Protein Science (1999), *8*:418–425. Cambridge University Press. Printed in the USA.

Simplified methods for pK_a and acid pH-dependent stability estimation in proteins: Removing dielectric and counterion boundaries

JAMES WARWICKER

Institute of Food Research, Reading Laboratory, Earley Gate, Whiteknights Road, Reading RG6 6BZ, United Kingdom

(Received July 20, 1998; Accepted November 3, 1998)

Abstract

Much computational research aimed at understanding ionizable group interactions in proteins has focused on numerical solutions of the Poisson–Boltzmann (PB) equation, incorporating protein exclusion zones for solvent and counterions in a continuum model. Poor agreement with measured pK_as and pH-dependent stabilities for a (protein, solvent) relative dielectric boundary of (4,80) has lead to the adoption of an intermediate (20,80) boundary. It is now shown that a simple Debye–Hückel (DH) calculation, removing both the low dielectric and counterion exclusion regions associated with protein, is equally effective in general pK_a calculations. However, a broad-based discrepancy to measured pH-dependent stabilities is maintained in the absence of ionizable group interactions in the unfolded state. A simple model is introduced for these interactions, with a significantly improved match to experiment that suggests a potential utility in predicting and analyzing the acid pH-dependence of protein stability. The methods are applied to the relative pH-dependent stabilities of the pore-forming domains of colicins A and N. The results relate generally to the well-known preponderance of surface ionizable groups with solvent-mediated interactions. Although numerical PB solutions do not currently have a significant advantage for overall pK_a estimations, development based on consideration of microscopic solvation energetics in tandem with the continuum model could combine the large ΔpK_as of a subset of ionizable groups with the overall robustness of the DH model.

Keywords: colicins; pH-dependence; pK_a; protein dielectric; protein electrostatics; protein stability

Calculated macromolecular electrostatic potential surfaces, particularly with the finite difference Poisson-Boltzmann (FDPB) continuum method (Warwicker & Watson, 1982; Klapper et al., 1986; Warwicker, 1986), are commonly accepted aids to graphical interpretations of structure. In contrast, calculations of pK_as and pH-dependent stabilities in the continuum approximation are more problematic. Assignment of low protein (Gilson & Honig, 1986) and high solvent relative dielectric values ($\epsilon_p \sim 4$, $\epsilon_s \sim 80$) does not give good agreement with measured pK_as (Bashford & Karplus, 1990; Antosiewicz et al., 1994). Factors that may contribute to this discrepancy include the use of a restricted conformational set or a single conformer (You & Bashford, 1995; Zhou & Vijayakumar, 1997), particularly in regard to relaxation in response to pH titration, lack of consideration of ionizable group tautomers (Antosiewicz et al., 1996), and inadequate modeling of charge burial and the Born (self) energy (Born, 1920; Warwicker, 1997).

Within the continuum framework, $\epsilon_p \sim 20$ is an effective method for improving agreement across a range of measured pK_as (Antosiewicz et al., 1994) and decreasing charge burial penalties. A study with more extensive variation of ϵ_p (Demchuk & Wade, 1996) differentiates groupings of ionizable residues that give better pK_a agreement with $\epsilon_p \sim 10$–20 (mostly buried) or with $\epsilon_p \sim 80$ (mostly solvent exposed). Reasonable agreement to measured lysine pK_as in calbindin has been found with $\epsilon_p = \epsilon_s$ (Kesvatera et al., 1996). On the other hand, interactions in an $\epsilon_p = 20$ or $\epsilon_p = \epsilon_s = 80$ model are too weak to give the notably low and functionally significant cysteine pK_as of DsbA and thioredoxin (Warwicker, 1998). These cysteine pK_as are matched by $\epsilon_p = 4$ calculations that allow for the hydration entropy associated with solvent ordering around an ionized group, and its variation with extent of charge burial, thereby counteracting the unfavorable Born energy (Warwicker, 1997). Substantially shifted pK_as will be important in uncovering continuum model inadequacies and directing potential improvements. Although a charge burial-linked hydration entropy term is simple to implement, further progress requires consideration of ionizable side-chain conformation and salt-bridging

Reprint requests to: James Warwicker, Institute of Food Research, Reading Laboratory, Earley Gate, Whiteknights Road, Reading RG6 6BZ, United Kingdom; e-mail: james.warwicker@bbsrc.ac.uk.

Abbreviations: DH, Debye-Hückel; FDPB, finite difference Poisson-Boltzmann; HEWL, hen egg white lysozyme; MC, Monte Carlo; RMSD, root-mean-square deviation.

in multiple, pH-dependent conformers. Such factors are related to the strong interactions that can arise from protein dipolar interactions, and their relaxation with changes in ionization (Sham et al., 1998). While continuum calculations need development to handle buried and surface groups equally well in a single model that covers the entire range of observed ΔpK_as, it is of interest to examine a range of higher ϵ_p-based continuum models, in terms of the overall predictive power that derives from the preponderance of ionizable groups with relatively low ΔpK_as.

The shapes of measured pH-dependent stability curves are matched by the $\epsilon_p = 20$ model, but magnitudes tend to be overestimated (Antosiewicz et al., 1994). Such discrepancy can be reduced when extended polypeptide models are used to estimate ionizable group interactions in the denatured state (Schaefer et al., 1997), consistent with measurements of significant interactions in the denatured state of barnase (Oliveberg et al., 1995). A simple sequence-based model is now introduced for these interactions. With regard to ΔpK_as, it is found that a Debye–Hückel (DH) description, without dielectric or counterion boundaries ($\epsilon_p = \epsilon_s = 80$), gives results at least as good as those from more complex FDPB models. In combination with the unfolded state model, DH modeling gives much improved agreement with measured acid-dependence of stability (ΔG_{NU}). Although the topic of a high effective dielectric or its ability to cover for model inadequacies is not new (Karshikov et al., 1991; Sham et al., 1997), this paper gives a clear picture of the utility of a DH model alongside FDPB calculations that incorporate dielectric and counterion boundaries, as well as showing that large discrepancies in the calculated pH-dependence of ΔG_{NU} can be significantly reduced with a simple unfolded state model. These results are important in the context of a rapidly-expanding structural database of potential applications, demonstrated with application to the biologically-relevant acid stability variation of the pore-forming domains of colicins A and N, firstly in qualitative comparison to melting temperature measurements, and secondly with prediction of the molecular origin of this difference.

