Coupling Between Conformation and Proton Binding in Proteins

Jorge A. Vila,[1,2] Daniel R. Ripoll,[3] Yelena A. Arnautova,[1] Yury N. Vorobjev,[4] and Harold A. Scheraga[1]*

[1]*Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca New York*

[2]*Universidad Nacional de San Luis, Facultad de Ciencias Físico Matemáticas y Naturales, Instituto de Matemática Aplicada San Luis, CONICET, San Luis-Argentina.*

[3]*Computational Biology Service Unit, Cornell Theory Center, Cornell University, Ithaca New York*

[4]*Institute of Chemical Biology and Fundamental Medicine of Siberian Brunch of Russian Academy of Science, Novosibirsk, Russia*

ABSTRACT **Interest centers here on whether the use of a fixed charge distribution of a protein solute, or a treatment that considers proton-binding equilibria by solving the Poisson equation, is a better approach to discriminate native from non-native conformations of proteins. In this analysis of the charge distribution of 7 proteins, we estimate the solvation free energy contribution to the total free energy by exploring the 2^{ζ} possible ionization states of the whole molecule, with ζ being the number of ionizable groups in the amino acid sequence, for every conformation in the ensembles of 7 proteins. As an additional consideration of the role of electrostatic interactions in determining the charge distribution of native folds, we carried out a comparison of alternative charge assignment models for the ionizable residues in a set of 21 native-like proteins. The results of this work indicate that (1) for 6 out of 7 proteins, estimation of solvent polarization based on the Generalized Born model with a fixed charge distribution provides the optimal trade-off between accuracy, with respect to the Poisson equation, and speed when compared to the accessible surface area model; for the seventh protein, consideration of *all* possible ionization states of the whole molecule appears to be crucial to discriminate the native from non-native conformations; (2) significant differences in the degree of ionization and hence the charge distribution for native folds are found between the different charge models examined; (3) the stability of the native state is determined by a delicate balance of *all* the energy components, and (4) conformational entropy, and hence the dynamics of folding, may play a crucial role for a successful ab initio protein folding prediction. Proteins 2005;61: 56–68.** © **2005 Wiley-Liss, Inc.**

Key words: charge distribution; electrostatics; solvation; protein conformation; proton binding

INTRODUCTION

An important aspect of polypeptides and proteins is that these biological molecules usually contain a large fraction of ionizable side-chain groups that can bind or release protons to become positively or negatively charged. When considering the ionizable residues, Asp, Glu, His, Lys, Tyr, and Arg, the average and standard deviation of the percent of ionizable groups in a sample of 106 proteins[1] is 26.3 ± 5.9%. The observed pK's of these ionizable groups depend on the conformation of the molecule and on the environment of these groups in the macromolecule.[2] Since charged groups may come spatially close to each other at intermediate conformations during folding, the equilibrium binding of protons should also vary. The conformation, in turn, is sensitive to the state of ionization of the individual amino acid residues,[3] and should accommodate to the particular state of charge. As a consequence, adoption of a fixed charge distribution during a simulation may introduce an undesired *bias* to the folding process. This means that the appropriate charge distribution should be assigned by solving the Poisson equation by considering the 2^{ζ} ionization states for every conformation, with ζ being the number of ionizable groups in the molecule. Although this approach has been used for an oligopeptide with a *limited* number of ionizable groups,[4] the inclusion of such level of detail in methodologies for ab initio protein structure prediction is not currently feasible for medium-size proteins (50–100 amino acids residues) with existing computational resources because, among other reasons, the fraction of ionizable residues is approximately 25% of the total number of residues in the protein. As a consequence, the number of conformations to be explored for a sequence containing N residues can be approximated by $\sim 2^{[N/4]}$

The Supplementary Materials referred to in this article can be found at http://www.interscience.wiley.com/jpages/0887-3585/suppmat/

Grant sponsor: National Institutes of Health; Grant numbers: GM-14312 and TW006335. Grant sponsor: National Science Foundation; Grant number: MCB00-03722. Grant sponsor: National Research Council of Argentina (CONICET); Grant number: PIP-02485. Grant sponsor: Universidad Nacional de San Luis (Argentina); Grant number: P-328402.

*Correspondence to: Harold A. Scheraga, Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca, NY 14853-1301. E-mail: has5@cornell.edu

Received 21 January 2005; Accepted 3 March 2005

Published online 3 August 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20531

ionization states (with the bracket representing the integer part of this fraction), and this number is quite large.

The theory of protein titration has been the subject of extensive research for many decades[5–19] because of its importance in the study of biological processes. Recent evidence[20] indicates that a correct description of electrostatic interactions that considers all states of ionization may be crucial for understanding protein stability and, consequently, to discriminate the native state from non-native conformations.

It should be noted that all physics-based scoring functions used recently to discriminate native from non-native folds make use of different levels of approximation to compute solvent polarization.[21–26] However, a common denominator of all of them is the assumption of a fixed charge distribution obtained by considering all ionizable residues as either neutral or as having a fixed charge of 0 or $\pm$1e. The state of charge of a given ionizable group is determined by the pH of the solution and the pK_a of the group according to the *Null* (or *Zero*) model[27] (i.e., the group titrates as if it is immersed in an aqueous solution with no perturbations from other ionizable groups).

To examine the relevance of the charge distribution to the discrimination of native folds, we carried out a theoretical analysis of the ensembles of conformations derived for 7 proteins, namely, 1e0l, 9api_B, 1gab, 1bdd, 1vii, 1res, and 1fsd. As a scoring function, we use 3 different forms of the potential energy function that differ in how solvation effects are treated. The following solvation models are considered: (1) a solvent-accessible surface area, represented here by the parameter set SRFOPT[28]; (2) a pairwise [Generalized Born (GB)] approximation to the solution of the Poisson–Boltzmann equation, as implemented, for example, by Ghosh et al.[29]; and (3) the Multigrid Boundary Element (MBE) method,[30–32] in which the free energy associated with the state of ionization of the ionizable groups, at a fixed pH value, is calculated by using the general multisite titration formalism. The latter approach has been used to predict several observables obtained from NMR at a given fixed pH, with good agreement.[4,33–35]

In these applications, many questions arise and are addressed here:

1. How cost-effective is the computation of the solvent effect when we consider the 2^{ζ} ionization states?
2. How does this cost-effectiveness change if we compute solvent polarization using faster and simpler models that ignore the proton binding/release equilibrium?
3. How elusive is the native fold?
4. Can the trivial *Null* (or *Zero*) model describe the charge distribution in native folds accurately?
5. Aside from errors in the potential function, and limitations on the efficiency of the search of the conformational space, what are the reasons that keep the ab initio protein folding problem from being solved despite the fact that existing scoring functions can discriminate the native from non-native folds?

This work is not intended to be a complete review of (1) methods for the computation of solvent polarization; (2) capabilities of different scoring functions, either knowledge- or physics-based, to discriminate native from non-native folds; or (3) approaches to compute proton binding/release equilibrium.

METHOD

Forms of the Potential Energy Function

Three alternative forms were used to compute the total free energy as a function of the coordinates $\mathbf{r_p}$ of the protein, namely,

1. *A gas phase potential plus a solvent-accessible surface area model to treat the solvent (GPSAS):*

$$E(\mathbf{r_p}) = E_{\text{int}}(\mathbf{r_p}) + F_{\text{sas}}(\mathbf{r_p}) \tag{1}$$

where $E_{\text{int}}(\mathbf{r_p})$ is the internal conformational energy of the molecule in the absence of solvent, assumed to correspond to the Emperical Conformational Energy Program for Peptides (ECEPP/3)[36–39] energy of a neutral molecule; and $F_{\text{sas}}(\mathbf{r_p})$ represents the solvation free energy as defined by Vila et al.[28]

2. *A gas phase potential plus a multigrid boundary element method to treat the solvent (GPBEM):*

$$E(\mathbf{r_p},\ \text{pH}) = E_{\text{int}}(\mathbf{r_p}) + F_{\text{cav}}(\mathbf{r_p}) + F_{\text{solv}}(\mathbf{r_p}) + F_{\text{inz}}(\mathbf{r_p},\ \text{pH}). \tag{2}$$

where $F_{\text{cav}}(\mathbf{r_p})$ is the free energy associated with the process of cavity creation when transferring the molecule from the gas phase into the aqueous solution, $F_{\text{solv}}(\mathbf{r_p})$ is the free energy associated with the polarization of the aqueous solution, and F_{inz} ($\mathbf{r_p}$, pH) is the free energy associated with the change in the state of ionization of the ionizable groups due to the transfer of the molecule from the gas phase to the solvent, at a fixed pH value. $F_{\text{cav}}(\mathbf{r_p})$ describes the free energy of creation of a cavity to accommodate a *zero-charge* peptide molecule (i.e., with all partial atomic charges set to *zero*). As shown previously,[40,41] $F_{\text{cav}}(\mathbf{r_p})$ can be considered as the free energy of transfer of a nonpolar molecule from the gas phase to water. This free energy is proportional to the solvent-accessible surface area of the molecule. The term $F_{\text{solv}}(\mathbf{r_p})$ is obtained by using the fast MBE method, and F_{inz} ($\mathbf{r_p}$, pH) is calculated by using the general multisite titration formalism.[30,42,43]

3. *A gas phase potential plus a GB model to treat the solvent (GPGB):*

$$E(\mathbf{r_p},\ \text{pH}) = E_{\text{int}}(\mathbf{r_p}) + F_{\text{GB}}(\mathbf{r_p}) \tag{3}$$

where F_{GB} ($\mathbf{r_p}$) represents the "pairwise" GB solvation model of Hawkins et al.,[44] as implemented by Ghosh et al.[29] (with the GB code provided by D. A. Case).

