On the orientation of the backbone dipoles in native folds

Daniel R. Ripoll*, Jorge A. Vila†‡, and Harold A. Scheraga†§

*Computational Biology Service Unit, Cornell Theory Center, Cornell University, Ithaca, NY 14853-3801; †Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca, NY 14853-1301; and ‡Facultad de Ciencias Físico Matemáticas y Naturales, Instituto de Matemática Aplicada San Luis, Universidad Nacional de San Luis, Consejo Nacional de Investigaciones Científicas y Técnicas de Argentina, Ejército de Los Andes, 950, 5700 San Luis, Argentina

Contributed by Harold A. Scheraga, April 4, 2005

The role of electrostatic interactions in determining the native fold of proteins has been investigated by analyzing the alignment of peptide bond dipole moments with the local electrostatic field generated by the rest of the molecule with and without solvent effects. This alignment was calculated for a set of 112 native proteins by using charges from a gas phase potential. Most of the peptide dipoles in this set of proteins are on average aligned with the electrostatic field. The dipole moments associated with α-helical conformations show the best alignment with the electrostatic field, followed by residues in β-strand conformations. The dipole moments associated with other secondary structure elements are on average better aligned than in randomly generated conformations. The alignment of a dipole with the local electrostatic field depends on both the topology of the native fold and the charge distribution assumed for all of the residues. The influences of (i) solvent effects, (ii) different sets of charges, and (iii) the charge distribution assumed for the whole molecule were examined with a subset of 22 proteins each of which contains <30 ionizable groups. The results show that alternative charge distribution models lead to significant differences among the associated electrostatic fields, whereas the electrostatic field is less sensitive to the particular set of the adopted charges themselves (empirical conformational energy program for peptides or parameters for solvation energy).

Charge–charge and ionizable group–peptide dipole interactions have been proposed to play a relevant role in the stabilization of proteins (1) and in determining the native fold (2). Although a quantitative analysis of the relative stability of the native fold requires consideration of a delicate balance among all physical forces, we concentrate here on how electrostatic interactions can be identified as an important determination of the backbone conformation of the native structure.

It is well known that the CO and NH dipoles in an α-helix are electrostatically aligned nearly parallel to the axis of the helix (3), and that the alignment of peptide dipoles in various types of secondary structure elements is significant in determining the 3D structure (4) and stability of globular proteins (5, 6). The concept of the preferential alignment of the backbone dipole moments (μ) with the local electrostatic field (E) (see Fig. 1) was used by Piela and Scheraga (7) and by Ripoll and Scheraga (8) to develop procedures to surmount the multiple-minima problem in the conformational analysis of peptides and proteins. The charges used in these procedures to define μ, shown in Fig. 1, are ECEPP/3 (empirical conformational energy program for peptides) charges (9) that differ from other charges, e.g., PARSE (parameters for solvation energy) charges (10).

Approximately 50% of the residues in proteins fit in the most common secondary structure classes (11), such as α-helix or β-sheet, and the rest are associated with statistical coil fragments or loop regions. The question then arises as to how the μ’s in proteins behave.

We view the native fold as the net effect of different interactions such as van der Waals, electrostatic, and solvent polarization, among others. If the electrostatic interactions were dominant, then it would be expected that all of the dipoles would be well aligned with the local electrostatic field. On the other hand, if the role of electrostatics is irrelevant, then the dipoles should not have a preferential orientation with respect to the local electrostatic field. The purpose of this work is to explore the extent of the alignment of the peptide units in the local electrostatic field and the dependence of this extent on the adopted representation of the electrostatic charge distribution. In an attempt to clarify the role of electrostatic interactions in determining the native conformation, we have analyzed a relatively large set of experimentally solved protein structures. Particular emphasis has been placed on comparing different representations of the electrostatic field.

Proteins contain ionizable groups whose ionization states depend on a number of variables, among them the pKa’s, the pH of the solution, and the interaction with neighboring groups. As a consequence, the charge distribution and the conformation of a protein are coupled (12). A computational treatment of this effect is quite difficult, and simplifications of this problem are adopted in most theoretical studies carried out at present. Typically, charge assignment in molecular mechanics and molecular dynamics calculations on proteins with explicit solvent are carried out by using the Null/Zero model (13). In this model, any ionizable residue is assigned a zero or ±1e charge (adding or subtracting a H⁺ atom) without consideration of nearby interacting groups. On the other hand, it is common in molecular mechanics calculations with implicit solvent to assume that all residues are neutral to avoid overestimating the electrostatic contributions.