Results and discussion

Model comparison to measured pK_as

Results of FDPB and DH calculations are compared for four proteins (Table 1), extending work that established the $\epsilon_p \sim 20$ model (Antosiewicz et al., 1994, 1996). Although current and reported $\epsilon_p = 20$ calculations (fd20 and fd20a) show some difference, presumably arising from differences in charge assignments and programming of the FDPB method, they give qualitatively similar results with moderately better RMSDs than the null model. Removal of the dielectric boundary, maintaining a counterion inaccessible envelope (fd80, $\epsilon_p = \epsilon_s = 80$), gives marginally lower RMSDs overall than fd20, so that $\epsilon_p = 80$ is as effective as $\epsilon_p = 20$ in modeling these ionizable group pK_as. The fd80 model approximates a system with uniform dielectric (ϵ_s) and Monte Carlo (MC) simulation of counterions around the protein that has been effective in studies of calcium and hydrogen ion binding to proteins (Kesvatera et al., 1994, 1996). Counterion boundary removal with DH calculations (dh80, $\epsilon_{DH} = 80$) shows relatively little difference to fd80, with this simplified model performing at least as well as those of more complexity that retain counterion (fd80) and dielectric (fd20) boundaries, and giving significant predictive power relative to the null model.

Comparison of dh80 and dh50 ($\epsilon_{DH} = 50$) (Warwicker et al., 1987) shows only slight variation in match to experiment, emphasizing that success lies in the preponderance of ionizable groups that are mostly solvent exposed and for which higher dielectric dominates interactions. It is also apparent that dh80 ΔpK_as are driven by interactions between ionizable groups, rather than between ionized and dipolar groups. For example, the dh80 RMSD of 0.48 for RNase A moves only to 0.53 upon neglect of the ionizable group-dipolar interactions. In relation to the view that, in the absence of microscopic consideration of protein relaxation, larger ϵ_p is suitable for charge-charge and smaller ϵ_p more appropriate for charge-dipole interactions (Sham et al., 1998), the dh80 model may fail to account adequately for dipolar solvation in the subset of significantly buried ionizable groups.

An example in this subset is E35 of HEWL, for which fd20, fd80, and dh80 calculations on the tricilinic form give pK_as of 4.7, 3.9, and 4.2, respectively, compared to 6.1 by experiment (Kuramitsu & Hamaguchi, 1980). In contrast, finite difference calculations with $\epsilon_p = 4$ (Bashford & Karplus, 1990; Yang & Honig, 1993; Antosiewicz et al., 1994) give an E35 pK_a substantially elevated from that of the model compound, although RMSDs across the set of HEWL pK_as are lower with $\epsilon_p > 4$ (Antosiewicz et al., 1994). We favor the development of $\epsilon_p \sim 4$ based (FDPB) continuum models with consideration of factors such as hydration entropy (Warwicker, 1997) and pH-dependent conformational sampling,

Table 1. *Comparison of RMSDs to experiment for pK_a calculations*

Protein	n_gps[a]	μ (M)[b]	Null[c]	fd20a[d]	fd20	fd80	dh50	dh80
RNaseA	16	0.20	0.86	0.75	0.66	0.51	0.47	0.48
RNaseT	4	0.15	1.21	1.14	0.94	0.75	0.75	0.78
Barnase	13	0.05	1.28	—	1.06	0.77	0.88	0.68
tri_Lys	10	0.15	1.14	0.78	0.92	0.98	0.94	0.90
tet_Lys	10	0.15	1.14	0.99	0.73	0.90	0.80	0.84

[a] Number of ionizable groups.
[b] Ionic strength.
[c] RMSDs for model compound pK_as.
[d] Reported $\epsilon_p = 20$ calculations of Antosiewicz et al. (1994), with nd1 proton in the neutral histidine form for $RNaseT_1$, barnase, and HEWL (tri_Lys and tet_Lys).

alongside applications of the dh80 model that are aimed at predicting overall pK_a effects, but with potential underestimation of a small subset of ionizable group effects. The view that lower ϵ_p-based FDPB models can be improved with detailed analysis of microscopic solvation energetics is consistent with analogous progress reported for the protein dipoles Langevin dipoles method (Sham et al., 1997).

Model comparison to measured pH-dependent stabilities

Modeling of unfolded state charge interactions (Fig. 1) is tested in Figures 2 and 3. For RNase A, Figure 2A shows that $\epsilon_p = 20$ (fd20a, fd20) significantly overestimates neutral and acid pH-dependent changes in the unfolding energy ΔG_{UN} (Antosiewicz et al., 1994), which is also the case for $\epsilon_p = 80$ (fd80). Calculations with DH modeling without unfolded state interactions (dh50, dh80), and with the u1 model (dh50u1, dh80u1), also give substantial overestimation of measured changes, while incorporation of the u2 model (dh50u2, dh80u2) gives significantly improved agreement (Fig. 2B). With relatively small differences between equivalent models that use $\epsilon_{DH} = 50$ and $\epsilon_{DH} = 80$, $\epsilon_{DH} = 80$ modeling (effectively a water solvent continuum) is used in further comparisons.

General agreement between fd20 and fd20a through the left panels of Figure 3 shows that the two finite difference implementations, with separate charge sets, give similar results. The $\epsilon_p = 20$ modeling again overestimates the pH-dependence of ΔG_{UN}, but to varying degrees (Antosiewicz et al., 1994). Replacement of fd20 with fd80 gives little difference for barnase and triclinic HEWL, and a moderate increase in pH-dependence for tetragonal HEWL.