It should be noted that re-evaluation of the total energy of a given conformation with the GPBEM and GPGB energy functions [Eqs. (2) and (3), respectively] does not include local energy minimization. In addition, our estima-

TABLE I. Details of the Proteins Analyzed[a]

Proteins[b] (PDB code)	Secondary Structure Content[c] α-helix (%)	β-sheet (%)	Residues[d] (*N*)	Ionizable residues[e] (%)	Experimental method	Backbone heavy atoms RMSD[f] (Å)
1e01	0.0	35.1	37 (17)	45.9	NMR (pH = 6.5)	0.09
9api_B	0.0	38.9	36 (9)	25.0	X-ray diffraction (pH = 7.0)[g]	0.15
1gab	56.6	0.0	53 (20)	37.7	NMR (pH = 6.0)	0.09
1bdd	61.7	0.0	46 (15)	32.6	NMR (pH = 5.0)	0.14
1vii	52.8	0.0	36 (12)	33.3	NMR (pH = 3.7)	0.11
1res	53.5	0.0	43 (15)	34.9	NMR (pH = 6.1)	0.09
1fsd	32.1	0.0	28 (16)	57.1	NMR (pH = 5.0)	0.09

[a]For which an EDMC conformational search was carried out.
[b]Protein code, as listed in the PDB.
[c]According to the Rost & Sander[1] classification: (α-helix, 3_{10}-helix, π-helix) → α-helix, and (extended strand) → β-sheet
[d]Total number of residues in the sequence. All the proteins were considered with the N- and C-terminal unblocked. The numbers of ionizable residues, including the N- and C-terminal groups, are given in parentheses.
[e]In boldface, we denote the highest percentage of ionizable residues seen among all proteins analyzed and listed in Tables I–III.
[f]RMSD for the superposition of the NMR or X-ray structure on the structure obtained from it by optimizing the geometry to that used in the ECEPP/3 potential.
[g]When the pH of the crystallization was not reported in the PDB, a value of 7.0 was adopted.

TABLE II. Summary of the EDMC Runs

Proteins (PDB code)	No. of (accepted) conformations[a]	RMSD Range[b] (Å)	Discrimination of native from non-native folds[c] GPSAS	GPGB	GPBEM
1e0l	4218	0.1–30.0	No	Yes	Yes
9api_B	692	0.1–16.0	No	No	Yes
1gab	2019	0.1–23.0	No	Yes	Yes
1bdd	1067	0.1–20.0	Yes	Yes	Yes
1vii	2114	0.9–13.0	Yes	Yes	Yes
1res	1952	0.1–18.0	No	Yes	Yes
1fsd	2754	0.1–13.0	No	Yes	Yes

[a]According to the Metropolis criterion. This column indicates the total number of conformations accepted after EDMC runs starting from the *native, helix,* and *random,* structures, respectively, as explained in the Method section.
[b]Minimum and maximum values of the RMSD, with respect to the native structure, seen for the ensemble of conformations generated.
[c]These columns contain the answer to the question: Do GPSAS, GPGB, and GPBEM discriminate nativelike folds from non-native (decoys)?

tion of the free energy, as given by Eqs. (1) through (3), does not contain terms accounting for vibrational entropy (due to local fluctuations) and conformational entropy (due to large-scale conformational variations). The conformational entropy is not computed, because a sufficient number of conformations is not obtained in the search procedure used here. The vibrational entropy can be treated by a harmonic approximation[45,46] for each conformation obtained by using the ECEPP/3 potential function. However, the values of the vibrational entropy for native, misfolded or denaturated conformations are similar[21,26,47] and small compared to the large energy contributions, while evaluation of this term is computationally expensive.

Generation of the Ensemble of Conformations (Decoys)

For each protein listed in Table I, we generated an ensemble of conformations, starting from (1) the native fold, (2) the canonical α-helix, and (3) a randomly generated conformation, by using the EDMC (Electrostatically Driven Monte Carlo) method,[4,48] with the GPSAS potential function given by Eq. (1). For each protein, the total generated ensemble from all 3 starting points varied from 692 to 4218 conformations and is characterized by a uniform distribution of root-mean-square deviation (RMSD) from the native fold in the range 0.1–30.0 Å. The results of the search are shown in Table II. A description of each starting conformation follows.

1. Native fold: conversion to fixed-geometry (ECEPP/3-optimized) conformation

The coordinates for the native structure of each protein used in this work (listed in the first column of Tables I through III) were obtained from the Protein Data Bank (PDB) and subsequently optimized to ECEPP/3 geometry (i.e., with fixed bond lengths and bond angles). This

TABLE III. Testing the Null Model[a]

Proteins (PDB code)	Secondary structure content α-helix (%)	Secondary structure content β-helix (%)	Residues[b] (*N*)	Experimental method[c]	Backbone heavy atoms RMSD[d] (Å)	Disagreement in the assignment of charge[e] (%)
1sh1	0.0	35.4	48 (16)	NMR (pH 5.0)	0.2	**50** (20)
4cpa_I	0.0	15.8	38 (9)	X-ray	0.2	22 (0)
6hir	0.0	12.3	65 (10)	NMR (pH 7.0)	0.1	10 (10)
2mhu	0.0	0.0	30 (7)	NMR (pH 7.0)	0.0	0 (0)
2or1_L	53.6	0.0	63 (14)	X-ray	0.3	7 (0)
2tgp_I	13.8	24.1	58 (18)	X-ray	0.2	0 (0)
1bds	0.0	27.9	43 (10)	NMR (pH 7.0)	0.2	20 (20)
4rxn	0.0	14.8	54 (23)	X-ray	0.3	30 (22)
1ppt	50.0	0.0	36 (13)	X-ray	0.2	8 (8)
1ovo_A	17.9	12.5	56 (15)	X-ray	0.2	0 (0)
3icb	50.7	0.0	75 (28)	X-ray	0.2	21 (11)
2ltn_B	7.7	0.0	46 (12)	X-ray	0.1	0 (0)
1tgs_I	16.1	12.5	56 (13)	X-ray	0.2	15 (15)
1mrt	0.0	0.0	31 (6)	NMR (pH 7.0)	0.1	0 (0)
1fc2_C	36.2	0.0	45 (11)	X-ray	0.2	27 (27)
1crn	43.5	8.7	46 (6)	X-ray	0.2	17 (0)
1cdt_A	0.0	43.3	60 (17)	X-ray	0.3	41 (35)
2mev_4	5.7	0.0	58 (9)	X-ray	0.5	22 (0)
3ebx	0.0	43.6	62 (15)	X-ray	0.1	13 (0)
1cka_A	0.0	36.8	57 (26)	X-ray	0.1	23 (8)
1p7e_A	25.0	42.9	56 (19)	NMR (pH 6.5)	0.1	5 (0)

[a]Computation of the average degree of charge for each of the 21 listed proteins was carried out *only* for the ECEPP/3-optimized structure.
[b]This column denotes the number of residues in the amino acid sequence. The corresponding numbers of ionizable residues excluding the N- and C-terminal groups are given in parentheses. These end groups were excluded, because N- and C-termini are usually free to move, and hence could introduce a *bias* in the comparison shown in column 7.
[c]Computation of the average degree of charge for each ionizable residue was carried out at the indicated pH. The value of 7.0 was adopted for the pH when the experimental conditions were not reported in the PDB database.
[d]RMSD for the superposition of the NMR or X-ray structure on the structure obtained from it by optimizing the geometry to that used in the ECEPP/3 potential.
[e]This column denotes the percentage of *all* ionizable groups in the sequence that are in disagreement between the estimated value from the *Null* (or *Zero*) model and the average degree of charge computed by the GPBEM model, at the pH indicated in column 5. The values listed represent disagreement of 10% and 30% (in parentheses) between both predictions. We highlight the highest disagreement found in bold.

conversion provides an all-atom representation, including hydrogen atoms, for each of the selected proteins. With the few exceptions of very high-resolution structures that include those from neutron studies, X-ray crystallography does not identify the positions of hydrogen atoms. However, the positions of all the atoms in the structure are *crucial* for the level of detail of the current analysis.

