other groups in their extended conformations can be approximately considered as coinciding with positions of the atoms in the extended lysine (for structures of side chains and notation of atoms see, e.g., Ref. [19], p. 6). We generated 40 000 self-avoiding conformations of the polylysine with L_r ranging from 2 to 20 by randomly choosing values of $\varphi, \psi \in [0,360°]$ for each peptide unit. For every configuration generated the coordinates of the C_β, C_γ, C_δ, C_ϵ, and N_ζ atoms were recorded for all lysines (for the notations of the lysine atoms see insert in Fig. 1) as well as the coordinates of the C and N terms. Relatively moderate amount of conformations considered, as it will be shown below, was sufficient to achieve our goals.

The radius of gyration calculated from Eq. (1) as a function of L_r (assuming $m_i = 1$) is shown in Fig. 1. For comparison, the distances h between the N_ζ atoms belonging to the terminal residues of the polylysine are also shown in the figure. Statistical noise is not observed in both plots which indicates that the amount of conformations generated is sufficient to obtain reliable information about the quantities of interest. The generation of random conformations of the polylysine chain can be viewed as a random walk of a chosen atom, e.g., of the N_ζ atom of the first residue. Indeed, both R_g and h obey for $L_r \geqslant 5$ the square-root law $y^2 = a_0 + a_1 \times L_r$. Also, the ratio of the coefficients a_1 for R_g and h gives the value 5.89, which is close to the value given by the fundamental relation $h^2 = 6 \times R_g^2$ derived by Tanford (p. 167 in Ref. [18]) for an ideal flexible chain (in this case h stands

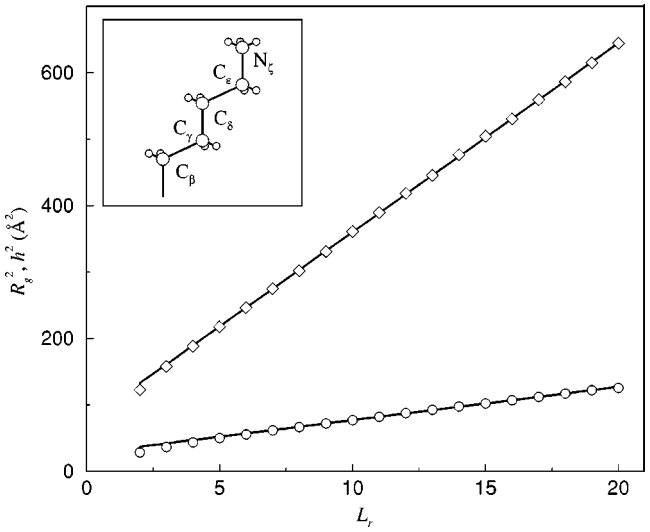

FIG. 1. Radius of gyration R_g calculated from Eq. (1) (circles) and average distance h between the N_ζ atoms (diamonds) for the polylysine chain as a function of number of polypeptide units L_r. Data points presented in the plot were obtained as an average over 40 000 random conformations of the polylysine chain. Solid lines are functions $y=a_0+a_1\times x$ with $a_0=27.22$ and $a_1=5.00$ for R_g and $a_0=76.09$ and $a_1=28.43$ for h. Inset shows schematic representation of the lysine side chain in extended conformation. Large circles denote carbon and nitrogen atoms while small circles stand for the hydrogen atoms. Notations of the atoms in the side chain are adopted from Ref. [20].

for an average end-to-end distance of a flexible chain). It is noteworthy that this ratio is insensitive of what lysine atom the distance h refers to. Thus, we suggest that R_g for unfolded proteins can be evaluated from relation

$$R_g=\sqrt{27.22+5\times L_r}, \quad (\text{Å}), \tag{2}$$

where L_r is the total number of residues in a protein molecule ($L_r \geqslant 5$). The coefficients in Eq. (2) were obtained from a least-squares linear fit to the R_g data presented in Fig. 1. These coefficients are independent of the internal structure of a given molecule.

In Table I the values of R_g calculated by Eq. (2) are compared to experimental values for proteins of different sizes in their denatured state. The calculated R_g values show a tendency of overestimation, which may result from the fact that Eq. (2) was obtained from pure geometrical considerations and does not take into account factors arising from secondary structure elements that at certain conditions may be present in denatured proteins. However, the magnitude of this overestimation is less then the variation of the experimental data and one can consider Eq. (2) as giving a reliable evaluation of R_g. This is also supported by the comparison of our calculations with the values obtained by another theoretical method given at the bottom of Table I.

C. Distribution of charges

In order to take into account a possible influence of protein sequence on the electrostatic interactions in denatured state we introduce a parameter

TABLE I. Comparison of the radius of gyration R_g calculated by Eq. (2) with both experimental measurements and the lattice chain-growth algorithm for various denatured proteins. Experimental values are taken, wherever possible, for the maximally unfolded state of a protein studied in the corresponding reference.