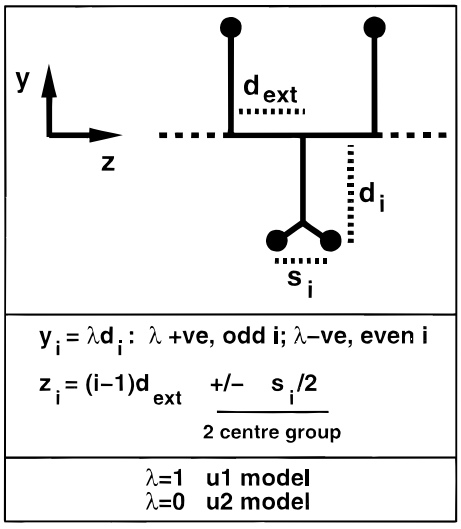

Fig. 1. A simple model for ionizable group interactions in the unfolded state. Ionizable groups are anchored along the *z*-axis according to residue number and C_α spacing in an extended polypeptide chain ($d_{ext} = 3.8$ Å). The ionizable charge of residue *i* is located at a distance of λd_i (on *y*) from the backbone, with λ scaling from 0 to 1 and positive for odd *i*/negative for even *i* (so that nearest neighbor side chains alternate), and d_i assigned from extended side-chain geometries: D, 2.8 Å; E, 4.2 Å; H, 3.9 Å; K, 6.3 Å; R, 6.8 Å. A distance s_i separates charges for two-site ionizable groups: D and E, 2.16 Å; H, 2.43 Å; R, 2.30 Å. Amino and carboxyl terminal ionizable groups are treated in analogous fashion, with $d_i = 0$.

For RNase T_1, fd80 gives better agreement with experiment than fd20. In this case the increased stability at mild acidic pH arises from protonation of D/E residues with elevated pK_as, and the associated unfavorable interactions between acidic groups are reduced by $\epsilon_p = 80$ relative to $\epsilon_p = 20$.

Results with the dh80 model are similar to those with fd80, leaving a large discrepancy to experiment for barnase in particular (right panels of Fig. 3). In contrast, dh80u2 gives a reasonable match to experiment for all proteins shown in Figure 3. The u2 model shifts ΔG_{UN} pH-dependence to a varying degree for each protein, but sufficiently to provide a reasonable estimate of measured values in each case (compare the scale of u2 modifications for RNase A and barnase to those for the other proteins). It is noted again that this solvent-based approach can be wrong in detail. The shape of the measured curve for HEWL is not precisely matched (recalling also that dh80 does not give the elevated E35 pK_a); dh80 or fd80 give a moderately better match to experiment than does dh80u2 for RNase T_1; and FD/DH model charge sets do not account for histidine-aromatic interactions in barnase (Loewenthal et al., 1992). However, the current work indicates that alongside general pK_a predictions with dh80 that are as effective as current FDPB methods, dh80u2 gives significant improvement in predictions of folding energy pH-dependence.

The u2 model is investigated in Figure 4 for those groups contributing to neutral-acid pH-dependence. The overall stabilizing effect of ionizable group interactions is clear for ΔpK_as calculated with (Δp$K_{a,NU}$) and without (Δp$K_{a,N}$) unfolded model subtraction. A strong Δp$K_{a,N} \sim \Delta$p$K_{a,NU}$ diagonal indicates relatively small Δp$K_{a,U}$ modification for most groups. All points substantially below this diagonal represent histidine or N-t groups with a neighboring basic residue in the protein sequence, giving unfavorable interactions in the unfolded state model. Groups substantially above the main diagonal tend to lie close to a further diagonal arising from favorable nearest neighbor interactions in the unfolded state model. RNase A and barnase each contribute five such groups, RNase T_1 three, and HEWL none, so that the variable scale of the unfolded model term, that is an important determinant of agreement with experiment in Figures 2 and 3, relates directly to sequential nearest neighbor ionizable group pairs. Residue D38 of RNase A has two nearest neighbor bases, making a proportionate contribution to the dh80/dh80u2 difference in Figure 2. The nearest neighbor interaction is 0.8–0.9 pK_a units for the u2 model with $\epsilon_{DH} = 80$ at 0.03 M ionic strength.

A rapid decline of interaction strength with sequence separation in the unfolded state model gives the clear prediction that nearest neighbor pairs will dominate, with an uneven distribution of Δp$K_{a,U}$s. Calculated (dh80u2) Δp$K_{a,U}$s for the 12 D/E residues of barnase range from -1.2 to $+0.1$, with an average of -0.4 that agrees with the experimentally determined mean value (Oliveberg et al., 1995). The calculated p$K_{a,U} = 3.9$ for D93 of barnase at 0.03 Mμ is consistent with reported protonation at around pH 4 and 0.05 Mμ (Oliveberg et al., 1995). Model development will probably include conformational sampling (Schaefer et al., 1997). For example, it is likely that charge interactions will contribute to conformational preferences, so that the magnitude of an unfavorable nearest neighbor interaction will be less than that of the corresponding favorable pair in the unfolded state.

The calculated distribution of Δp$K_{a,NU}$s (Fig. 4) indicates the extent to which ionizable group contributions to stabilization and