The RMSD for backbone heavy atoms between the native structure before and after the ECEPP/3 optimization is very low, as can be seen from column 7 of Table I and column 6 of Table III (i.e., all the RMSDs are lower than 0.6 Å). For illustrative purposes only, superpositions of native and ECEPP/3-optimized structures for proteins 1fsd and 4cpa_I are shown in Fig. S1(A and B, respectively) in the Supplementary Material.

When more than one structure was present in the PDB, one of them was selected arbitrarily, and its geometry was ECEPP/3-optimized. In one case, all 10 NMR structures of 1e0l were ECEPP/3-optimized for later evaluation of the average charge distribution.

2. *Canonical α-helix*

For each protein listed in Table I, a canonical α-helix conformation was generated by assigning the values (−60.0°, −40.0°) to the dihedral angles (ϕ, ψ) of each amino acid residue, while keeping all values of ω as 180.0°.

3. *Randomly generated conformations*

For each protein listed in Table I, a randomly generated conformation was produced by assigning to the dihedral angles of each amino acid residue values chosen randomly from the ranges $-180.0° \leq \phi \leq 180.0°$, $-180.0° \leq \psi \leq 180.0°$ and assuming all values of ω to be 180.0°.

Experimental Data Used

Seven proteins used for generating an ensemble of conformations are listed in Table I. Two of them are β-sheet folds (1e0l and 9api_B). The remaining 5 are α-helical folds. The highest percentage of ionizable residues was found for the α-helical protein 1fsd (57.1%) and for the β-sheet protein, 1e0l (45.9%).

To test whether the *Null* (or *Zero*) model is an accurate approximation to assign the charge distribution for native-like (ECEPP/3-optimized) structures, a test of 21 proteins selected from the PDB, namely, those listed in Table III, with less than 30 ionizable residues, was carried out. The upper limit of 30 ionizable residues was adopted because energy evaluations with the GPBEM function require

exploring the 2^{ζ} ionization states for each conformation, with ζ being the number of ionizable residues. As an example, for one conformation of protein 3icb, with 30 ionizable groups (2 of them representing the unblocked N- and C-terminal groups) a single energy evaluation required ~6 h and 17 min on a single Athlon 2800 processor.

RESULTS AND DISCUSSION

Discrimination of Native From Non-Native Folds: Advantages and Disadvantages of Different Approaches Used to Compute Solvent Polarization

Six of the seven proteins, listed in Table I were determined by NMR spectroscopy. The experimental conditions, also listed in Table I, cover a broad range of pH, ranging from 3.7 (for 1vii) to 7.0 (for 9api_B). For each of the proteins, we carried out an EDMC conformational search starting from native, helix and randomly generated conformations, as described in the Method section. Table II, columns 2 and 3, gives specific information about the number of accepted conformations (by the Metropolis criterion) for all of the mentioned starting structures, as well as the whole range of backbone heavy-atom RMSDs covered by the corresponding ensemble of conformations. Figure 1(a) (for 1e0l) and Figure 2(a) (for 9api_B) illustrate the wide range of RMSD covered.

The last three columns of Table II indicate whether the energy of the models used to compute solvent polarization discriminates native from non-native folds. We discriminate in terms of fragments of the protein in regular secondary structure, as in α-helices or β-sheets. The word "YES" means that the identified lowest-energy structure in the ensemble (1) belongs to the same secondary structure class as the native one, (2) contains its secondary structure motifs in the correct location within the amino-acid sequence (with an allowed shift of up to 3 residues), and (3) is packed in the same way as in the native structure in more than 50% of these secondary structure motifs. If any of these 3 conditions is not met, the word "NO" is used. It should be noted that the criterion to discriminate native from non-native folds is not based on an RMSD analysis.

It is clear from Table II that use of the GPBEM potential function, which explicitly considers the coupling between conformation and degree of ionization, provides the best results. Our implementation of the GB model, in the GPGB potential function, also leads to very good scores (i.e., 6 of 7 native folds were also discriminated). Comparison of (a) and (b) in Figure 1 for 1e0l shows that both models behaved quite similarly in this test [illustrated in Fig. S2(A and B) in the Supplementary Material]. On the other hand, comparison of (a) and (b) in Figure 2 for 9api_B shows the only case for which the GPGB potential function failed to discriminate the native from the non-native conformation [illustrated in Fig. S3(A and B) in the Supplementary Material]. From Table II, it is clear that the poorest results are obtained with the GPSAS potential, in accord with a recent analysis made by Lee and Duan.[26] Even though we did not try any other existing solvent-accessible surface area models, such as OONS,[49] it seems that their applicability is, as far as we know, limited to some α-helix motifs. Nevertheless, failure to discriminate other α-helix motifs such as 1gab, 1res, or 1fsd, as well as β-sheet folds (as can be seen from Table II) may indicate that the effectiveness of GPSAS as a discriminative tool is limited.

A comparison of the efficiency of the GPBEM and GPGB potential functions to discriminate native from non-native conformations shows that there is no correlation between total free energy and RMSD [see Figs. 1 and 2] in accord with previous work.[23,50] In fact, we observed a significant anticorrelation between solvent polarization [$F_{solv}(\mathbf{r_p})$ in Eq. (2)] and total internal energy [E_{int} ($\mathbf{r_p}$) in Eqs. (1) through (3)], as shown in Fig. 3. A similar anticorrelation, but between solvent polarization and internal electrostatic energy, was already noted by Vorobjev and Hermans.[21–23] As indicated clearly in Figure 3, for protein 1gab, the solvent polarization free energy favors unfolded (random) conformations, whereas more compact conformations, such as α-helical structures, have more favorable total internal energy. A similar pattern to the one illustrated in Figure 3 was observed for *all* 7 proteins analyzed in this work.

The computed solvent polarization for the accepted ensemble of conformations of protein 1e0l by using both the BEM and GB models, is well correlated (i.e., with a value of $R = 0.99$, and the slope of the correlation equal to 0.65). An analysis regarding structural similarity between the lowest energy conformations identified with both the GPBEM and GPGB methods shows that (1) for two proteins (i.e., 1bdd and 1res), the *same* conformation in the ensemble was identified as the lowest energy one using both the GPGB and GPBEM potential energy functions [Fig. S4(A and B) in the Supplementary Material]; (2) for three proteins, namely, 1e0l, 1gab, 1vii, and 1fsd, *different* conformations, but having the native-like fold (see Figs. S2 and S5 in the Supplementary Material), were identified as the lowest energy structures using the GPGB or GPBEM functions; and (3) for the remaining protein (i.e., 9api_B), the lowest energy conformation identified by GPGB does not have the nativelike fold [see Fig. S3(A) in the Supplementary Material], while the one identified by GPBEM does [see Fig. S3(B) in the Supplementary Material]. In all the runs that make use of the GB model to estimate solvent polarization, all the ionizable groups were first assumed to be uncharged. To test whether the assignment of full charges to the ionizable groups could improve the discriminating power of the GPGB potential energy function, new energy evaluations of the whole ensemble of conformations generated for 9api_B were carried out. The charge distribution was estimated with the *Null* (or *Zero*) approximation model. Despite considering either two such charge distribution models or different salt concentrations, namely, 5.0 *M* and 0.001 *M*, it was not possible to discriminate the native from the non-native structures of 9api_B using our current implementation of the GB model. A better treatment, involving the GB model, would treat all 2^{ζ} possible ionization states explicitly for all conformations, but we have not yet developed such a treatment.