Consequently, calculations of E in this work range from simple representations, in which the molecule is considered in isolation with all residues in a neutral state, to more sophisticated ones in which the problem of ionization equilibria, as well as the electrostatic effects of solvation, are considered (14).

The present study has also enabled us (i) to assess the accuracy of using a fixed (neutral or charged) distribution in molecular modeling and simulations, instead of the appropriate but more complex and time-consuming representation that considers variable ionization states of the ionizable residues; and (ii) to test the validity of the assumption that μ’s are locally optimally oriented with E. The latter assumption is used in protein structure prediction approaches such as the self-consistent electric field (SCEF) (7) and the electrostatically driven Monte Carlo (EDMC) (8) methods.

Abbreviations: BEM, boundary element method; ECEPP, empirical conformational energy program for peptides; LW, long window; PARSE, parameters for solvation energy; PDB, Protein Data Bank; SW, short window.

*To whom correspondence should be addressed. E-mail: has@cornell.edu.

© 2005 by The National Academy of Sciences of the USA
Materials and Methods

Conversion of the Native Structures from Flexible to Rigid ECEPP/3 Geometry. To carry out the present study, the experimentally determined conformations (i) were regularized, i.e., all residues were replaced by the standard ECEPP/3 residues in which bond angles and bond lengths are fixed (rigid geometry approximation), and (ii) hydrogen atoms were added.

The conversion process was carried out by generating the new rigid-geometry conformation from the N terminus by adding one residue at a time and minimizing the rms deviation (rmsd) between all heavy atoms in the generated fragment and the corresponding fragment in the experimental structure. The procedure was iterated until the C-terminal group was added to the chain. The final conformations resulting from this regularization procedure are quite close to the experimental ones in practically all of the cases, with rmsd values for all of the heavy atoms ranging from 0.1 to 0.8 Å.

The experimental set of structures contains 112 proteins and was obtained from the Protein Data Bank (PDB) (15). These proteins are identified by a four-letter PDB code followed in some cases by an extra single number or letter to indicate the proteins are identified by a four-letter PDB code followed in was obtained from the Protein Data Bank (PDB) (15). These atoms ranging from 0.1 to 0.8 Å.

Stiffness of the Polypeptide Chain. As for polymers in general (17), bond-angle restrictions in real polypeptide chains determine that the direction assumed by a given main chain bond is strongly influenced by the direction of a few preceding bonds in the chain. These restrictions give rise to a local stiffness of the protein backbone. To investigate the effect of this local stiffness of the polypeptide chain on the alignment between $\mu$ and $E$, we selected seven proteins (PDB ID codes 1BDD, 1EOL, 1FSD, 3ICB, 4RXN, 1RBP, and 2GBP) and generated 1,000 random conformations for each.

All backbone and side-chain dihedral angles of these random conformations were chosen randomly between 180.0° and $-180.0°$, with the exception of the dihedral angles $\omega$ of the peptide group that were always chosen in the trans (180.0°) conformation. Local energy minimization was carried out for each conformation by using a simplified force field based on a soft-sphere model (18) to remove atomic overlaps. No electrostatic term was included in these energy minimizations to avoid bias in the alignment of $\mu$ and $E$.

Computation of the Alignment Between $\mu$ and $E$. The current study was carried out by analyzing the alignment of the $\mu$'s with $E$'s in native proteins generated by (i) the protein molecule itself, and (ii) the protein molecule plus its aqueous solvent.

To assess the robustness of the computed results, we considered two different sets of partial atomic charges as given by ECEPP/3 (9, 19–21) and PARSE (10), respectively. In addition, we investigated the effects of the following alternative assignments of the charge distribution in the protein. (i) The ionizable residues were assigned charges obtained by solving the multiple titration binding problem, computed at the experimental pH. Computations of the average degree of ionization of ionizable residues in a subset of 22 proteins was carried out by using the multiboundary element method of Vorobjev and Scheraga (22). (ii) The ionizable residues were assigned charges dictated by the Null or Zero model (13), by adding or subtracting a proton (H$^+$), depending on the group's intrinsic pK$a$ and the pH of the solution, and by neglecting the effects of nearby interacting groups. (iii) All residues were assigned neutral charges.