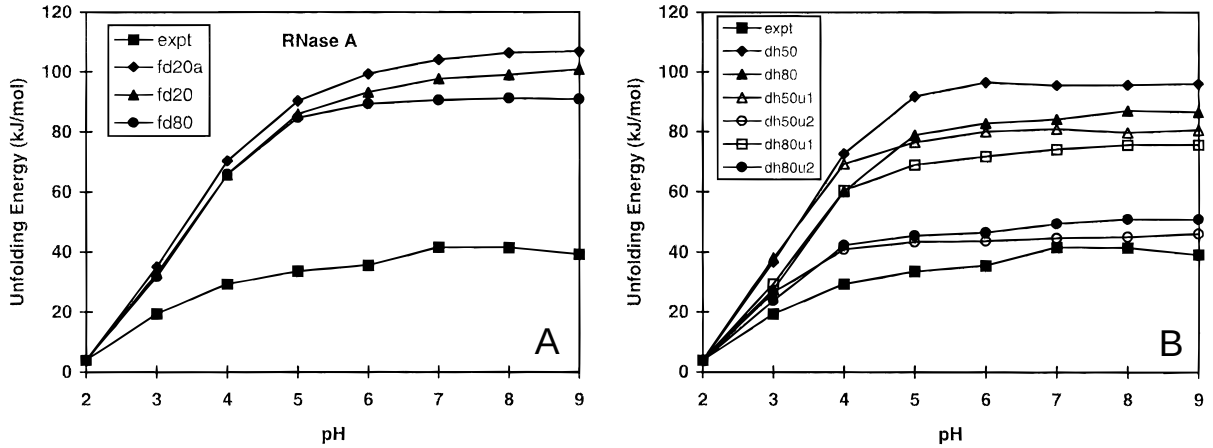

Fig. 2. Calculated and measured pH-dependence of unfolding energy for RNase A. **A:** FD calculations compared to experiment for $\epsilon_p = 20$ (fd20a, Antosiewicz et al., 1994; fd20, current), and $\epsilon_p = \epsilon_s = 80$ (fd80). **B:** DH calculations, with and without ionizable group interactions in the unfolded state (u1 and u2 models), compared with experiment for $\epsilon_{DH} = 50$ (dh50, dh50u1, dh50u2), and for $\epsilon_{DH} = 80$ (dh80, dh80u1, dh80u2).

acid pH-dependence vary in the dh80u2 model. While a relatively small set of residues have stabilizing $\Delta pK_{a,NU} > 1.5$, a larger number maintain significant contributions to ΔG_{NU}, predicting a generally broad structural base to acid pH-dependence. Agreement with experiment (Figs. 2, 3) suggests that although a small subset of more extreme pK_a shifts may be missing in the dh80u2 model, the values in Figure 4 provide a reasonable overall estimate of acid pH-dependent stability in these proteins. It will be important to develop methods that combine the detail required for the subset of ionizable groups that interact strongly within the native protein with the robustness of the more simple algorithms. In addition, studies of relative energetics for macromolecule conformers and ligand binding will require complete descriptions of polar and nonpolar interactions (Schaefer et al., 1997). However, the pH-dependent component is of considerable interest, for example with regard to acid denaturation or molten globule formation (Fink et al., 1994), and the next section gives dh80u2 model-based predictions for a system in which acid stability is physiologically relevant.

Colicin pore-forming domains

The pore-forming domains of colicins A and N (ThColA and ThColN) exhibit different acid pH sensitivities, as measured by melting temperature (T_m), that may be related to membrane insertion mediated by a molten globule intermediate in the lower effective pH close to a negatively-charged membrane (Lakey et al., 1994; Evans et al., 1996). With a PDB coordinate set available for ThColA but not ThColN, and with 58% identity between the two molecules, comparative modeling is used together with DH-based calculations. Although direct comparison between T_m and unfolding energy is not possible, modeling without ionizable group interactions in the unfolded state (dh80, Fig. 5B) gives substantial pH-dependence that is similar for ThColA and ThColN, and not in accord with differences in the acid pH-dependence of T_m (Evans et al., 1996) (Fig. 5A). In contrast, the dh80u2 model (Fig. 5C) gives a smaller overall scale for acid pH-dependence within which the ThColA/ThColN difference is more significant and in line with experiment.

Figure 6 shows contributions to acid pH-dependence with $\Delta pK_{a,NU}$ values for D and E residues of ThColA and ThColN. The histidines (H29 and H88 of ThColA and H49 of ThColN) are not predicted to contribute significantly, noting also that chemical modification of histidines does not alter the pH-dependence of membrane insertion for ThColA (Evans et al., 1996). Over the three panels of Figure 6, the difference between $\Delta pK_{a,NU}$ sums for ThColA and ThColN is -1.0, corresponding approximately to the difference in relative unfolding energy at pH 2 (Fig. 5C). The dh80u2 model predicts a ThColA/ThColN differential spread over a number of acidic groups, with an emphasis on those unique to ThColA (Fig. 6B). It has been noted that the ThColA/ThColN acid pH-dependence difference cannot be a simple net charge effect since ThColN is more positively-charged than ThColA, and further suggested that a subset of acidic residues could mediate the difference (Evans et al., 1996). The current work predicts individual group contributions and indicates that the overall ThColA/ThColN difference is a cumulative effect from summation of relatively small differences. A ThColA molecule with the D78N mutation binds faster to negatively-charged vesicles than wild-type, but inserts more slowly into the membrane (Lakey et al., 1994). Binding has been related to ThColA/membrane orientation prior to insertion, whilst slower insertion may signal increased native state stability (Lakey et al., 1994). The dh80u2 model predicts a destabilizing effect for the negative charge of D78 in ThColA (positive $\Delta pK_{a,NU}$ in Fig. 6A), but a stabilizing influence ($\Delta pK_{a,N} = -0.2$ in ThColA) is predicted for the dh80 model, supporting incorporation of the unfolded state model in calculations of unfolding energy.

Calculations with colicin pore-forming domains show the predictive utility of DH-based modeling for ionizable group interactions in native and unfolded state proteins. This modeling extends reports of the effectiveness of incorporating higher ϵ_p (Antosiewicz et al., 1994) and extended chain estimates of the unfolded state (Schaefer et al., 1997). The robustness and simplicity of the current model make it suitable for overall analyses of pH-dependence in proteins, while improved continuum methods based on lower ϵ_p (e.g., Warwicker, 1997), will be valuable in studies that include the