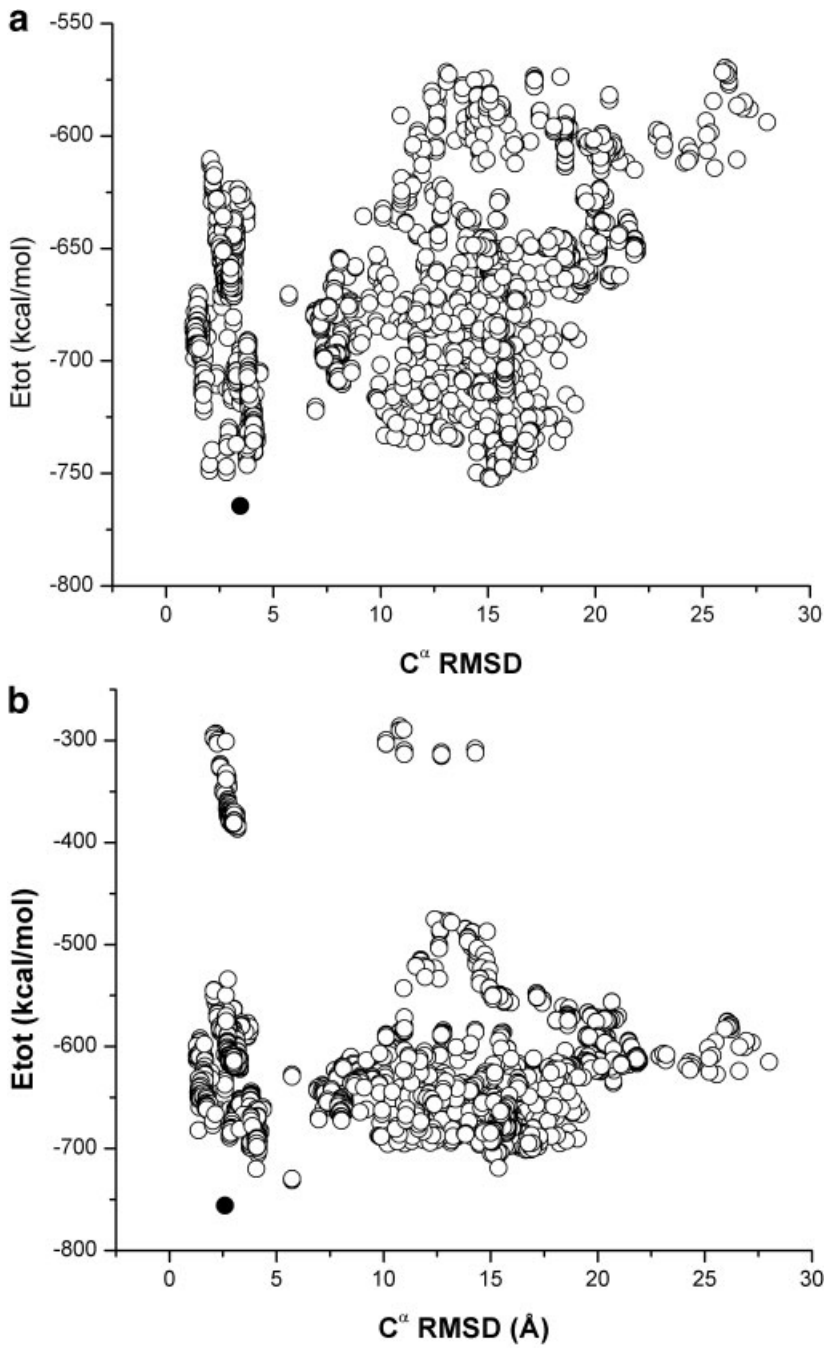

Fig. 1. (**a**) Total energy versus C^{α} RMSD for 1e0l, computed by Eq. (3) (i.e., using the GB solvation model to estimate solvent polarization). Each point in the graph corresponds to an accepted conformation (see row 1 of Table II). The filled black circle indicates the lowest-energy conformation identified with this potential energy function [superposition of this structure with the native ECEPP/3-opptimized one is shown in Fig. S2(A) in the Supplementary Material]. (**b**) Total energy versus C^{α} RMSD for 1e0l, computed by Eq. (2) using the BEM solvation model that considers all 2^{17} degrees of ionization, at pH 6.5, to estimate solvent polarization. Each point in the graph corresponds to an accepted conformation (see row 1 of Table II). The filled black circle indicates the lowest energy conformation identified with this potential energy function [superposition of this structure with the native ECEPP/3-optimized one is shown in Fig. S2(B) in the Supplementary Material].

How Elusive Is the Native Fold?

A few years ago, Lazaridis and Karplus[50] carried out an extensive analysis of the discrimination of the native from misfolded protein models by using an energy function that includes implicit solvation. In their study they concluded that "…Once certain side chains are displaced from the correct orientation, the energy is nearly equivalent to many other unfolded or misfolded

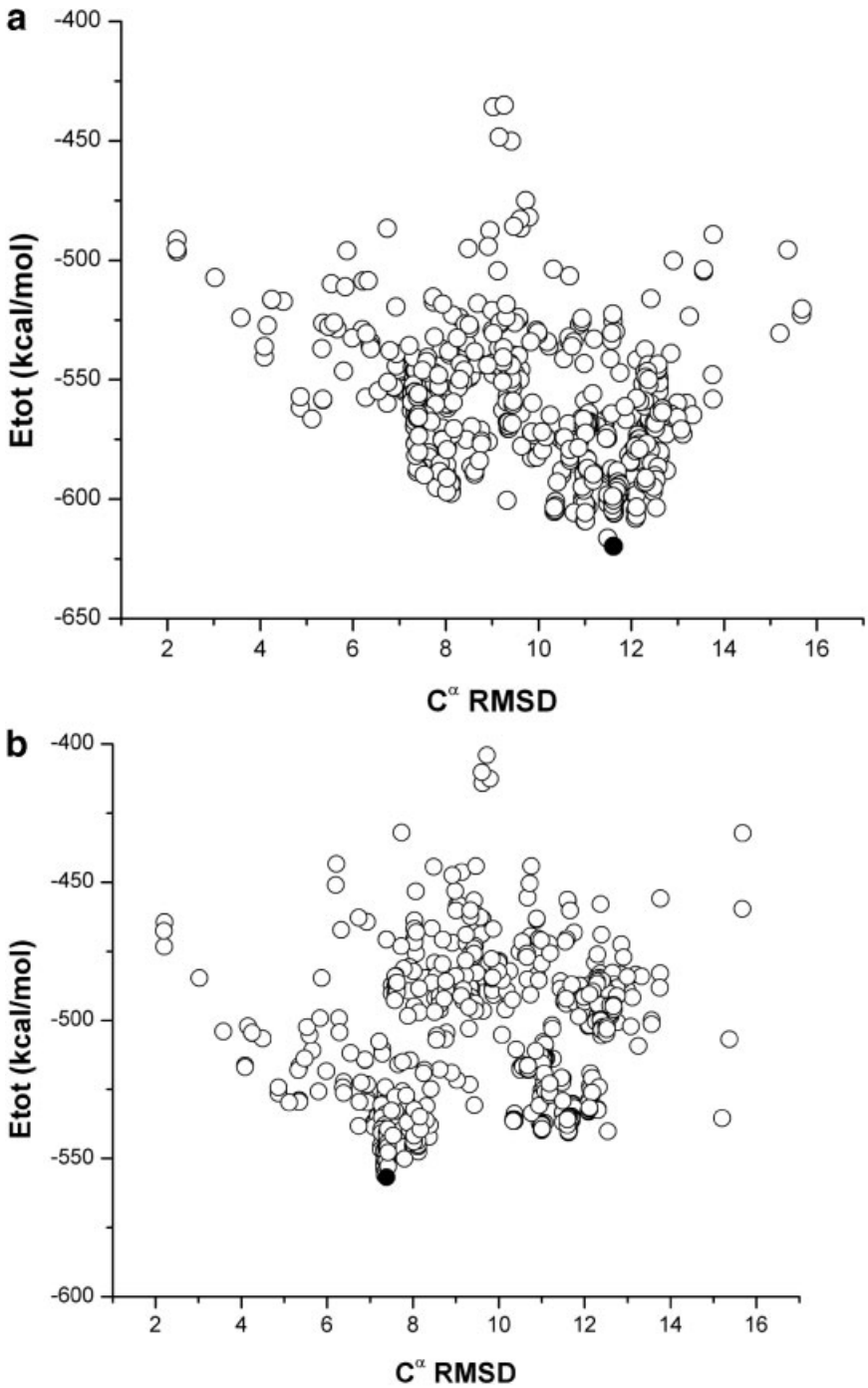

Fig. 2. (**a**) Total energy versus C^{α} RMSD for 9api_B, computed by Eq. (3) (i.e., using the GB solvation model to estimate solvent polarization). Each point in the graph corresponds to an accepted conformation (see row 2 of Table II). The filled-black circle indicates the lowest-energy conformation identified with this potential energy function [superposition of this structure with the native ECEPP/3-optimized one is shown in Fig. S3(A) in the Supplementary Material]. (**b**) Total energy versus C^{α} RMSD for 9api_B, computed by Eq. (2) using the BEM solvation model that considers all 2^9 degrees of ionization, at pH 7.0, to estimate solvent polarization. Each point in the graph corresponds to an accepted conformation (see row 2 of Table II). The filled-black circle indicates the lowest energy conformation identified with this potential energy function [superposition of this structure with the native ECEPP/3-optimized one is shown in Fig. S3(B) in the Supplementary Material].