The angle $\theta_i$ between the local electrostatic field, $E_i(r_i)$, and the $i$th dipole moment, $\mu_i(r_i)$ (see Fig. 1), is used as a measure of the alignment between the two vectors, and is computed as

$$\theta_i = \arccos(\mathbf{\mu}(r_i) \cdot E_i(r_i)/|\mathbf{\mu}(r_i)||E_i(r_i)|),$$

where

$$\mathbf{\mu}(r_i) = q_C(r_{C_i} - r_i) + q_O(r_{O_i} - r_i) + q_N_i(r_{N_i} - r_i) + q_{H_i}(r_{H_i} - r_i).$$

The quantities $E_i(r_i)$ and $\mu_i(r_i)$ are evaluated at position $r_i$ (with respect to an arbitrary origin). The vector $r_i$ is chosen, by convention, as a charge-weighted average of the positions of the C', O, N, and H atoms forming the $i$th peptide plane, i.e.,

$$r_i = (q_C(r_{C_i}) + q_O(r_{O_i}) + q_N_i(r_{N_i}) + q_{H_i}(r_{H_i}))/(|q_C| + |q_O| + |q_N_i| + |q_{H_i}|).$$

The computation of the local electrostatic field due to the protein molecule itself is carried out by using Coulomb's law:

![Fig. 1. Ball and stick representation of a peptide unit showing the atoms and bonds that lie in the peptide plane. The ECEPP/3 charges (in electronic charge units) of those atoms associated with $\mu$ are indicated. The expressions given by Eqs. 2 and 3, respectively, are used to compute $\mu$ and the position, $n_i$, where $\mu$ is located. The value of $E_i(n_i)$, the local electrostatic field, is computed either by solving the Poisson equation numerically when solvent is present, or by using Eq. 4 in the absence of solvent.](image-url)
\[
E_i(r_i) = \sum_j q_j (r_i - r_j) / (\varepsilon|r_i - r_j|^3),
\]

where the sum runs over all of the atoms of the molecule except (as indicated by the prime in the summation symbol) for (i) those forming the peptide unit \(i\) and (ii) a predetermined set associated with the local stiffness of the chain, \(q_j\) is the charge of atom \(j\) located at position \(r_j\) and \(\varepsilon\) is the dielectric constant in which the charges are embedded (the protein environment), assumed to be 4.

Calculations of \(E\) were carried out by considering two alternative groups of excluded atoms: (i) the short window (SW) containing atoms forming the peptide unit \(i\), as well as the preceding and following C\(^\alpha\) atoms (Fig. 1); and (ii) the long window (LW) containing atoms in the SW plus the atoms of the preceding and following peptide units. Our LW definition resembles the one used by Avbelj and Moult (23) to analyze the role of electrostatics of the main chain atoms on the conformational preference of amino acid residues in proteins.

Computation of the local electrostatic field due to both the protein molecule and its aqueous environment was carried out by solving the Poisson equation numerically with the program DELPHI (24). In these calculations, the molecule was considered as a low-dielectric cavity (\(\varepsilon_{prot} = 4\)) immersed in a cubic grid (65 \times 65 \times 65 grid points) assumed to be filled with water (\(\varepsilon_{H_2O} = 78.3\)). Partial charges were assigned to each of the atoms of the protein. The focusing feature from the program DELPHI with Coulombic boundary conditions was used to obtain precise values of the electrostatic field. The value of \(E\) at a given position (e.g., the position at which a particular \(\mu\) is computed) was obtained by interpolation between the grid values from the final DELPHI run.

Because the calculation of the local electrostatic field, \(E\), requires the exclusion of the charges associated with the \(i\)th dipole moment and the neighboring atoms (i.e., SW or LW; see the prime on \(\Sigma\) of Eq. 4), \(N - 1\) runs with the program DELPHI were needed to compute \(E\) at the position of each dipole in a protein that contains \(N\) residues.