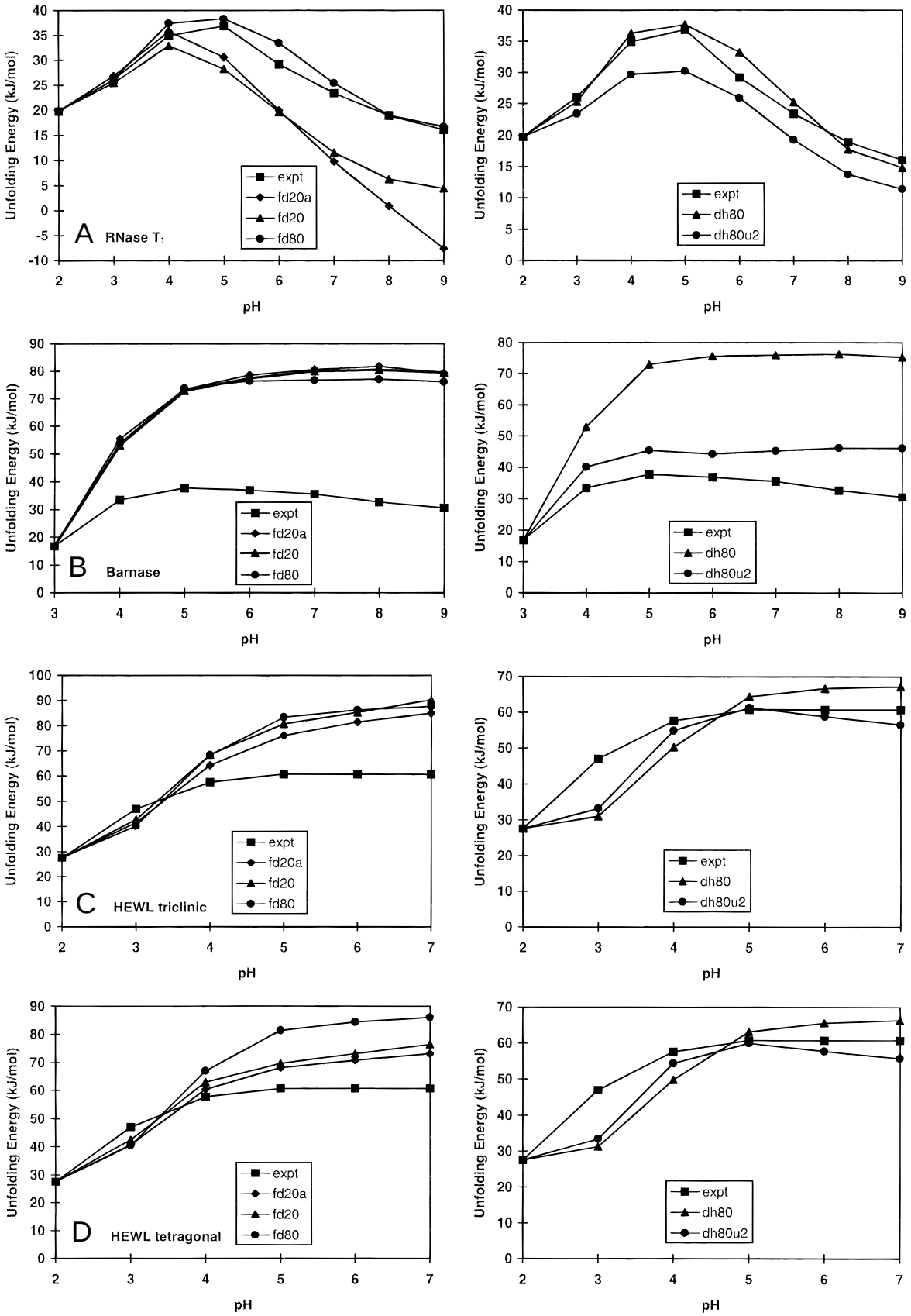

Fig. 3. Calculated and measured pH-dependence of unfolding energy for RNaseT_1, barnase, and HEWL. The left panels give the same calculations as Figure 2A, and the right panels show a subset of the calculations in Figure 2B, based on $\epsilon_{DH} = 80$ with (dh80u2) and without (dh80) unfolded state ionizable group interactions. **A:** RNase T_1. **B:** Barnase. **C:** Triclinic HEWL. **D:** Tetragonal HEWL.

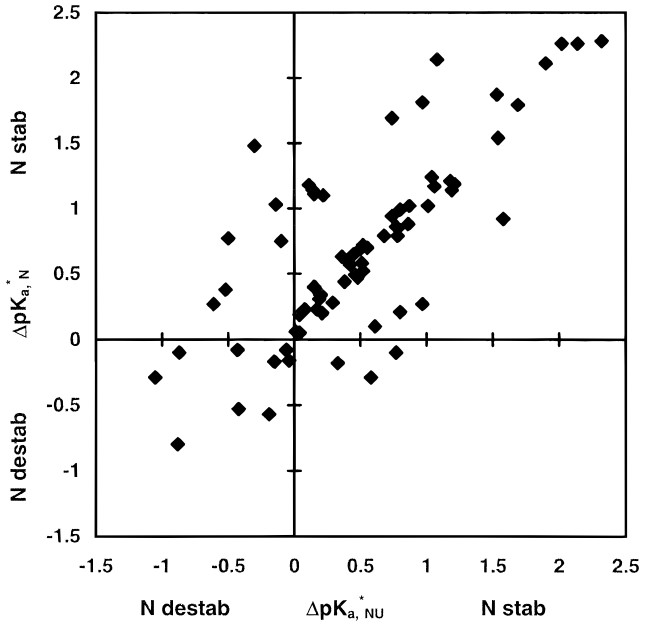

Fig. 4. A scatter plot of calculated $\Delta pK^*_{a,N}$ against calculated $\Delta pK^*_{a,NU}(\Delta pK^*_{a,N} - \Delta pK^*_{a,U})$ for D, E, H, N-t, and C-t residues in RNase A, RNase T_1, barnase, and HEWL for the u2 unfolded state model. ΔpK^*_a indicates that ΔpK_a signs for D, E, and C-t residues have been reversed so that positive ΔpK_a relates uniformly to stabilization (relative to model compound values), and negative ΔpK_a to destabilization.