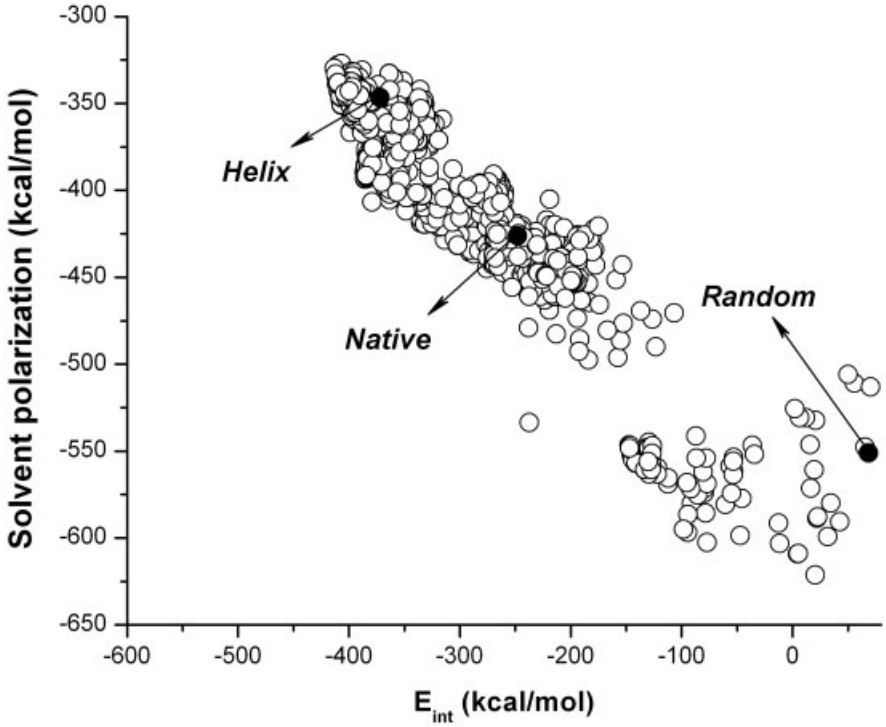

Fig. 3. Solvent polarization given by the term $F_{solv}(\mathbf{r_p})$ in Eq. (2) versus total internal energy as given by the term E_{int} ($\mathbf{r_p}$) in Eqs. (1) through (3). Filled black circles indicate the positions of the initial randomly-generated, native and all-α-helical structures used during the EDMC runs for protein 1gab.

conformations." We investigated this behavior to obtain some insight about the accuracy that a given energy function should have in order to recognize a native fold, in an attempt to shed more light on the inherent difficulties in protein structure prediction.

To carry out this analysis, the lowest energy conformation identified for the protein 1e0l with the GPBEM function was chosen. This particular protein was selected because (1) it represents the smallest monomeric triple-stranded β-sheet protein domain stable in the absence of disulfide bonds, bound ions or ligands[51]; (2) it does not have a formal hydrophobic core, even though some hydrophobic residues are highly conserved; (3) it contains a large number of ionizable residues (45.9%) and would enable us to monitor changes in the average charge as a consequence of variable ionizable side-chain positions; (4) the NMR experiments carried out at pH 6.5 in 30 m*M* NaCl, so the corresponding Debye length is ~17 Å, suggest that screening should not affect the electrostatic interactions significantly; and (5) analysis of the lowest-energy conformation identified in the ensemble of 4218 conformations mentioned in Table II revealed that Lys13 is fully deprotonated at pH 6.5 (see Fig. 4); visual inspection of this structure revealed that the Lys13 side-chain is partially buried leading to an unfavorable solvation of the amino group (see Fig. S6 in the Supplementary Material). To test the Lazaridis and Karplus conclusions, we varied the dihedral angle χ^1 by ~90°, followed by local energy-minimization with Eq. (1); this exposes the amino group to the solvent completely, while maintaining the rest of the structure

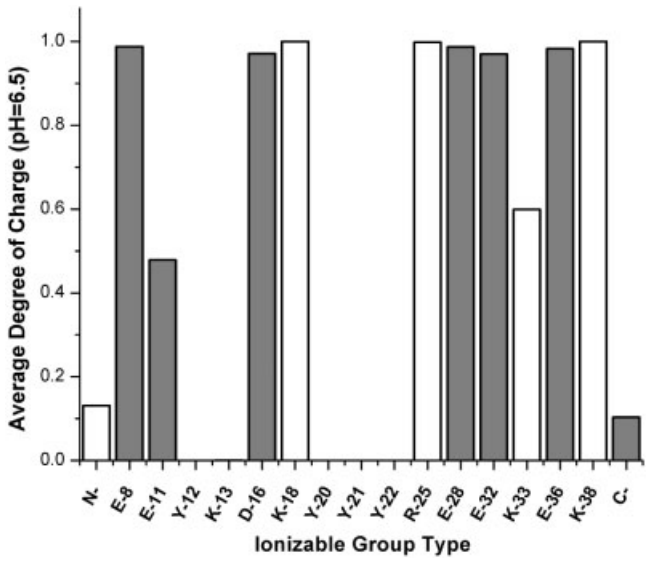

Fig. 4. Bars indicate the average value of the charge for each ionizable group, computed at pH 6.5, for the lowest energy conformation identified in the ensemble of 4218 conformations generated for the unblocked protein 1e0l. Gray-filled bars pertain to acid groups and white bars to basic groups. The N- and C-termini are indicated as N- and C-, respectively. The one-letter code is used for the ionizable groups, followed by a number that represents its position in the 1e0l sequence. The values of 3.90, 4.3, 10.50, 10.10, and 12.50 were adopted as pK_a^o for the ionizable groups for residues Asp, Glu, Lys, Tyr, and Arg, respectively, as an average from the data of Perrin.[52] Values of 7.80 and 3.75 for the α-amino and α-carboxyl groups[53] were used for the pK_a^o of the ionizable N- and C-terminal groups, respectively.

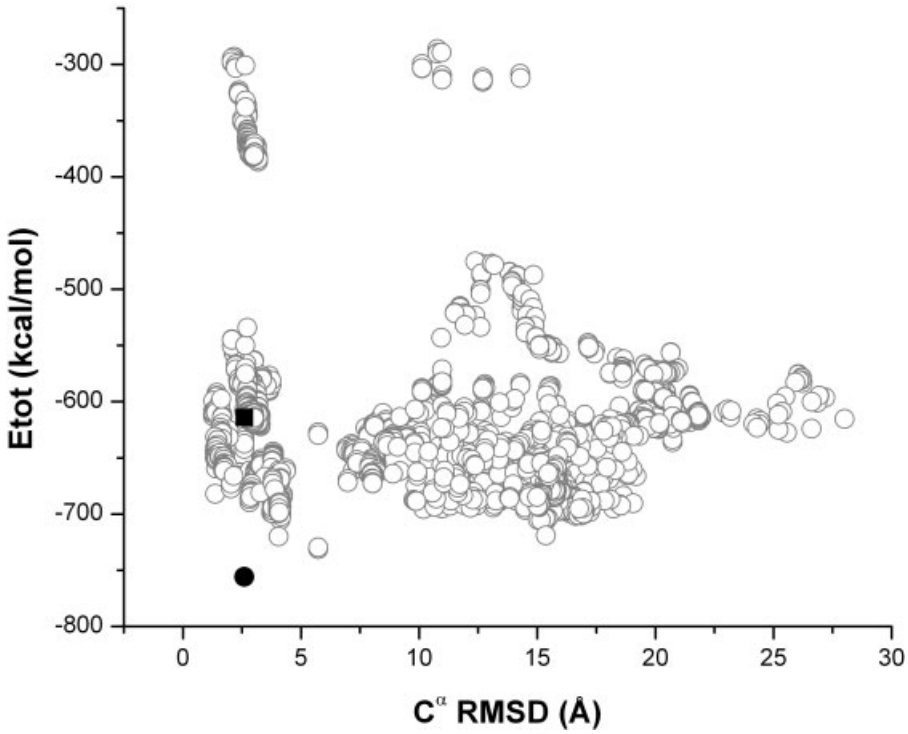

Fig. 5. Total energy versus C^{α} RMSD, computed by Eq. (2) (i.e., using the GPBEM energy function) for protein 1e0l. Each accepted conformation is indicated by an open circle. The black filled circle indicates the position for the lowest energy conformation identified with this potential energy function during the EDMC runs. The black filled square corresponds to the conformation derived from the lowest energy one *after* the side-chain of Lys13 was modified (see Results and Discussion section for details).

unchanged. Re-evaluation of the energy of the side-chain-modified conformation with the GPBEM showed that (1) Lys13 is fully protonated; (2) the total free energy of the new conformation is ~140 kcal/mol higher (black-filled square in Fig. 5) than the lowest energy structure from which it was derived (black-filled circle in Fig. 5); and (3) the all heavy-atom RMSDs with respect to the native structure for both the new and the lowest energy conformations from which it was derived, are *almost* identical (i.e., RMSD of 2.60 Å). In summary, the side-chain-modified conformation is now part of the ensemble containing misfolded structures, as shown in Figure 5 as a black-filled square, and in line with the Lazaridis and Karplus[50] observation that different side-chain conformations can lead to large changes in the total energy.