In treating the ionizable residues, the charges were assigned by the following procedure. An atom \(i\), belonging to an ionizable residue \(j\), was assigned a charge, \(q_i\), that resulted from a linear interpolation between the uncharged and fully charged states, i.e.,

\[
q_i = q_{i0}^0 + (q_{i1}^1 - q_{i0}^0)\alpha_j,
\]

where \(q_{i0}^0\) is the partial charge of the \(i\)th atom in the uncharged state of the \(j\)th ionizable residue, \(q_{i1}^1\) is the partial charge of the \(i\)th atom in the charged state of the \(j\)th ionizable residue, and \(\alpha_j\) is the degree of ionization of the \(j\)th ionizable residue computed by the procedure adopted to obtain the charge distribution. A special treatment was needed for ionizable residues in the first and last positions in the sequence when the end-groups of the polypeptide chain were also ionizable. In those cases, a double interpolation scheme was derived to assign the charges described by

\[
q_i = q_{i0}^{00} + (q_{i1}^{01} - q_{i0}^{00})\alpha_j^{01} + [q_{i1}^{10} - (q_{i1}^{01})\alpha_j^{01}]\alpha_j^{10},
\]

where \(q_{i0}^{00}\) is the partial charge of the \(i\)th atom that depends on the state of ionization of the end group and the neighboring residue (if it is ionizable); the first superscript, \(x\), indicates the ionization state of the end group (\(x = 0\) indicates that it is in the uncharged state, and \(x = 1\) indicates that the end group is fully charged). Similarly, the second superscript, \(y\), indicates the ionization state of the neighboring ionizable residue (\(y = 0\) indicates that it is in the uncharged state, and \(y = 1\) indicates that the residue is fully ionized). Finally, \(\alpha_j^0\) and \(\alpha_j^1\) represent the degree of ionization of the end group and neighboring ionizable residue, respectively.

**Results**

**Analysis of the Dipole-Field Alignment in the Gas Phase. Randomly generated conformations.** The average angle between the \(\mu\)’s and \(E\)’s computed with the SW for the set of 112 proteins in the gas phase is \(\approx 46^\circ\). This result indicates that the local \(E\), on average, has a component parallel to \(\mu\) and pointing in the same direction. However, the bond-angle restrictions in the polypeptide chain can be largely responsible for the observed alignment between the \(\mu\)’s and \(E\)’s. To investigate further the effect of the local environment on the direction of the electrostatic field, a set of
randomly generated conformations was studied. The influence of the local stiffness on the direction of the electrostatic field in 1,000 randomly generated conformations is illustrated by an analysis of the alignment between the $\mu$'s and $\mathbf{E}$'s for seven proteins (Fig. 2).

Comparison of the average alignment, $\Theta$, using the SW ($\approx60^\circ$) and LW ($\approx84^\circ$), indicates that the local electrostatic field is strongly influenced by the local structure around the peptide unit. Because our aim is to investigate the effect of the long-range electrostatic interactions on the organization of the polypeptide chain in native folds, from here on, the analysis will be focused on the electrostatic field computed by using the LW in Eq. 4.

**Native proteins.** Analysis of the alignment of $\mu$ and $\mathbf{E}$ was carried out for different secondary structure elements in the set of 112 protein structures. The electrostatic field was computed by using Eq. 4 with the LW set of excluded atoms. The $\alpha$ structural class encompasses 30.7% of the total number of $\mu$'s in the set, while the $\beta$ and $\Lambda$ classes contain 16.1% and 53.2% of the $\mu$'s, respectively. The $\mu$'s in the $\alpha$ class show the best alignment with the electrostatic field with an average value of $31.1^\circ\pm16.4^\circ$. Even though $\mu$'s associated with the $\beta$ and $\Lambda$ classes show average values (50.5$^\circ\pm31.1^\circ$ and 61.5$^\circ\pm35.4^\circ$, respectively) lower than randomly generated conformations ($84^\circ\pm34.7^\circ$; see Fig. 2), their standard deviations are significantly larger than those of the $\alpha$ class.

**Analysis of the Dipole-Field Alignment in Aqueous Solvent. Alignment of $\mu$ and $\mathbf{E}$ in native proteins.** As already mentioned, the native structure should be the result of a balance of different interactions. In particular, the electrostatic effects include contributions from the solvent. To account for these contributions, the Poisson equation was solved numerically by using the program DELPHI, and the local electrostatic field at each dipole position was computed for a reduced set of proteins.