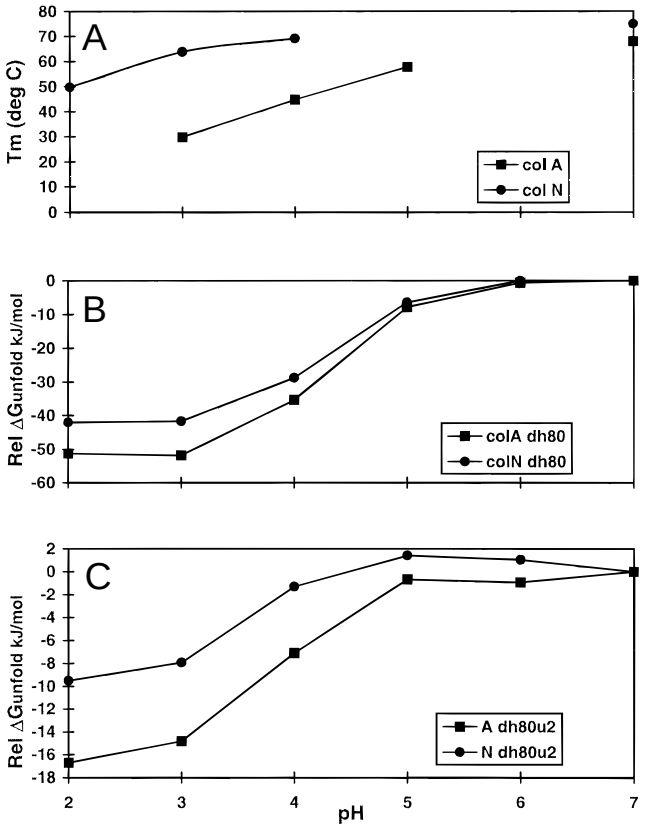

Fig. 5. Comparison of the measured pH-dependence of T_m (Evans et al., 1996) and the calculated pH-dependence of the unfolding energy for the pore-forming domain of colicin A relative to that of colicin N. All calculated curves are set to zero at pH 7. **A:** Measured T_m. **B:** Calculations (dh80) without unfolded state interactions. **C:** Calculations (dh80u2) with unfolded state interactions.

small subset of ionizable groups having relatively strong interactions within a protein.

Materials and methods

Calculations

Charge interactions were calculated either with finite difference solutions to the Poisson–Boltzmann equation (Warwicker & Watson, 1982; Warwicker, 1986), in the program FDCALC, or with a DH screened potential for charge q at distance r in a uniform relative dielectric $\epsilon_{\rm DH}$ $[q/4\pi\epsilon_0\epsilon_{\rm DH}r]e^{-\kappa r}$, with ϵ_0 the zero permittivity and $\kappa^2 = 2{,}000e^2N_A\mu/\epsilon_0\epsilon_{\rm DH}kT$, with μ the molarity of univalent counterions, k is Boltzmann's constant, N_A is Avagadro's number, e is the unit charge, and $T = 300$ K. Ionizable groups included were: D, E, H, K, R, and N-t, C-t as appropriate; with two charge centers for D, E, H, R, and C-t residues. Tyrosine (model compound pK_a of 9.6) ionization may be a factor in protein stability at alkaline pH. The current work excludes tyrosines from pK_a comparisons and studies pH-dependence at pH < 9, concentrating on changes from neutral to acidic pH. Partial (background) charges were allocated from the GROMOS library (van Gunsteren & Berendsen, 1987). Ionization has been modeled as the transition from a neutral form with zero partial charge. Analysis of partial charge distribution in ionizable group neutral forms is of potential importance for a subset of residues, but is not the target of this study.

A statistical treatment of interacting ionizable groups (Bashford & Karplus, 1990), extended with an MC method for computations with large numbers of such groups (Beroza et al., 1991), and coded in the program PKCALC, was used to derive pK_as. The MC method used 10,000 steps and a modification that allows for multiple site transitions in pairs that are coupled by an interaction equivalent to more than two pK_a units (Beroza et al., 1991). Finite difference calculations were made with $\epsilon_p = 20$ and $\epsilon_p = \epsilon_s = 80$, and DH calculations with $\epsilon_{\rm DH} = 50$ and $\epsilon_{\rm DH} = 80$. The molecular surface drawn by a 1.4 Å radius solvent probe formed the FDPB dielectric boundary, noting that a smaller probe radius (Demchuk & Wade, 1996) or use of the van der Waals surface (Antosiewicz et al., 1994) would tend toward smaller pK_a shifts. The pK_a calculations include Born/self-energy and background charge terms, differenced between protein and isolated amino acid, as well interactions between ionizable groups (Bashford & Karplus, 1990). Self-energy differences are zero for DH calculations. Model compound pK_as, relating to the isolated amino acids, are: D, 4.0; E, 4.4; H, 6.3; K, 10.4; R, 12.0; N-t, 7.5; C-t, 3.8. Unfolded state pK_as either follow model compound values or derive from a simple model for ionizable group interactions. The pH-dependence of the free energy difference between the native (N) or unfolded (U) state and the appropriate set of model compounds (M) is calculated either directly from the partition function (for smaller numbers of groups) or, for larger numbers of groups, through pH summation of the difference in average net charge (Yang & Honig, 1993; Antosiewicz et al., 1994), for example following $\partial(\Delta G_{NM})/\partial{\rm pH} = 2.303RT(\langle q_M\rangle - \langle q_N\rangle)$ for the N/M difference. Then $\Delta G_{NU} = \Delta G_{NM} - \Delta G_{UM}$, so that $\Delta G_{NU} = \Delta G_{NM}$ when U is modeled by M.