From this analysis, some questions arise:

1. Is the observed difference of ~140 kcal/mol in the total free energy a consequence *only* of the change in the protonation of Lys13? Figure 6 illustrates the changes in the average degree of charge *after* modification of the Lys13 side-chain. From Figure 6, it can be seen that, among other small changes, a sizable decrease in the average degree of charge for both Glu11 and Lys33 takes place. In this conformation of 1e0l, the distance between Lys33 N^{ζ} and Glu11 $O^{\epsilon 2}$ is 5 Å. Certainly, they participate in a salt bridge and, what is more important, are 14 Å away from Lys13 N^{ζ} (i.e., at a distance shorter than the Debye length), which means that the electrostatic potential between these groups is not screened significantly. This represents an example of the cooperative interactions that take place in proteins.
2. Can Lys13 be deprotonated in the native fold? Evaluation of the average degree of charge of Lys13 for the 10

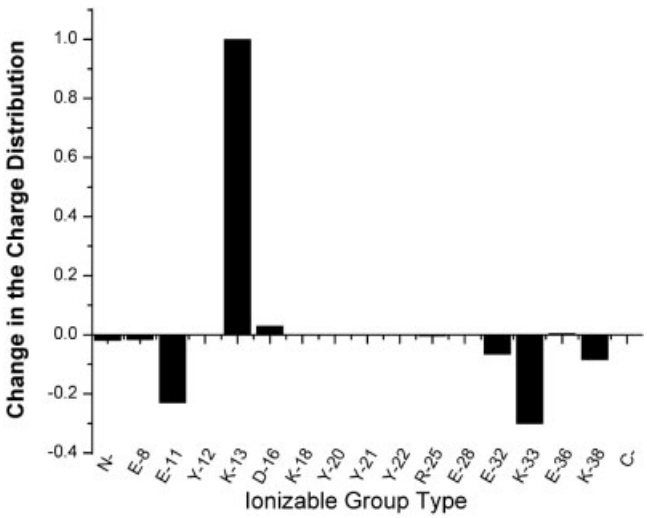

Fig. 6. Bars indicate the difference in the net charge of each ionizable group between the lowest energy conformation described in Figure 5 and the conformation obtained after the side-chain of Lys13 was modified (black-filled circle and square, respectively, in Fig. 5).

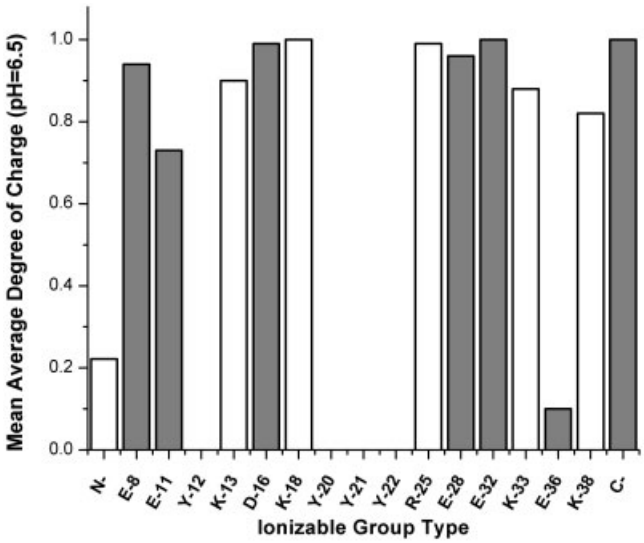

Fig. 7. Bars indicate the value of the charge averaged over 10 conformations of the NMR-derived models for the native structure of 1e0l, computed at pH 6.5. Gray filled bars pertain to acid groups and white bars to basic groups. The N- and C-termini are indicated as N- and C-, respectively. The 1-letter code is used for the ionizable groups, followed by a number that represents their position in the 1e0l sequence. The values of 3.90, 4.3, 10.50, 10.10, and 12.50 were adopted as pK_a^o for the ionizable groups for residues Asp, Glu, Lys, Tyr, and Arg, respectively, as an average from the data of Perrin.[52] Values of 7.80 and 3.75 for the α-amino and α-carboxyl groups[53] were used for the pK_a^o of the ionizable N- and C-terminal groups, respectively.

models representing the native structure satisfying the NMR constraints revealed that only 1 out of 10 appears to be fully deprotonated. This calculation leads to a value of 0.9 for the average degree of ionization of Lys13 (as shown in Fig. 7). To be more specific, the only deprotonated Lys13 side-chain is found for the initial NMR structure from which the ensemble of conformations shown in Figure 1(a and b) was derived.

These results obtained for the β-sheet protein (1e0l) are in complete agreement with the similar calculations carried out for an α-helical protein (1vii).[20] Results from 1vii for the computed average degree of charge from the structures satisfying the NMR constraints show that assignment of an integer number to the charge of each ionizable group [as in the *Null* (or *Zero*) model] may represent a poor approximation, because the *native fold* (which is an ensemble of NMR structures) cannot be represented by a *unique* structure. To investigate further whether an integer charge value (0 or 1) for each ionizable residue agrees with our estimation of the averaged degree of charge, we analyzed the charge distribution of *all* the ionizable amino acid residues of 21 proteins listed in Table III, and the results are discussed in the next subsection.

Indeed, all the accumulated evidence indicates that the stability of the native state is determined by a delicate balance of *all* the energy components, in line with experimental[54,55] and theoretical[50] observations.

Charge Distribution in Native Folds: Is the *Null* (or *Zero*) Model Accurate Enough?

Based on the calculation of the pK_a's of ionizable groups in proteins, [11,12,27] the accumulated evidence shows that the *Null* (or *Zero*) model is frequently a fairly good approximation, even though it is physically unrealistic. For example, it is common to observe titration curves displaying unusual shapes, when compared with the typical Henderson–Hasselbalch-type curves. This behavior is a consequence of the strong site–site couplings between ionizable groups.[2,56] Such interactions are expected to occur frequently in proteins because (1) they exhibit a high percentage of ionizable amino acid residues; (2) charged groups may come spatially close to each other during folding, especially for intermediates for which the equilibrium binding of protons should play a very important role; and (3) the long-range nature of the electrostatic interactions, as discussed for protein 1e0l.

To further understand the changes in the charge distribution of native folds by using different models to assign charges (*Null* or BEM, respectively), 21 ECEPP/3-optimized proteins (listed in Table III) have been used to compute the average degree of charge for each ionizable amino acid residue. To decide if the average degree of charge computed by the BEM is *significantly* different from the assignment of the *Null* (or *Zero*) model (1/0), the following criterion was adopted: If the average degree of charge computed for a given ionizable residue differs by more than a predefined percentage from the integer values 1 or 0, the charge assignment is considered *different*. Predefined difference values, Δ, of 10% or 30% between both assignments were used here. These values were adopted by analogy with the average degree of charge computed for residues Lys13 and Glu11, respectively, from 10 native conformations of protein 1e0l (as described in the previous subsection).

From Table III, we can conclude that the *Null* model correctly predicts the charge distribution for only 25–50% of the analyzed proteins (see Table III) depending on the value of Δ adopted.

A Comparison of CPU Times Between GPGB and GPBEM

A comparison of the CPU time required to compute solvent polarization for 9api_B shows that evaluation of the total free energy with the GPBEM function is 2 orders of magnitude more time-consuming than the evaluation using GPGB (the time for GPGB is too short to show in Fig. 8). In open squares, Figure 8 shows how the computation of the solvent polarization with the BEM scales with the number of residues (proteins 1fsd, 9api_B, 1res, 4rxn, and 3icb containing 28, 36, 43, 54, and 75 residues, respectively). To scale a BEM application to larger proteins, it would be necessary to speed up the computation of the solution of the Poisson equation. With this goal in mind, we are currently developing a new algorithm to estimate the last 2 terms in Eq. (2) (i.e., F_{solv} and F_{inz}) faster. The new algorithm, Fambe2[57] for solving the Poisson equation is a multilevel generalization of the current 3-level bound-

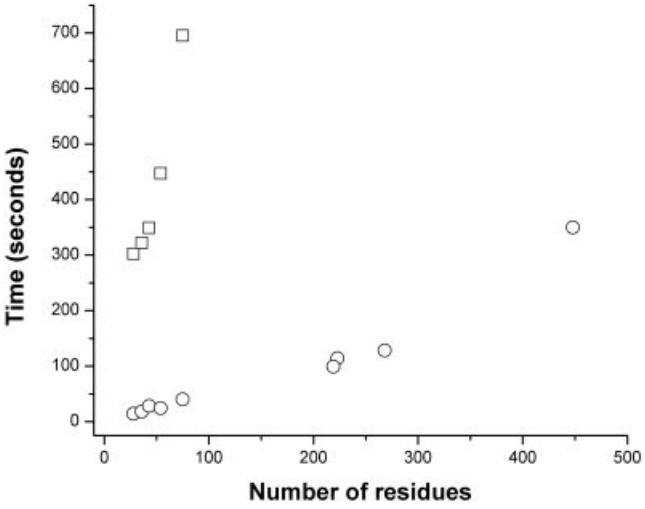

Fig. 8. Time for computation of solvent polarization as a function of the number of residues, using (**a**) the BEM (open squares: proteins 1fsd, 9api_B, 1res, 4rxn, and 3icb), and (**b**) the new Fambe2 algorithm[57] (open circles: proteins 1fsd, 9api_B, 1res, 4rxn, 3icb, 1jc9, 1bty, 1k82, 1eal). Calculations were carried out on an Athlon 2800 processor.