Protein molecule	L_r	R_g (Å) Eq. (2)	R_g (Å) Literature	Ref.
	Experiment			
Bovine pancreatic trypsin inhibitor	56	17.5	11–15	[21]
Streptococcal protein *G*	56	17.5	23±1	[22]
Horse heart cytochrome *c*	104	23.4	30.1	[23]
			17.7	[23]
α-lactalbumin	123	25.3	30	[24]
Hen egg-white lysozyme	129	26.0	22.1–23.5	[25]
			22	[26]
			24.9	[27]
			21–22	[28]
Bovine β-lactoglobulin	162	28.9	35	[29]
Streptomyces subtilisin inhibitor	226	34.0	29.3–29.8	[30]
DnaK	777	62.6	55±1.5	[31]
Yeast phosphoglycerate kinase	1660	91.2	66	[29]
			89±4	[32]
	Lattice-chain growth algorithm			
Cd-7 Metallothionein-2	30	13.31	≈14	[33]
434 Repressor	63	18.5	≈20	
434 Cro	65	18.77	>20	

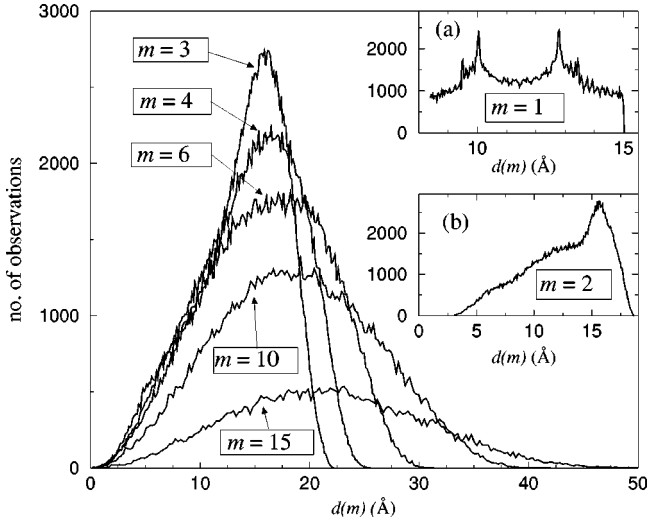

FIG. 2. Histograms of $d(m)$ [formula (3)] for the N_ζ atoms of the polylysine chain for several m values. For clarity, cases $m=1$ and $m=2$ are shown separately in insets (a) and (b), respectively.

$$d(m)=\frac{1}{L_r-m}\sum_{l=1}^{L_r-m}\langle d_{l,l+m}\rangle, \tag{3}$$

which represents a statistically expected distance between two ionizable sites separated by $(m-1)$ nontitratable peptide units along the protein sequence. In the expression (3) index l enumerates peptide units along the sequence, $\langle d_{l,l+m}\rangle$ is the average distance between titratable sites at positions l and $(l+m)$. Values of $d(m)$ for m ranging from 1 to 15 were calculated for the polylysine with $L_r=20$. Histograms of $d(m)$ for the N_ζ atoms and for several m are shown in Fig. 2. Histograms of $d(m)$ obtained for other sidechain atoms do not differ considerably from those presented in Fig. 2. For small m histograms of $d(m)$ differ essentially from those for large m. For $m=1$ two peaks can be distinguished [inset (a) in Fig. 2], which reflect the interplay of the two rotational degrees of freedom, φ and ψ, determining the mutual disposition of the side chain atoms in neighboring peptide units. This effect decays very rapidly and for $m\geqslant 3$ the $d(m)$ histograms are close to a Gaussian distribution characteristic to a completely stochastic quantity. Also, the histogram maxima shift towards larger distances very slowly with increasing m (e.g., for $m=4$ the maximum appears at distance ≈ 18 Å whereas for $m=15$ at ≈ 21 Å). Thus, the distance between two titratable sites, d_τ in the denatured state can be simply related to the sequence separation, m. For simplicity, $d_\tau(m)$ was determined as the distance at which a maximum in the corresponding $d(m)$ histogram appears. In the case $m=1$ a random choice between two possible values was made. For $m>6$ the value of $d(m=6)$ was taken.

Note that for several m we estimated d also for a polypeptide chain with random choice of titratable side chains. As expected, the results do not differ significantly compared to the results for the polylysine. However, such calculations are much more demanding from computational point of view.