To assess the robustness of the computed results, the average value of the angle between $\mu$ and $\mathbf{E}$ was computed with two different sets of charges: the ECEPP/3 and PARSE charges, respectively. In addition, to estimate the differences between different charge-parameter sets and different representations of the charge distributions (i.e., alternative ionization states of the molecule), the angle between their respective electrostatic fields

---

**Table 1. Average $\Omega$ angle between alternative representations of the electrostatic field considering solvent.**

<table>
<thead>
<tr>
<th>Protein name,* PDB code</th>
<th>Disagreement in the assignment of charge† (%)</th>
<th>Average angle $\Omega$ between electrostatic field vectors,‡ degrees</th>
<th>Null/Zero model and BEM method§</th>
<th>Neutral and BEM method¶</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>ECEPP/3</td>
<td>PARSE</td>
<td>ECEPP/3</td>
</tr>
<tr>
<td>1BDI</td>
<td>65 (50)</td>
<td>21.8</td>
<td>12.4</td>
<td>19.4</td>
</tr>
<tr>
<td>1BDJ</td>
<td>20 (20)</td>
<td>18.7</td>
<td>17.6</td>
<td>20.2</td>
</tr>
<tr>
<td>1CDT.A</td>
<td>41 (35)</td>
<td>28.5</td>
<td>29.2</td>
<td>31.8</td>
</tr>
<tr>
<td>1CRN</td>
<td>17 (0)</td>
<td>7.6</td>
<td>5.0</td>
<td>17.9</td>
</tr>
<tr>
<td>1E0L</td>
<td>23 (18)</td>
<td>17.1</td>
<td>11.6</td>
<td>40.5</td>
</tr>
<tr>
<td>1FC2.C</td>
<td>27 (27)</td>
<td>7.7</td>
<td>4.4</td>
<td>25.6</td>
</tr>
<tr>
<td>1GA8</td>
<td>15 (15)</td>
<td>18.1</td>
<td>11.6</td>
<td>19.6</td>
</tr>
<tr>
<td>1MRT</td>
<td>0 (0)</td>
<td>2.7</td>
<td>2.1</td>
<td>29.8</td>
</tr>
<tr>
<td>1OVO.A</td>
<td>0 (0)</td>
<td>11.3</td>
<td>6.4</td>
<td>33.8</td>
</tr>
<tr>
<td>1PPT</td>
<td>8 (8)</td>
<td>8.7</td>
<td>4.7</td>
<td>24.0</td>
</tr>
<tr>
<td>1SH1</td>
<td>50 (20)</td>
<td>34.0</td>
<td>24.8</td>
<td>26.6</td>
</tr>
<tr>
<td>1TGS.1</td>
<td>15 (15)</td>
<td>18.0</td>
<td>9.1</td>
<td>33.7</td>
</tr>
<tr>
<td>1VII</td>
<td>25 (25)</td>
<td>4.6</td>
<td>2.6</td>
<td>16.2</td>
</tr>
<tr>
<td>2LTN.B</td>
<td>0 (0)</td>
<td>15.1</td>
<td>27.0</td>
<td>33.3</td>
</tr>
<tr>
<td>2MHU</td>
<td>0 (0)</td>
<td>11.4</td>
<td>8.6</td>
<td>27.4</td>
</tr>
<tr>
<td>2OR1.L</td>
<td>7 (0)</td>
<td>2.2</td>
<td>1.5</td>
<td>21.6</td>
</tr>
<tr>
<td>2TGP.J</td>
<td>0 (0)</td>
<td>32.3</td>
<td>21.9</td>
<td>32.9</td>
</tr>
<tr>
<td>3ICB</td>
<td>21 (11)</td>
<td>16.4</td>
<td>13.7</td>
<td>25.5</td>
</tr>
<tr>
<td>4CPA.J</td>
<td>22 (0)</td>
<td>11.9</td>
<td>12.4</td>
<td>55.3</td>
</tr>
<tr>
<td>4RXN</td>
<td>30 (22)</td>
<td>27.5</td>
<td>18.8</td>
<td>44.9</td>
</tr>
<tr>
<td>6HJR</td>
<td>10 (10)</td>
<td>14.8</td>
<td>8.7</td>
<td>38.9</td>
</tr>
<tr>
<td>9API.B</td>
<td>0 (0)</td>
<td>1.9</td>
<td>2.6</td>
<td>37.2</td>
</tr>
</tbody>
</table>

*Four-letter code used to designate the protein in the PDB. An extra single letter that follows the PDB code indicates the particular monomer used in the calculation.