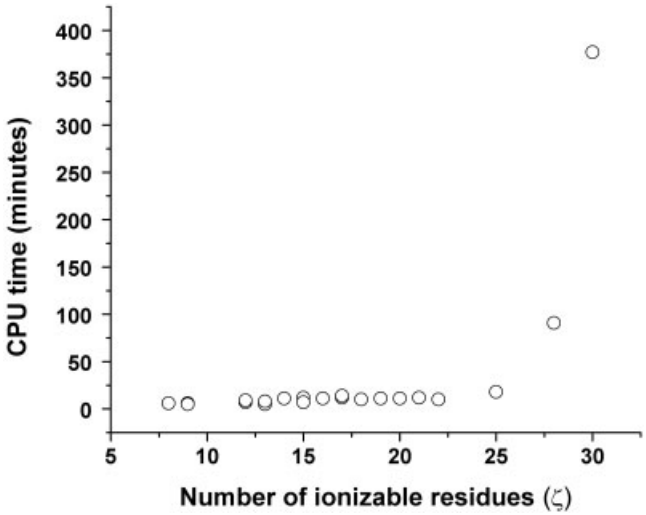

Fig. 9. CPU time required for the computation of solvent polarization by the BEM as a function of the number of ionizable residues (ζ), with ζ ranging from 7 to 28 (for the proteins listed in Table III).

ary element method.[32] The new algorithm also utilizes a smoother dielectric surface interface calculation.[32] An effective multigrid tessellation of the dielectric surface guided by the size of the protein molecule provides a method with linear scaling over protein size. Figure 8 shows a comparison, in terms of CPU time between the new Fambe2 algorithm (open circles) and the current BEM used in this work (open squares), for calculating F_{solv} in Eq. (2). Significant speedups, by 1 order of magnitude, in the computation of the solvent polarization are observed with the new Fambe2 algorithm.[57]

Figure 9 shows how the CPU time with BEM scales with the number of ionizable residues. Here, it can be observed that there is a plateau in the CPU time corresponding to the sequences containing less than $\zeta = 20$ ionizable residues. This plateau is a consequence of the fact that the computation of the 2^{ζ} states of ionization requires only a small fraction of the CPU time needed to compute solvent polarization when the number of ionizable groups is less than 20. A faster method for calculating the ionization energy when the molecule contains more than 20 such groups is in progress. It is based on a Monte Carlo calculation of the equilibrium degree of ionization and would provide a calculation of the free energy of ionization with the Tanford–Schellman titration approach.[58,59] The cost-effectiveness of this new method, together with the faster Fambe2 algorithm, is expected to make this procedure competitive with faster, though less accurate, alternative approximations such as the GB models.

CONCLUSIONS

There are at least 2 different but related issues that emerged from the current study. One has to do with the determination of the optimal function, among those used in this work, that discriminates native from non-native folds while representing the best trade-off between computational speed and accuracy. The results shown in Table II demonstrated that the GPSAS model has limited success in discriminating native from non-native structures when compared to GPGB or GPBEM. On the other hand, a fast model for estimation of solvent polarization, such as the GB model, improves the success of the potential energy function remarkably. A comparison with the more precise, but also more CPU-time-consuming method, that considers all possible degrees of ionization, as given by the GPBEM function, indicates that the GPGB approximation may be the best alternative. It provides the optimal trade-off between accuracy, with respect to the GPBEM model, and speed, when compared to the GPSAS model. Certainly, the speed of the calculations is a very important factor in many applications. However, in many others, such as the determination of a protein-folding pathway for a given sequence of amino acids, the indispensable *accuracy* of the calculations may require consideration of the coupling between folding and proton binding/release equilibrium. From this point of view, the *Null* model may not be a sufficiently accurate approximation, since it predicts the charge distribution correctly for only 25–50% of the analyzed proteins.

Another issue discussed in this work relates to the prediction of tertiary structure by physics-based methods. It is known that a *necessary* condition for a successful ab initio protein folding prediction is that the energy function used should be capable of discriminating the native fold among non-native ones. From this point of view, the current work provides some hints about the weakness and strength of different approaches for computing solvent polarization. However, the inclusion of the polarization contribution to the potential function is not a *sufficient* condition. The potential energy function must also be able to *guide* the formation of the native fold, when starting from randomly generated initial conformations. This requirement represents a challenging, yet unsolved, problem.

To predict tertiary structure, starting from an arbitrary conformation, an accurate computation of the total free energy should include a proper estimation of *all* free-energy contributions responsible for the stability of proteins. This means that an accurate estimation of the *entropy*, which is missing in our current estimation of the total free energy, has to be included. By considering how polymer structures are distributed in the conformational space, Sullivan and Kuntz[60] provided an estimate of the entropy change during folding. Their findings indicate that the entropy contribution appears to be essential. Changes in the vibrational entropy contribution for native, misfolded, or denaturated conformations appear to be roughly the same[21,26,47]; hence, attention should be focus on the conformational entropy contribution to the total free energy. Then, the following question arises: Why are energy functions that ignore this contribution able to distinguish native from non-native folds? To provide a tentative answer, we should consider that (1) explicit and implicit models used to estimate solvent polarization correlate very well among themselves; however, an implicit solvent model systematically overestimates the magnitude of this contribution[61]; and (2) conformational entropy and solvent polarization both favor unfolded versus folded structures. The preference of solvent polarization for unfolded conformations is clearly illustrated in Figure 3. Conceivably, an overestimated contribution of solvent polarization could, to some extent, compensate for the absence of the conformational entropy component. This fortuitous cancellation of errors may provide good results, as in some applications discussed here. However, solvation free energy *favors* conformations where all the polar and ionizable groups are well exposed to the solvent. During a conformational search, this preference imposes severe restrictions as to which *unfolded* conformations will be sampled. Clearly, in a conformational search, in which the conformational entropy is missing, the overestimation of the free energy of solvation could mislead the folding process.

We can conclude that a successful ab initio prediction would require (1) developing faster and more accurate methods for computing solvent polarization, and (2) a better estimation of the conformational entropy (e.g., by molecular dynamics).

ACKNOWLEDGMENTS

Part of this research was conducted using the resources of the Cornell Theory Center, which receives funding from Cornell University, New York State, federal agencies, foundations, and corporate partners, the National Partnership for Advanced Computational Infrastructure at the Pittsburgh Supercomputing Center which is supported in part by the National Science Foundation (MCA99S007P).

REFERENCES

1. Rost B, Sander C. Prediction of protein secondary structure at better than 70-percent accuracy. J Mol Biol 1993;232:584–599
2. Laskowski M Jr, Scheraga HA. Thermodynamic considerations of protein reactions: I. Modified reactivity of polar groups. J Am Chem Soc 1954;76:6305–6319.
3. Neurath H, Greenstein JP, Putnam FW, Erickson JO. The chemistry of protein denaturation. Chem Rev 1944;34:157–265.
4. Ripoll DR, Vorobjev YN, Liwo A, Vila JA, Scheraga HA. Coupling between folding and ionization equilibria: effects of pH on the conformational preferences of polypeptides. J Mol Biol 1996;264: 770–783.
5. Linderstrøm-Lang KU. On the ionization of proteins. Comp Rendus du Trav Lab Carlsberg Sér Chim 1924;15:1–29.
6. Hill TL. Titration curves and ion binding on proteins, nucleic acids, and other macromolecules with a random distribution of binding sites of several types, J Am Chem Soc 1956;78;5527–5529.
7. Tanford C, Kirkwood JG. Theory of protein titration curves: 1. General equations for impenetrable spheres. J Am Chem Soc 1957;79:5333–5339.
8. Warshel A. Calculations of enzymatic-reactions—calculations of pK_a, proton-transfer reactions, and general acid catalysis reactions in enzymes. Biochemistry 1981;20:3167–3177.
9. Bashford D, Karplus M. PK_a's of ionizable groups in proteins—atomic detail from a continuum electrostatic model. Biochemistry 1990;29:10219–10225.
10. Beroza P, Fredkin DR, Okamura MY, Feher G. Protonation of interacting residues in a protein by a Monte-Carlo method—application to lysozyme and the photosynthetic reaction center of Rhodobacter–Sphaeroides. Proc Natl Acad Sci USA 1991;88:5804–5808.
11. Yang A-S, Honig B. On the pH-dependence of protein stability. J Mol Biol 1993;231:459–474.
12. Antosiewicz J, McCammon JA, Gilson MK. The determinants of pK(a)s in proteins. Biochemistry 1996;35:7819–7833.
13. Zhou H-X, Vijayakumar M. Modeling of protein conformational fluctuations in pK(a) predictions. J Mol Biol 1997;267:1002–1011.
14. van Vlijmen HWT, Schaefer M, Karplus M. Improving the accuracy of protein pK(a) calculations: Conformational averaging versus the average structure. Proteins 1998;33:145–158.
15. Koumanov A, Karshikoff A, Friis EP, Borchert TV. Conformational averaging in pK calculations: improvement and limitations in prediction of ionization properties of proteins. J Phys Chem B 2001;105:9339–9344.
16. Gorfe AA, Ferrara P, Caflisch A, Marti DN, Bosshard HR, Jelesarov I. Calculation of protein ionization equilibria with conformational sampling: pK(a) of a model leucine zipper, GCN4 and barnase. Proteins 2002;46:41–60.
17. Georgescu RE, Alexov EG, Gunner MR. Combining conformational flexibility and continuum electrostatics for calculating pK(a)s in proteins. Biophys J 2002;83:1731–1748.
18. Alexov E. Role of the protein side-chain fluctuations on the strength of pair-wise electrostatic interactions: comparing experimental with computed pK(a)s. Proteins 2003;50:94–103.
19. Laurents DV, Huyghues-Despointes BMP, Bruix M, Thurlkill RL, Schell D, Newsom S, Grimsley GR, Shaw KL, Treviño S, Rico M, Briggs JM, Antosiewicz JM, Scholtz JM, Pace CN. Charge–charge interactions are key determinants of the pK values of ionizable groups in ribonuclease Sa (pI = 3.5) and a basic variant (pI = 10.2). J Mol Biol 2003;325:1077–1092.
20. Ripoll DR, Vila JA, Scheraga HA. Folding of the villin headpiece subdomain from random structures: analysis of the charge distribution as a function of pH. J Mol Biol 2004;339:915–925.
21. Vorobjev YN, Almagro JC, Hermans J. Discrimination between native and intentionally misfolded conformations of proteins: ES/IS, a new method for calculating conformational free energy that uses both dynamics simulations with an explicit solvent and an implicit solvent continuum model. Proteins 1998;32:399–413.
22. Vorobjev YN, Hermans J. Estimation of conformational free energy by combining dynamics simulations with explicit solvent with an implicit solvent continuum model. Biophys Chem 1999;78: 195–205.
23. Vorobjev YN, Hermans J. Free energies of protein decoys provide insight into determinants of protein stability. Prot Sci 2001;10: 2498–2506.
24. Felts AK, Gallicchio E, Wallqvist A, Levy RM. Distinguishing conformations of proteins from decoys with an effective free energy estimator based on the OPLS all-atom force field and the surface generalized Born solvent model. Proteins 2002;48:404–422.
25. Hsieh M-J, Luo R. Physical scoring function based on AMBER force field and Poisson–Boltzmann implicit solvent for protein structure prediction. Proteins 2004;56:475–486.
26. Lee MC, Duan Y. Distinguish protein decoys by using a scoring function based on a new AMBER force field, short molecular