III. METHOD

A. Electrostatic energy

In the present paper we employ the ideology of the theory of electrostatic interactions in native proteins (see, for review Ref. [34]). Within this theory, a microscopic state $\mathbf{X}$ of a protein molecule with L_t titratable groups can be determined by a set of deprotonation variables $\mathbf{X}=(x_1,\ldots,x_{L_t})$, with $x_i=0$ $(1\leqslant i\leqslant L_t)$ when ith titratable site is protonated and with $x_i=1$ when it is deprotonated. The electrostatic energy in a state $\mathbf{X}$ is written as

$$\mathcal{E}(\mathbf{X})=\frac{1}{2}\sum_{i,j} W_{ij}(\mathcal{P}_i-x_i)(\mathcal{P}_j-x_j)$$

$$-\ln(10)kT\sum_i (p\mathrm{H}-p\mathrm{K}_i^{\mathrm{int}})x_i, \tag{4}$$

where indexes i and j number titratable groups, $\mathcal{P}_i=1(0)$ for a basic (acidic) group, k is the Boltzmann constant, T is the temperature, and

$$p\mathrm{K}_i^{\mathrm{int}}=p\mathrm{K}_i^0+\Delta p\mathrm{K}_i^{\mathrm{per}}+\Delta p\mathrm{K}_i^{\mathrm{sol}}. \tag{5}$$

Here $p\mathrm{K}_i^0$ is the pK value of the ith residue corresponding to the null approximation (the standard pK), $\Delta p\mathrm{K}_i^{\mathrm{per}}$ and $\Delta p\mathrm{K}_i^{\mathrm{sol}}$ are pK shifts arising from influence of protein permanent charges, such as peptide dipoles, and due to the desolvation penalty, respectively. Derivation of Eq. (4) is given in the Appendix. We applied Eq. (4) for calculation of electrostatic energy in denatured proteins with the only difference that $\mathbf{X}$ in this case is a function not only of the deprotonation variables, but of spatial positions of charges as well.

Because the polypeptide chain is considered highly flexible we neglected the interaction of the titratable site with other charged components, such as peptide dipoles. For simplicity, the influence of the desolvation energy on the protonation/deprotonation equilibrium of the individual site was assumed to be insignificant. It should be noted, however, that the pK change due to desolvation in this case is 0.2–0.3 pH units (see also Ref. [13]). Thus, in the present paper $p\mathrm{K}_i^{\mathrm{int}}=p\mathrm{K}_i^0$ in Eq. (4).

For calculations of pairwise interactions W_{ij} in Eq. (4) we employed a formalism originally developed by Kirkwood and Tanford [18,35,34] for a spherical representation of native protein molecules. Thereupon W_{ij} splits into three parts

$$W_{ij}=W_{ij}^{(1)}+W_{ij}^{(2)}+W_{ij}^{(3)}. \tag{6}$$

The first term in the right-hand side of Eq. (6) is the simple Coulomb interaction

$$W_{ij}^{(1)}=\frac{e^2}{\varepsilon_p r_{ij}}, \tag{7}$$

where r_{ij} is the distance between the charges of the ith and jth titratable sites and e is the elementary charge. The second term in Eq. (6) arises from the fact that protein and surrounding solvent are different dielectric media

$$W_{ij}^{(2)}=\frac{e^2}{\varepsilon_p R_g}\sum_{n=0}^{\infty}\frac{(n+1)(\varepsilon_s-\varepsilon_p)}{(n+1)\varepsilon_s+n\varepsilon_p}P_n(y), \tag{8}$$

where $P_n(y)$ are the Legendre polynomials and $y=1-(r_{ij}^2/2R_g^2)$. The third term in Eq. (6) reflects the influence of ionic strength

$$W_{ij}^{(3)}=\frac{e^2}{\varepsilon_s(R_g+\Delta R)}\left\{\frac{x}{1+x}+\sum_{n=1}^{\infty}\frac{2n+1}{2n-1}\left[\frac{\varepsilon_s}{(n+1)\varepsilon_s+n\varepsilon_p}\right]^2 \times\frac{z^{2n}x^2P_n(y)}{\frac{K_{n+1}(x)}{K_{n-1}(x)}+z^{2n+1}\frac{n(\varepsilon_s-\varepsilon_p)}{(n+1)\varepsilon_s+n\varepsilon_p}\frac{x^2}{4n^2-1}}\right\}, \tag{9}$$

where ΔR is a radius of a salt ion of the solvent, $z=R_g/(R_g+\Delta R)$, $x=\kappa(R_g+\Delta R)$, κ is the Debye-Hückel parameter, and

$$K_n(x)=\sum_{s=0}^{n}\frac{2^s n!(2n-s)!)}{s!(2n)!(n-s)!}x^s.$$

B. Monte Carlo algorithm

In order to reduce computational time a spherical grid was used, i.e., a set of points uniformly distributed on the surface of the sphere. The number of points in the spherical grid, L_p, was chosen so, that a minimum distance between the grid points, $d_{\min}$, is 2–3 Å. This corresponds to minimum distance between the charges of an ion pair in proteins. The polypeptide molecule is represented as an virtual chain with L_t elements that connects the titratable sites only. The bond length connecting two neighboring sites in this virtual chain (not necessarily neighbors in the protein sequence), $d_\tau(m)\pm d_{\min}$, is determined by the distance constraints (Sec. II C). The deviation $d_{\min}$ arises from the discreteness of the spherical grid and does not affect the quality of the statistics and the final results. To each titratable site, *i*, a deprotonation variable x_i is assigned and an initial distribution of the titratable sites on the spherical grid is generated as follows. The first point of the virtual chain is arbitrary placed on the spherical grid and the position of each next point is chosen randomly so that: (i) the virtual chain does not become self-intersecting and (ii) distance from the previous point satisfies the constraint $d_\tau(m)\pm d_{\min}$.