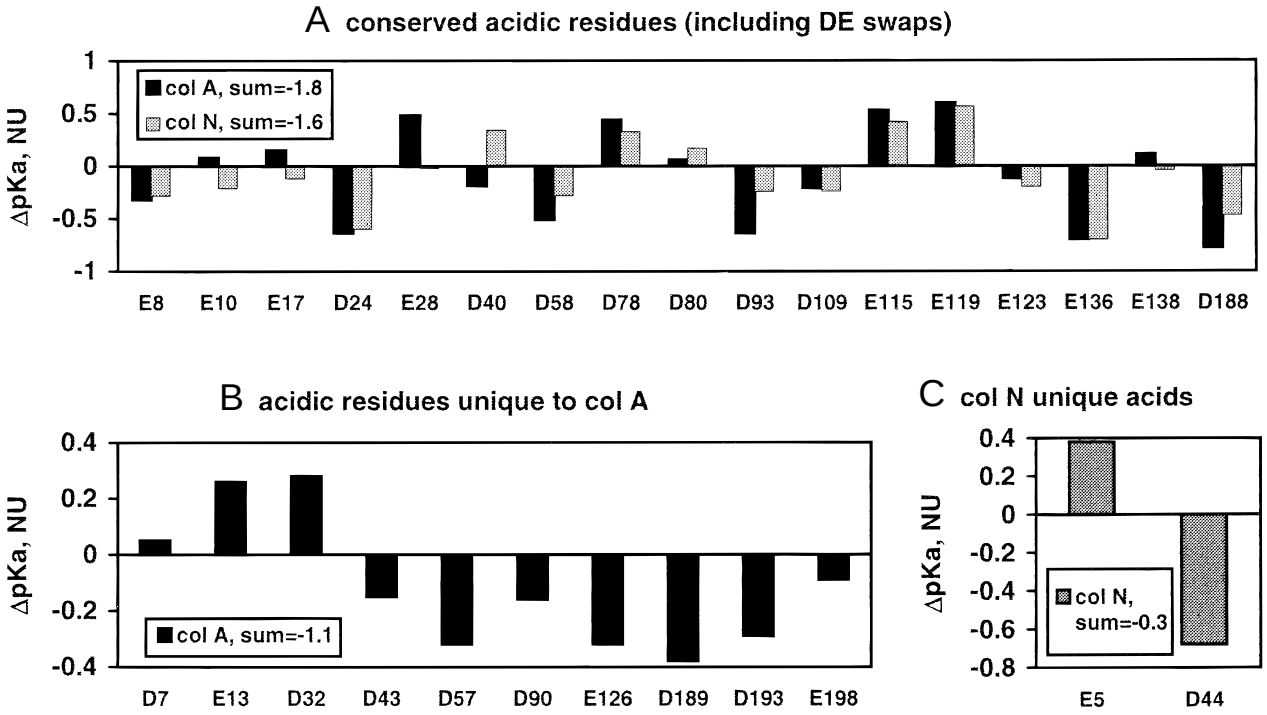

Fig. 6. Indications of ionizable group contributions to the relative acid pH-dependent stabilities of ThColA and ThColN. Acidic residue $\Delta pK_{a,NU}$s were calculated as $\Delta pK_{a,N} - \Delta pK_{a,U}$ with the u2 unfolded state model. Each panel gives sums over the displayed $\Delta pK_{a,NU}$ values. **A:** Acidic residues present in ThColA and ThColN. **B:** Acidic residues present in ThColA but not ThColN. **C:** Acidic residues present in ThColN but not ThColA.

A simple model for ionizable group interactions in the unfolded state

Energy-minimized extended chain structures have been introduced as models of the unfolded state for HEWL (Schaefer et al., 1997), giving significant pK_a shifts in FDPB calculations with $\epsilon_p = 20$, that are in accord with experiments on barnase (Oliveberg et al., 1995). The effectiveness of a relatively simple calculation scheme is now assessed, with DH modeling applied to a sequence-based model for ionizable group interactions in an extended polypeptide (Fig. 1). Results for two values of λ are compared with the measured pH-dependence of folding energy (ΔG_{NU}). The u1 model denotes $\lambda = 1$, while for the u2 model ionizable group charge is compressed onto the backbone ($\lambda = 0$). Dipolar interactions are neglected in the absence of specific unfolded state conformers, and the DH model lacks a self-energy term, leaving only interactions between ionizable groups. The programs written for this study may be obtained from the author.

Proteins

Test systems are ribonuclease A (RNase A), ribonuclease T_1 (RNase T_1), hen egg white lysozyme (HEWL), and barnase (Antosiewicz et al., 1994), with the inclusion of measured pK_as for barnase (Oliveberg et al., 1995). Protein Data Bank (PDB) (Bernstein et al., 1977) coordinate sets 3rn3 (RNase A), 3rnt (RNase T_1), 1bni (barnase), 2lzt (triclinic HEWL), and 1lyz (tetragonal HEWL) were used. Matching Antosiewicz et al. (1994), ionizable residues included in the RMSD to measured pK_as are: all D,E,H with N-t and C-t for RNase A (16 groups); H27, H40, H92, and E58 for RNase T_1; all D,E,H, except D119, with C-t for HEWL (10 groups). Matching Oliveberg et al. (1995), all D,E with C-t are included for barnase (13 groups). For two groups with pK_as < 2.0 (RNase A D14 and barnase D93), RMSD contributions are zero for calculated values <2.0 and the difference to 2.0 otherwise. Ionic strength for pK_a calculations matches experiment: 0.2 M for RNase A, 0.05 M for barnase, and 0.15 M for RNase T_1 and HEWL. Plots of the pH-dependence of stability are given as unfolding energies (ΔG_{UN} rather than ΔG_{NU}), and are set equal to experiment at the low pH limit since these calculations do not account for any terms other than the pH-dependent electrostatic energy. Ionic strength for pH-dependent stability calculations matches previous work (Antosiewicz et al., 1994): 30 mM for RNase A, RNase T_1, and barnase; 150 mM for HEWL.