dynamics simulations, and the generalized Born solvent model. Proteins 2004;55:620–634.
27. Antosiewicz J, McCammon JA, Gilson MK. Prediction of pH-dependent properties of proteins. J Mol Biol 1994;238:415–436.
28. Vila J, Williams RL, Vásquez M, Scheraga HA. Empirical solvation models can be used to differentiate native from near-native conformations of bovine pancreatic trypsin-inhibitor. Proteins 1991;10:199–218.
29. Ghosh A, Elber R, Scheraga HA. An atomically detailed study of the folding pathways of protein A with the stochastic difference equation. Proc Natl Acad Sci USA 2002;99:10394–10398.
30. Vorobjev YN, Scheraga HA, Hitz B, Honig B. Theoretical modeling of electrostatic effects of titratable side-chain groups on protein conformation in a polar ionic solution: 1. Potential of mean force between charged lysine residues and titration of poly(L-lysine) in 95-percent methanol solution. J Phys Chem 1994;98:10940–10948.
31. Vorobjev YN, Scheraga HA, Honig B. Theoretical modeling of electrostatic effects of titratable side-chain groups on protein conformation in a polar ionic solution: 2. pH-induced helix-coil transition of poly(L-lysine) in water and methanol ionic-solutions. J Phys Chem 1995;99:7180–7187.
32. Vorobjev YN, Scheraga HA. A fast adaptive multigrid boundary element method for macromolecule electrostatic computations in solvent. J Comp Chem 1997;18:569–583.
33. Vila JA, Ripoll DR, Villegas ME, Vorobjev YN, Scheraga HA. Role of hydrophobicity and solvent-mediated charge–charge interactions in stabilizing alpha-helices. Biophys J 1998;75:2637–2646.
34. Vila JA, Ripoll DR, Vorobjev YN, Scheraga HA. Computation of the structure-dependent pK(a) shifts in a polypentapeptide of the poly[f(v)(IPGVG),f(E)(IPGEG)] family. J Phys Chem B 1998;102:3065–3067.
35. Makowska J, Bagińska K, Kasprzykowski F, Vila JA, Jagielska A, Liwo A, Chmurzyński L, Scheraga HA. Interplay of charge distribution and conformation in peptides: comparison of theory and experiment. Biopolymers 2005;80:214–224.
36. Momany FA, McGuire RF, Burgess AW, Scheraga HA. Energy parameters in polypeptides: 7. Geometric parameters, partial atomic charges, nonbonded interactions, hydrogen-bond interactions, and intrinsic torsional potentials for naturally occurring amino acids. J Phys Chem 1975;79:2361–2381.
37. Némethy G, Pottle MS, Scheraga HA. Energy parameters in polypeptides: 9. Updating of geometrical parameters, nonbonded interactions, and hydrogen-bond interactions for the naturally-occuring amino acids. J Phys Chem 1983;87:1883–1887.
38. Sippl MJ, Némethy G, Scheraga HA. Intermolecular potentials from crystal data: 6. Determination of empirical potentials for O-H…O=C hydrogen bonds from packing configurations. J Phys Chem 1984;88:6231–6233.
39. Némethy G, Gibson KD, Palmer KA, Yoon CN, Paterlini G, Zagari A, Rumsey S, Scheraga HA. Energy parameters in polypeptides: 10. Improved geometrical parameters and nonbonded interactions for use in the ECEPP/3 algorithm, with application to proline-containing peptides. J Phys Chem 1992;96:6472–6484.
40. Sitkoff D, Sharp KA, Honig B. Accurate calculation of hydration free energies using macroscopic solvent models. J Phys Chem 1994;98:1978–1988.
41. Simonson T, Brünger AT. Solvation free-energies estimated from macroscopic continuum theory—an accurate assessment. J Phys Chem 1994;98:4683–4694.
42. Bashford D, Karplus M. pK(a)s of ionizable groups in proteins—atomic detail from a continuum electrostatic model. Biochemistry 1990;29:10219–10225.
43. Yang A-S, Gunner MR, Sampogna R, Sharp K, Honig B. On the calculation of pK(a)s in proteins. Proteins 1993;15:252–265.
44. Hawkins GD, Cramer CJ, Truhlar DG. Pairwise solute descreening of solute charges from a dielectric medium. Chem Phys Lett 1995;246:122–129.
45. Gō N, Scheraga HA. Analysis of contribution of internal vibrations to statistical weights of equilibrium conformations of macromolecules. J Chem Phys 1969;51:4751–4767.
46. Zimmerman SS, Pottle MS, Némethy G, Scheraga HA. Conformational-analysis of 20 naturally occurring amino-acid residues using ECEPP. Macromolecules 1977;10:1–9.
47. Lee MR, Duan Y, Kollman PA. Use of MM-PB/SA in estimating the free energies of proteins: application to native, intermediates, and unfolded villin headpiece. Proteins 2000;39:309–316.
48. Ripoll DR, Scheraga HA. On the multiple-minima problem in the conformational-analysis of polypeptides: 2. An electrostatically driven Monte-Carlo method—tests on poly(L-alanine). Biopolymers 1988;27:1283–1303.
49. Ooi T, Oobatake M, Némethy G, Scheraga HA. Accessible surface areas as a measure of the thermodynamic parameters of hydration of peptides. Proc Natl Acad Sci USA 1978;84:3086–3090.
50. Lazaridis T, Karplus M. Discrimination of the native from misfolded protein models with an energy function including implicit solvation. J Mol Biol 1998;288:477–487.
51. Macias MJ, Gervais V, Civera C, Oschkinat H. Structural analysis of WW domains and design of a WW prototype. Nat Struct Biol 2000;7:375–379.
52. Perrin DD. Dissociation constants of organic bases in aqueous solution: supplement. London: Butterworths; 1972. p 402.
53. Edsall, JT, Wyman J. Biophysical chemistry. Vol. I. New York: Academic Press; 1958. p 536 (Table IX).
54. Matthews B. Genetic and structural analysis of the protein stability problem. Biochemistry 1987;26:6885–6888.
55. Alber T. Mutational effects on protein stability. Annu Rev Biochem 1989;58:765–798.
56. Bashford D. Macroscopic electrostatic models for protonation states in proteins. Front Biosci 2004;9:1082–1099
57. Vorobjev YN. Fambe2: an improved version of the Fast Adaptive Boundary Element Method (unpublished).
58. Tanford C. Protein denaturation: Part C. Theoretical models for the mechanism of denaturation. Adv Protein Chem 1970;24:1–95.
59. Schellman JA. Macromolecular binding. Biopolymers 1975;14:999–1018.
60. Sullivan DC, Kuntz ID. Distributions in protein conformation space: implications for structure prediction and entropy. Biophys J 2004;87:113–120.
61. Wagoner J, Baker NA. Solvation forces on biomolecular structures: a comparison of explicit solvent and Poisson–Boltzmann Models. J Comp Chem 2004;25:1623–1629.