After the virtual chain was generated, MC simulations using the Metropolis algorithm [36] were started with a probability p of transition from a state $\mathbf{X}$ to a state $\mathbf{X}'$ given by $p=\min\{1,\exp(-\Delta\mathcal{E}/kT)\}$ with $\Delta\mathcal{E}$ being the difference in the energies $\mathcal{E}$ [Eq. (4)] between the states $\mathbf{X}$ and $\mathbf{X}'$. When generating $\mathbf{X}'$ from $\mathbf{X}$ either the protonation state or allocation of a given titratable site can be altered. In the case of reallocation of site *i*, the distance constrains with respect to both the $(i-1)$th and the $(i+1)$th sites are taken into account. The choice between the two types of alterations is made randomly. After repeating this procedure L_t times the set $\mathbf{X}$ and the energy $\mathcal{E}$, are calculated and stored. This comprises one Monte Carlo step per site (MCS/S).

All results presented below were obtained as average over 20 independent (i.e., with different initial configuration of the virtual chain) MC runs each of them consisting of 10 000 MCS/S with first 2000 MCS/S being discarded from consideration. The average over independent runs reduces computational time essentially since it extinguishes problems related to a trapping of the system in a metastable state. For instance, a significant scattering of points is observed in a single-residue titration curve obtained from one very long (10^7 MCS/S) MC run while average over 20 relatively short MC runs gives smooth titration curve. Values of other parameters used in calculations are presented in Table II. The outputs, averaged values of deprotonation variables $\langle x_i\rangle$ as a function of pH, are used to calculate an overall titration curve of a protein molecule of interest and the pK values of individual groups (in the present paper the pK value was considered to be equal to the $p\mathrm{K}_{1/2}$ value, i.e., a pH value at which the average deprotonation variable $\langle x_i\rangle=0.5$).

TABLE II. Values of the model parameters used in calculations.

Parameter	Notation	Value
Ion exclusion radius	ΔR, Å	2
Ionic strength	I	0.1
Temperature	T, °C	20
Solvent dielectric constant	ε_s	80
Protein dielectric constant	ε_p	4/40
Standard pK values (from Refs. [37,38])	$p\mathrm{K}_i^0$	4.0 (Asp)
		4.4 (Glu)
		6.3 (His)
		9.1 (Cys)
		9.4 (Tyr)
		10.4 (Lys)
		12.0 (Arg)
		3.67 (*C* term)
		8.20 (*N* term)
Distance constraints (Sec. II C)	$d_\tau(m)$, Å	10 or 12.5 ($m=1$)
		15.0 ($m=2$)
		15.5 ($m=3$)
		17.0 ($m=4$)
		18.0 ($m=5$)
		20.0 ($m\geqslant 6$)

IV. RESULTS AND DISCUSSION

In this section we demonstrate applicability of the model described above for the calculations of pK values of titratable groups in an unfolded protein molecule and of the protein net charge as a function of pH. The correct prediction of these quantities is a prerequisite for the analysis of phenomena such as folding/unfolding processes, pH-dependent stability of proteins, etc. To our knowledge, there are only few proteins for which experimental data on electrostatic properties in denatured state are available. In the present paper

TABLE III. pK values of titratable groups of barnase. Experimental data (except for His18) are taken from Ref. [5]. Note that the experimental pK values for denatured barnase are averaged over all residues of a given type.