†To decide whether the average degree of charge computed by the BEM is different from the assignment of the Null/Zero model (1/0), the following criterion was adopted. If the average degree of charge computed at a fixed pKa 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 of 10% or 30% between both assignments were used here. The values listed correspond to the percentage of all ionizable groups for which the BEM and Null/Zero model disagree in the charge assignment by >10% or, within parentheses, by >30%.

‡Angles between representations of electrostatic fields as explained in Difference between alternative electrostatic field representations in native proteins. For each protein, two values are listed; the first corresponds to calculations with the ECEPP/3 charge set, and the second one corresponds to calculations with the PARSE charge set.

§Values obtained with electrostatic fields computed by using the distributions of charges as given by the Null/Zero model or the BEM model.

¶Values obtained with electrostatic fields computed by using the distributions of charges as given by neutral approximation or the BEM model.
was computed. A distribution based on the solution of the Poisson equation that considers the ionization states of a given protein, with \( z \), the number of ionizable residues, is taken as the most accurate representation of electrostatic effects. Consequently, the values of \( E \) obtained in DELPHI calculations with the boundary element method (BEM) states of ionization (22) were used as a reference with which to measure the quality of the other two simplified representations of the charge distribution.

These particular calculations were carried out on a set of 22 native proteins. Average values of \( \theta \) over 22 proteins, for \( \mu \)'s associated with the \( \alpha \), \( \beta \) and \( \Lambda \) classes, were obtained by using charge distributions given by (i) a solution of the problem of multiple titration as given by the BEM, (ii) the Null/Zero approximation, and (iii) assuming neutral charges for all residues, respectively. Averages computed by using ECEPP/3 (see Fig. 3) and PARSE charges (data not shown) led to very similar values. Both sets of charges show that dipole moments are preferentially oriented with respect to the local electrostatic field, with \( \theta (\alpha) < \theta (\beta) < \theta (\Lambda) \), i.e., with a similar tendency as observed for gas phase calculations on the larger set of 112 proteins. Analysis of the angle between \( \mu \) and \( E \) (see Fig. 3) indicates also that the average values are affected only slightly by the assumed charge distribution.

**Difference between alternative electrostatic field representations in native proteins.** The previous analysis might, at first sight, lead to the conclusion that the optimal orientation of \( \mu \) and \( E \) does not depend on either the charge set or the adopted charge distribution or even on the solvent. However, this analysis, although important, should be complemented with the results of the relative orientations of the \( E \)'s because of the different sets of charges, their corresponding distribution, and by analysis of the solvent effect. The angles (\( \Omega \)) between representations of the electrostatic fields vary from protein to protein and are presented in Table 1. As expected, for most of the proteins, the Null/Zero distribution leads to \( E \)'s that are more similar to those obtained from the BEM distribution (\( \Omega \) values listed in the third and fourth columns of Table 1) than the \( E \)'s generated by the neutral distribution (\( \Omega \) values listed in the fifth and sixth of Table 1). Among all 22 proteins, the best agreements between the associated \( E \) and the reference from the BEM distribution were found for protein 9APLB with \( \Omega = 1.9^\circ \), using ECEPP/3 charges, and for protein 2OR1.L with \( \Omega = 1.5^\circ \), using PARSE charges.

Fig. 4 shows the values of the angle, \( \Omega_\alpha \), between \( E \) vectors computed at position \( \alpha \) (where \( \mu_\alpha \) is located), by using the Null/Zero and the BEM distributions (black bars) and the Neutral and BEM distributions (gray bars), using PARSE charges in both cases, for the protein 10VO_A. For this protein, using the Null/Zero charge distribution as an approximation of the more reliable BEM distribution leads to a more accurate representation of the electrostatic field than the use of the neutral charge distribution.