The structure of the pore-forming domain of colicin A (Parker et al., 1989, 1992) was obtained from the PDB (1col) with coordinates for residues 5–201, missing 4 N-t and 3 C-t residues. This thermolytic fragment is denoted ThColA (Evans et al., 1996). A model for the equivalent domain of colicin N (ThColN) was built from ThColA and regularized around the single amino acid deletion (ThColA 165), using the program QUANTA (Molecular Simulations Inc., Burlington, Massachusetts) on a Silicon Graphics workstation. No other regularization or energy minimization was performed, giving a ThColN model as close as possible to the ThColA structure, ensuring that interactions between conserved pairs of ionizable groups cancel on comparison. Changes of most interest to acid pH-dependence are largely D/E deletions going from ThColA to ThColN, so that calculated differences will be based mostly on experimental, rather than modeled, structure. Colicin calculations were made at 0.3 M ionic strength to match experiment (Evans et al., 1996).

References

Antosiewicz J, McCammon JA, Gilson MK. 1994. Prediction of pH-dependent properties of proteins. *J Mol Biol 238*:415–436.

Antosiewicz J, McCammon JA, Gilson MK. 1996. The determinants of pK_as in proteins. *Biochemistry 35*:7819–7833.

Bashford D, Karplus M. 1990. pK_as of ionizable groups in proteins: Atomic detail from a continuum electrostatic model. *Biochemistry 29*:10219–10225.

Bernstein FC, Koetzle TF, Williams GJB, Meyer EF, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. 1977. The Protein Data Bank: A computer-based archival file for macromolecular structures. *J Mol Biol 112*:535–542.

Beroza P, Fredkin DR, Okamura MY, Feher G. 1991. Protonation of interacting residues in a protein by a Monte-Carlo method: Application to lysozyme and the photosynthetic reaction centre of *Rhodobacter sphaeroides*. *Proc Natl Acad Sci USA 88*:5804–5808.

Born M. 1920. Volumes and heats of hydration of ions. *Z Phys 1*:45–48.

Demchuk E, Wade RC. 1996. Improving the continuum dielectric approach to calculating pK_as of ionizable groups in proteins. *J Phys Chem 100*:17373–17387.

Evans LJA, Goble ML, Hales KA, Lakey JH. 1996. Different sensitivities to acid denaturation within a family of proteins: Implications for acid unfolding and membrane translocation. *Biochemistry 35*:13180–13185.

Fink AL, Calciano LJ, Goto Y, Kurotsu T, Palleros DR. 1994. Classification of acid denaturation of proteins: Intermediates and unfolded states. *Biochemistry 33*:12504–12511.

Gilson MK, Honig B. 1986. The dielectric constant of a folded protein. *Biopolymers 25*:2097–2119.

Karshikov A, Duerring M, Huber R. 1991. Role of electrostatic interaction in the stability of the hexamer of constitutive phycocyanin from *Fremyella diplosiphon*. *Protein Eng 4*:681–690.

Kesvatera T, Jönsson B, Thulin E, Linse S. 1994. Binding of Ca^{2+} to calbindin D_{9k}: Structural stability and function at high salt concentration. *Biochemistry 33*:14170–14176.

Kesvatera T, Jönsson B, Thulin E, Linse S. 1996. Measurement and modelling of sequence-specific pK_a values of lysine residues in calbindin D_{9k}. *J Mol Biol 259*:828–839.

Klapper I, Hagstrom R, Fine R, Sharp K, Honig B. 1986. Focussing of electric fields in the active site of Cu-Zn superoxide dismutase: Effects of ionic strength and amino acid modification. *Proteins 1*:47–59.

Kuramitsu S, Hamaguchi K. 1980. Analysis of the acid-base titration curve of hen lysozyme. *J Biochem 87*:1215–1219.

Lakey JH, Parker MW, González-Mañas JM, Duché D, Vriend G, Baty D, Pattus F. 1994. The role of electrostatic charge in the membrane insertion of colicin A: Calculation and mutation. *Eur J Biochem 220*:155–163.

Loewenthal R, Sancho J, Fersht AR. 1992. Histidine-aromatic interactions in barnase: Elevation of histidine pK_a and contribution to protein stability. *J Mol Biol 224*:759–770.

Oliveberg M, Arcus VL, Fersht AR. 1995. pK_A values of carboxyl groups in the native and denatured states of barnase: The pK_A values of the denatured state are on average 0.4 units lower than those of model compounds. *Biochemistry 34*:9424–9433.

Parker MW, Pattus F, Tucker AD, Tsernoglou D. 1989. Structure of the membrane-pore-forming fragment of colicin A. *Nature 337*:93–96.

Parker MW, Postma JPM, Pattus F, Tucker AD, Tsernoglou D. 1992. Refined structure of the pore-forming domain of colicin A at 2.4 Å resolution. *J Mol Biol 224*:639–657.

Schaefer M, Sommer M, Karplus M. 1997. pH-dependence of protein stability: Absolute electrostatic free energy differences between conformations. *J Phys Chem B 101*:1663–1683.

Sham YY, Chu ZT, Warshel A. 1997. Consistent calculations of pK_a's of ionizable residues in proteins: Semi-microscopic and microscopic approaches. *J Phys Chem B 101*:4458–4472.

Sham YY, Muegge I, Warshel A. 1998. The effect of protein relaxation on charge-charge interactions and dielectric constants of proteins. *Biophys J 74*:1744–1753.

van Gunsteren WF, Berendsen HJC. 1987. *GROMOS manual*. Groningen, Netherlands: University of Groningen.

Warwicker J. 1986. Continuum dielectric modelling of the protein-solvent system, and calculation of the long-range electrostatic field of the enzyme phosphoglycerate mutase. *J Theoret Biol 121*:199–210.

Warwicker J. 1997. Improving pK_a calculations with consideration of hydration entropy. *Protein Eng 10*:809–814.

Warwicker J. 1998. Modeling charge interactions and redox properties in DsbA. *J Biol Chem 273*:2501–2504.

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

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

Yang AS, Honig B. 1993. On the pH dependence of protein stability. *J Mol Biol 231*:459–474.

You TJ, Bashford D. 1995. Conformation and hydrogen-ion titration of proteins: A continuum electrostatic model with conformational flexibility. *Biophys J 69*:1721–1733.

Zhou HX, Vijayakumar M. 1997. Modeling of protein conformational fluctuations in pK_a predictions. *J Mol Biol 267*:1002–1011.