Amino acid	pK_i^0	The pK values					
		Experiment		MC calculations			
		Native	Denatured	$\varepsilon_p=4$ [a]	$\varepsilon_p=10$ [a]	$\varepsilon_p=40$ [a]	$\varepsilon_p=40$ [b]
Asp8	4.0	3.1	3.50	3.00	3.05	3.31	3.36
Asp12	4.0	3.5	3.50	2.86	2.91	3.20	3.35
Asp22	4.0	3.3	3.50	2.81	2.88	3.17	3.35
Asp44	4.0	3.6	3.50	3.02	3.08	3.33	3.36
Asp54	4.0	2.2	3.50	2.96	3.03	3.28	3.36
Asp75	4.0	3.1	3.50	2.74	2.79	3.09	3.36
Asp86	4.0	4.2	3.50	2.95	3.02	3.27	3.36
Asp93	4.0	1.5	3.50	2.79	2.84	3.15	3.36
Asp101	4.0	2.0	3.50	2.98	3.03	3.28	3.37
Average				2.90	2.96	3.23	3.36
Glu29	4.4	3.75	3.70	3.40	3.47	3.78	3.87
Glu60	4.4	3.40	3.70	3.24	3.32	3.64	3.87
Glu73	4.4	2.10	3.70	3.23	3.31	3.62	3.87
Average				3.29	3.37	3.68	3.87
Tyr13	9.4			9.02	9.09	9.36	9.30
Tyr17	9.4			8.93	8.99	9.25	9.30
Tyr24	9.4			8.98	9.02	9.28	9.30
Tyr78	9.4			9.00	9.08	9.28	9.30
Tyr90	9.4			9.02	9.06	9.29	9.30
Tyr97	9.4			9.02	9.06	9.29	9.30
Tyr103	9.4			9.00	9.04	9.27	9.30
Average				9.00	9.05	9.29	9.30
His18	6.3	7.61 [c]	6.59 [d]	6.64	6.59	6.42	6.38
His102	6.3			6.59	6.55	6.35	6.38
Average				6.67	6.57	6.39	6.38
Lys19	10.4			11.76	11.69	11.34	11.02
Lys27	10.4			11.61	11.56	11.21	11.02
Lys39	10.4			11.52	11.43	11.10	11.02
Lys49	10.4			11.51	11.43	11.08	11.02
Lys62	10.4			11.63	11.55	11.17	11.02
Lys66	10.4			11.74	11.66	11.25	11.02
Lys98	10.4			11.60	11.52	11.19	11.01
Lys108	10.4			11.65	11.56	11.20	11.02
Average				11.63	11.55	11.19	11.02
Arg59	12.0			13.11	13.07	12.87	12.89
Arg69	12.0			13.33	13.28	13.02	12.89
Arg72	12.0			13.31	13.25	13.02	12.89
Arg83	12.0			13.24	13.20	12.95	12.89
Arg87	12.0			13.28	13.24	13.01	12.89
Average				13.25	13.21	12.97	12.89
C term	3.67	3.3		2.56	2.62	2.83	2.94
N term	8.20			8.87	8.77	8.47	8.37

[a]With distance constraints.
[b]Without distance constraints.
[c]From Ref. [41].
[d]From Ref. [42].

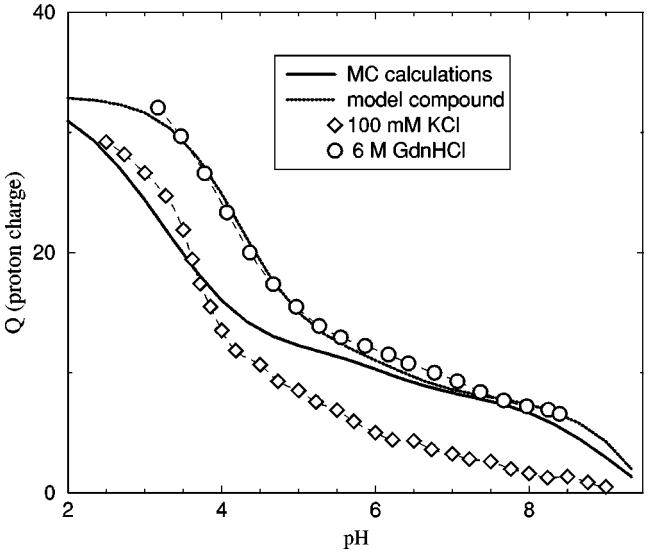

FIG. 3. The net charge of SNase as function of pH. Points denote the experimental data in different solvents taken from Fig. 1 in Ref. [45], while thick lines show results of calculations for different representations of the denatured state. Thin dashed lines are guides for the eye.

we consider two such proteins, barnase from *Bacillus amyloliquefaciens* (the PDB code 1A2P) and staphylococcal nuclease (SNase, the PDB code 1EY0) [39].

A. Barnase

Barnase is a protein with $L_r=110$ residues, which gives a radius of gyration $R_g=24$ Å [Eq. (2)]. This value is between the experimental observations, which vary from 15.9 Å to 34 Å [40]. The protein contains 37 titratable groups, which are distributed over 997 points of the spherical grid with the distance between neighboring points $d_{\min}=2.7$ Å. The pK values for all titratable groups of barnase calculated for three different values of ε_p are presented in Table III. The only amino acid for which the pK value was measured directly is the His18 group [42]. For this residue a good agreement with the experimental result was obtained. The experimental pK values for the acidic groups both for native and for denatured barnase as reported by Oliveberg *et. al* [5] are also given for comparison. Note, that the experimental pK values for the denatured state were averaged values obtained from a theoretical analysis of pH-dependent difference ΔQ in the net charge Q of barnase molecule in the native and denatured states. ΔQ, in turn, was calculated as the derivative of unfolding free energy ΔG obtained from thermal unfolding experiments [4]. Such methodology does not provide Q neither for native nor for the denatured state separately while the MC simulations give Q only for the denatured state. Thus, the only values, which can be compared here, are the pK values.