However, straightforward use of the Null/Zero charge distribution does not always guarantee an accurate representation of \( E \). Fig. 5 shows a similar diagram of the \( \Omega \) between \( E \) vectors computed by using the Null/Zero and the BEM distributions and ECEPP/3 charges for the protein 1SH1. The structure of this protein was obtained by NMR methods at pH 5. Computation of the average degree of charge of the ionizable groups shows that 6 of 18, namely the N-terminal group, Asp-7, Glu-8, Lys-32, Asp-42, and Arg-45, differ by at least 30% from the estimation provided by the Null/Zero model. For this protein, the agreement in the relative orientations of \( E \) between the Null/Zero and BEM distributions shown in Fig. 5 is quite poor.

These two examples illustrate the difficulties in replacing the actual charge distribution (i.e., the distribution derived from the BEM method) by an alternative one such as that based on neutral charges or the one derived from the Null/Zero model. Therefore, effort should be directed to include the most accurate charge distribution, such as that obtained with the BEM, for molecular simulations.

**Energetic analysis of the orientation of the backbone dipoles with the electrostatic field in native proteins.** The average dipole energy, \( -\langle \mu \cdot E \rangle \), and the average magnitude of \( E \), \( \langle |E| \rangle \), for each structural class was computed for the proteins listed in Table 1. Results from this analysis are presented in Table 2, which shows that the average alignment energies are significantly more negative for dipoles of the \( \alpha \) class than for those in \( \beta \) or \( \Lambda \) class,
The results from this work (see Analysis of the Dipole-Field Stabilization of the
ECEPP and PARSE sets, 4.8 Debye) for the ECEPP and PARSE sets,
respectively.

Table 2. Energetics of the alignment between \( \mu \) and \( E \)
in solution

<table>
<thead>
<tr>
<th>Structural class</th>
<th>Average dipole energy* ((-\langle \mu \rangle E)), kcal ( \cdot ) mol(^{-1} )</th>
<th>Average field† ( \langle E \rangle ), kcal mol(^{-1} ) ( \cdot ) e( \cdot )Å(^{-1} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha )</td>
<td>(-1.11 (-4.72) )</td>
<td>(2.52 (5.28)) 2.89 (5.52)</td>
</tr>
<tr>
<td>( \beta )</td>
<td>(-0.42 (-1.47) )</td>
<td>(1.12 (1.96)) 1.24 (2.00)</td>
</tr>
<tr>
<td>( \Lambda )</td>
<td>(-0.19 (-1.98) )</td>
<td>(3.45 (4.84)) 3.79 (5.05)</td>
</tr>
</tbody>
</table>

The set of 22 proteins listed in Table 1 was used for these computations.

*Values were computed as \(\langle 1/N \rangle \sum E_{i}^{\mu} \), where \( E_{i}^{\mu} \) is the dipole energy associated with each residue in the sequence. \( N \) is the number of residues in the sequence.

†Values were computed as \(\langle 1/N \rangle \sum E_{i} \), where \( E_{i} \) is the electric field strength at each residue.

The results from this work are consistent with previous studies (1–5) that postulate a relevant role for electrostatic interactions in stabilizing the native fold; (ii) argue strongly for a relevant role for electrostatic forces in determining the direction of the main chain in the native conformation; (iii) go beyond current knowledge by showing that the backbone dipole orientation with respect to the electrostatic field is not restricted to \( \alpha \)-helices and \( \beta \)-strands; and (iv) show the importance of an adequate choice of the charge distribution to represent the electrostatics.

In the same way that burial of nonpolar amino acids is an indication of the importance of hydrophobic effects in proteins, we consider that the alignment of the backbone dipole moments with the electrostatic field generated by the whole molecule and the surrounding solvent (but excluding the local environment of the dipole), described in this work, is a footprint that highlights the relevance of the electrostatic forces in determining the native fold.

Finally, although the multiple-minima problem remains as one of the most formidable unsolved challenges in the conformational analysis of proteins, the results of the present study validate the assumption of optimal alignment of \( \mu \) and \( E \) on which the SCEF and EDMC global optimization methods are based.

We thank A. Liwo for helpful discussions. This work was supported by grants from the National Institutes of Health (GM-14312 and TW006355) and the National Science Foundation (MCB00-03722). Support was also received from the National Research Council of Argentina (Consejo Nacional de Investigaciones Científicas y Técnicas de Argentina) (Grant PIP-02485) and from the Universidad Nacional de San Luis, Argentina (Grant P-1220402). This research was conducted in part by using the resources of the Cornell Theory Center, which receives funding from Cornell University, New York State, federal agencies, foundations, and corporate partners.

References