The pK values calculated for the acidic groups are sufficiently close to, but systematically below the experimental pK values. This discrepancy is most likely due to an overestimation of electrostatic energy, which can have several sources. One of them is the dielectric constant of the protein material that is an adjustable parameter in the present calculations. As seen from Table III, with increasing ε_p the agreement with the experimental data improves and for $\varepsilon_p=40$ the calculated and experimental pK values practically coincide. For His18 residue the agreement between the calculated and measured pK values is good for all ε_p because in this pH region the net charge of the molecule is close to zero, i.e., electrostatic influence is minimal. Considering ε_p as an adjustable parameter, values ~ 40 seem to be most appropriate for the model. Another source for the observed overestimation may be the fact that the radius of the dielectric (protein) sphere is kept constant for the whole pH range considered. Indeed, one expects that electrostatic repulsion between titratable sites may induce a swelling of the protein moiety, especially at extreme pH. Experimental evidences for such an effect have been provided by Stigter *et al.* [43], who have shown that the average coil radius of the myoglobin molecule increases dramatically with reduction of the ionic strength. Neglect of ΔpK_i^{sol} in Eq. (5) may also lead to an overestimation of electrostatic interactions because the desolvation penalty usually shifts pK values towards pK_i^0.

A noteworthy result is that small but detectable variation of the pK values calculated for the individual groups of the same kind is observed. This is not consistent with the concept of the denatured state as a state where the spread of pK's is lost [44]. Consequently, the question arises whether the variations of the pK values is due to insufficient statistics or whether this is a feature of the model proposed. In order to answer this question we performed MC simulations without distance restrictions which is equivalent to the removal of information about protein sequence. All other parameters were kept the same. The calculations for this case clearly showed that there is no dispersion in the pK values (the rightmost column of Table III). This suggests that the observed variation of the pK values is not due to statistical insufficiency, but rather reflects the influence of the protein sequence on the titration properties of individual titratable groups in denatured proteins. More experimental data are needed to ensure that the model proposed describes this influence correctly.

B. Staphylococcal nuclease

Staphylococcal nuclease (SNase) is a highly charged protein containing 31 basic and 28 acidic groups including the amino- and the carboxyl terminal groups ($L_r=149$). The denatured state of SNase was modeled by a flexible chain with 61 segments and a dielectric sphere with $R_g=27.79$ Å. The distance between the neighboring points in this case was $d_{\min}=3.1$ Å. To our knowledge there are no experimental pK values for denatured SNase available in the literature. Therefore, we limited our consideration to comparison of protein titration curves only. In Fig. 3 the titration curve calculated by the model is juxtaposed with the experimental data on H^+ binding obtained by Whitten and Garcia-Moreno [45]. As seen, the null approximation describes SNase well within the wide pH range if the protein is dissolved in 6 M GdnHCl which is very strong denaturate. In this case SNase is chemically denatured within the whole pH interval inves-

tigated in Ref. [45]. It is known that, GdnHCl is highly ionized agent and, therefore, it screens any electrostatic interactions.

However, for SNase dissolved in 100 mM KCl the calculated net charge agrees well with the experimental points for $p\mathrm{H} \lesssim 4$ where the protein is known to be thermally unfolded [45]. The observed, though small, discrepancy between the calculated and experimental net charges for $p\mathrm{H} < 4$ can be due to the factors discussed above, such as the assumption that R_g does not depend on pH. However, the agreement with experimental data in this case is significantly better than that of the null approximation (dotted line in Fig. 3). The net charge predicted by the null approximation is between 5 and 10 charge units larger compared to the experimental data [45]. The seeming agreement of the results obtained by the model proposed and by the null approximations at pH between 6 and 9 is due to the fact that both models are expected to predict all carboxyl, lysines and arginines being charged. The residual difference in the net charge in this pH region arises from the pK shifts of histidines predicted by our model.

V. CONCLUSIONS

The results of calculations presented in the paper suggest an important conclusion, namely, that the influence of electrostatic interactions on titration properties of denatured proteins might be predetermined by the amino-acid sequence of a given protein. The fact that individual titratable groups may have different pK values, i.e., may interact differently with the electrostatic surrounding, is of importance for a better understanding of the factors regulating protein stability.

One of the key assumptions of the model is that the unfolded protein is approximated by a sphere with a radius equal to the radius of gyration of a flexible polypeptide chain. In the present calculations this radius reflects only geometrical features of the polypeptide chain. We suppose that this is the main source of the observed tendency of overestimation of electrostatic energy. It is known, for instance, that at different conditions (such as pH and temperature) proteins may adopt different denatured states characterized by different compactness and by different residual secondary structures. These factors are not reflected in Eq. (2). Nevertheless, the model gives a more realistic physical description of thermally unfolded proteins and predicts electrostatic properties essentially better than the commonly used null approximation. An extension of Eq. (2) by including empirical terms determining pH- and temperature dependencies of R_g is in progress.

ACKNOWLEDGMENTS

This work was financially supported by Grant No. BIO4CT970129 from the IV Biotechnology Program of the European Communities and by Grant No. A1-5/2286 from the Swedish Council for Planning and Coordination of Research. A. Karshikoff thanks Professor R. Ladenstein (Dept. Biosciences at Novum, Karolinska Institutet) for useful discussions.

APPENDIX

Electrostatic energy of a polypeptide chain in a microscopic state **X** (see Sec. III A) assuming that all charges are point charges, can be written in general form as

$$\mathcal{E}(\mathbf{X}) = e\sum_i (\mathcal{P}_i - x_i)\left\{\frac{1}{2}\sum_j \varphi_i(q_j) + \varphi_i(q_{per}) + \Delta\varphi_i(q_i)\right\} + \sum_i \mu_i x_i \,, \tag{A1}$$

where e is the elementary charge. The potential $\varphi_i(q_j)$ is created by a charge of jth group ($j \neq i$), $\varphi_i(q_{per})$ is a potential created by all permanent (nontitratable) charges of a protein molecule and $\Delta\varphi_i(q_i)$ is a ''potential'' arising from difference between the self-energies of ith group in water and in protein (the Born energy or desolvation penalty).

The chemical potential μ_i in Eq. (A1) is related to the concentration of the protons $[\mathrm{H}^+]$ in surrounding solvent [46]

$$\mu_i = \mu_i^0 - kT\ln(\gamma_{\mathrm{H}^+}[\mathrm{H}^+]), \tag{A2}$$

where γ_{H^+} is the activity coefficient of the protons, T is the temperature, and k is the Boltzman constant. μ_i^0 is a residue-specific constant that might be viewed as the change of the self-energy of the ith titratable group (when alone in the solvent) during proton binding

$$\mu_i^0 = kT\ln\mathcal{K}_i \tag{A3}$$

with $\mathcal{K}_i$ being an equilibrium constant for the chemical reaction of the proton binding/unbinding to the ith group. Usually, relations (A2) and (A3) are expressed in terms of

$$p\mathrm{H} = -\log_{10}(\gamma_{\mathrm{H}^+}[\mathrm{H}^+]) \tag{A4}$$

and

$$p\mathrm{K}_i^{\mathrm{mod}} = -\log_{10}\mathcal{K}_i \,. \tag{A5}$$

Then, introducing notations

$$\Delta p\mathrm{K}_i^{\mathrm{per}} = e\varphi_i(q_{\mathrm{per}})/kT\ln(10) \tag{A6}$$

and

$$\Delta p\mathrm{K}_i^{\mathrm{sol}} = e\Delta\varphi_i(q_i)/kT\ln(10), \tag{A7}$$

Eq. (A1) might be rewritten in more customary, frequently used in the literature, form (see, e.g., Refs. [47,48])

$$\mathcal{E}(\mathbf{X}) = \mathcal{E}_0 + \frac{1}{2}\sum_{i,j} W_{ij}(\mathcal{P}_i - x_i)(\mathcal{P}_j - x_j) - \ln(10)kT\sum_i (p\mathrm{H} - p\mathrm{K}_i^{\mathrm{int}})x_i \,, \tag{A8}$$

where W_{ij} is a pairwise interaction potential,

$$\mathcal{E}_0 = kT \ln(10) \sum_i (\Delta pK_i^{per} + \Delta pK_i^{sol}) \mathcal{P}_i , \qquad \text{(A9)}$$

is contributions of the permanent charges and of the desolvation penalties in a fully protonated state of a polypeptide (all $x_i = 0$) and pK_i^{int} described by the Eq. (5).

Note that Eq. (A8) very often in the literature is written without the term $\mathcal{E}_0$. This does not lead to any mistakes when one works within only one conformation of a polypeptide since $\mathcal{E}_0$ is uniquely determined by structures (primary, secondary, tetriary, etc.) of a molecule. However, the term $\mathcal{E}_0$ become important when either different conformations of the same molecule or the same (or similar) conformations of different molecules (e.g., a wild-type protein and its mutants) are compared.

[1] A. Karshikoff and R. Ladenstein, Trends Biochem. Sci. **26**, 550 (2001).
[2] G.M. Ullmann and E.W. Knapp, Eur. Biophys. J. **28**, 533 (1999).
[3] C.N. Pace, R.W. Alston, and K.L. Shaw, Protein Sci. **9**, 1395 (2000).
[4] M. Oliveberg, S. Vuilleumier, and A.R. Fersht, Biochemistry **33**, 8826 (1994).
[5] M. Oliveberg, V.L. Arcus, and A.R. Fersht, Biochemistry **34**, 9424 (1995).
[6] Y.-J. Tan, M. Olivegerg, B. Davis, and A.R. Fersht, J. Mol. Biol. **254**, 980 (1995).
[7] M. Schaefer, M. Sommer, and M. Karplus, J. Chem. Phys. **101**, 1663 (1997).
[8] J. Warwicker, Protein Sci. **8**, 418 (1999).
[9] J. Warwicker (private communication).
[10] A.-S. Yang and B. Honig, J. Mol. Biol. **237**, 602 (1994).
[11] A.H. Elcock, J. Mol. Biol. **294**, 1051 (1999).
[12] A. Warshel and S.T. Russell, Q. Rev. Biophys. **17**, 283 (1984).
[13] A. Warshel, S.T. Russell, and A.K. Churg, Proc. Natl. Acad. Sci. U.S.A. **81**, 4785 (1984).
[14] J. Warwicker and N.C. Watson, J. Mol. Biol. **175**, 527 (1984).
[15] I. Klapper *et al.*, Proteins: Struct. Funct., Genet. **1**, 47 (1986).
[16] A. Nicholls and B. Honig, J. Comput. Chem. **12**, 435 (1991).
[17] B. Honig, K. Sharp, and A.-S. Yang, J. Phys. Chem. **97**, 1101 (1993).
[18] C. Tanford, *Physical Chemistry of Macromolecules* (Wiley, New York, 1961).
[19] C. Branden and J. Tooze, *Introduction to Proteine Structure* (Garland Publishing, New York, 1998), Chap. 1, p. 8.
[20] T.E. Creighton, *Proteins: Structure and Molecular Properties*, 2nd ed. (Freeman, New York, 1993).
[21] S.L. Kazmirski and V. Daggett, J. Mol. Biol. **277**, 487 (1998).
[22] C.K. Smith *et al.*, Protein Sci. **5**, 2009 (1996).
[23] Y.O. Kamatari, T. Konno, M. Kataoka, and K. Akasaka, J. Mol. Biol. **259**, 512 (1996).
[24] M. Kataoka, K. Kuwajima, F. Tokunaga, and Y. Goto, Protein Sci. **6**, 422 (1997).
[25] D.J. Segel *et al.*, J. Mol. Biol. **288**, 489 (1999).
[26] L. Chen *et al.*, J. Mol. Biol. **225**, 225 (1998).
[27] Y.O. Kamatari, T. Konno, M. Kataoka, and K. Akasaka, Protein Sci. **7**, 681 (1998).
[28] B. Ibarra-Molero and J.M. Sanchez-Ruiz, Biochemistry **36**, 9616 (1997).
[29] G.V. Semisotnov *et al.*, J. Mol. Biol. **262**, 559 (1996).
[30] T. Konno, Y. Kamatari, M. Kataoka, and K. Akasaka, Protein Sci. **6**, 2242 (1997).
[31] L. Shi, M. Kataoka, and A.L. Fink, Biochemistry **35**, 3297 (1996).
[32] V. Receveur *et al.*, FEBS Lett. **426**, 57 (1998).
[33] H.H. Gan, A. Tropsha, and T. Schlick, J. Chem. Phys. **113**, 5511 (2000).
[34] C. Tanford and J.G. Kirkwood, J. Am. Chem. Soc. **79**, 5333 (1957).
[35] J.G. Kirkwood, J. Chem. Phys. **2**, 351 (1934).
[36] K. Binder, in *Monte Carlo Methods in Statistical Physics*, edited by K. Binder (Springer-Verlag, Berlin, 1979).
[37] J.B. Matthew, Annu. Rev. Biophys. Biophys. Chem. **14**, 387 (1985), review.
[38] D. Bashford and M. Karplus, Biochemistry **29**, 10 219 (1990).
[39] The PDB files for a particular protein using the PDB code can be obtained via Internet on, e.g., http://www.rcsb.org/pdb/
[40] K.-B. Wong *et al.*, J. Mol. Biol. **296**, 1257 (2000).
[41] J.L. Neira and A.R. Fersht, J. Mol. Biol. **285**, 1309 (1999).
[42] D. Sali, M. Bycroft, and A.R. Fersht, Nature (London) **335**, 740 (1988).
[43] D. Stigter, D.O.V. Alonso, and K.A. Dill, Proc. Natl. Acad. Sci. U.S.A. **88**, 4176 (1991).
[44] R. Roxby and C. Tanford, Biochemistry **10**, 3348 (1971).
[45] S.T. Whitten and B. Garcia-Moreno, Biochemistry **39**, 14 292 (2000).
[46] P.H. McPherson, M.Y. Okamura, and G. Feher, Biochim. Biophys. Acta **934**, 348 (1988).
[47] D. Bashford and M. Karplus, J. Phys. Chem. **95**, 9556 (1991).
[48] P. Beroza, D.R. Fredkin, M.Y. Okamura, and G. Feher, Biophys. J. **68**, 2233 (1995).