Parametrization of Backbone−Electrostatic and Multibody Contributions to the UNRES Force Field for Protein-Structure Prediction from Ab Initio Energy Surfaces of Model Systems†

Adam Liwo,[‡,§,||] Stanisław Ołdziej,[‡,§] Cezary Czaplewski,[‡,§] Urszula Kozłowska,[‡,§] and Harold A. Scheraga*[,§]

Faculty of Chemistry, University of Gdańsk, Sobieskiego 18, 80-952 Gdańsk, Poland, Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853-1301, and Academic Computer Center in Gdańsk TASK, Narutowicza 11/12, 80-952 Gdańsk, Poland

Received: July 14, 2003; In Final Form: October 3, 2003

The multibody terms pertaining to the correlation between backbone−local and backbone−electrostatic interactions in the UNRES force field for energy-based protein-structure prediction, developed in our laboratory, were reparametrized on the basis of the results of high-level ab initio calculations on relevant model systems. MP2/6-31G(d,p) ab initio calculations were carried out to evaluate the energy surfaces of pairs consisting of *N*-acetyl-*N*′-methylacetamide molecules (AcNHMe, which model a regular peptide group) and *N*-acetyl-*N*′,*N*′-dimethylacetamide molecules ($AcNMe_2$, which model a peptide group preceding proline) at various intermolecular distances and orientations. For each pair, the calculated ab initio energy surface was subsequently fitted by a sum of Coulombic and Lennard-Jones components. Then, the restricted free-energy (RFE) surfaces of pairs of free peptide groups as well as the RFE factors corresponding to the coupling of backbone−local and backbone−electrostatic interactions in model tetrapeptides were calculated by numerical integration, with the use of the ab initio-fitted simplified energy functions and the ab initio energy maps of model terminally blocked amino acid residues calculated recently (Ołdziej, S.; Kozłowska, U.; Liwo, A.; Scheraga, H. A. *J. Phys. Chem. B*, in press, 2003). Next, analytical expressions based on Kubo's generalized cumulant theory from our previous work were fitted to the resulting RFE surfaces to parametrize the backbone−electrostatic and multibody terms in the UNRES force field. The computed coefficients of the cumulant-based expressions are different from those derived earlier, which had been based on the ECEPP/3 force field. To complete the force-field parametrization, the weights of the energy terms were determined, and the coefficients of the cumulant-based expressions were refined simultaneously by using our recently developed method of hierarchical optimization of a protein energy landscape using the protein 1IGD. The resulting force field was able to predict significant portions of the structures of proteins with α, β as well as both α and β structure correctly.

1. Introduction

In our continuing effort to develop a physics-based method for protein-structure prediction from amino acid sequence alone and to understand the mechanism of protein folding,[1−7] in the past few years we have been developing a physics-based united-residue force field, hereafter referred to as UNRES.[8−16] The force field is derived as a restricted free energy (RFE) function, which corresponds to averaging the energy over the degrees of freedom that are neglected in the united-residue model.[8−10,12−14,17−19] An inherent feature of such force fields, resulting from the fact that they correspond to an averaged energy, is the appearance of correlation or multibody terms, the inclusion of which is necessary to describe the system of interest.[10,14,20−22] The early version of the UNRES force field contained only a limited number of multibody terms corresponding to the correlation between hydrogen-bonding interactions of pairs of adjacent peptide groups.[10] Nevertheless, this version was capable of predicting the structures of helical proteins, as assessed during the CASP3 experiment,[23,24] although it was unable to model β structures.[12]

Recently,[14,19] we developed a general theory for deriving united-residue force fields from the all-atom energy surfaces through the factorization of the RFE function and the expansion of the factors in the series of Kubo's generalized cumulants,[25] which enabled us to derive analytical expressions for the multibody terms corresponding to the coupling between backbone−local and backbone−electrostatic interactions. With these terms, the UNRES force field is now able to treat proteins containing both α and β structure, as assessed in the CASP4 experiment.[26]

Until now, the multibody and also some other terms in the UNRES force field have been parametrized by averaging all-atom energy surfaces of model systems[14,18,19] computed with the ECEPP/3[27] force field. In general, empirical all-atom force fields are far from a perfect description of the details of interactions; additionally, the ECEPP/3 force field assumes rigid valence geometry, which can lead to an even worse description of the regions of the conformational space outside energy minima, which are included in the averaging. Therefore, and also encouraged by the fact that the model systems necessary

† Part of the special issue "Jack H. Freed Festschrift".

* Corresponding author. E-mail: has5@cornell.edu. Tel: (607) 255-4034. Fax: (607) 254-4700.

‡ University of Gdańsk.

§ Cornell University.

|| Academic Computer Center in Gdańsk TASK.

10.1021/jp030844f CCC: $27.50
Published on Web 03/26/2004

to derive RFE surfaces are relatively small, we recently decided to reparametrize crucial terms that involve interactions within a protein backbone by using ab initio methods of molecular quantum mechanics. In our most recent study,[16] we computed ab initio energy surfaces of terminally blocked glycine, proline, and alanine and used them to reparametrize torsional and double-torsional potentials. The goal of our present work was to reparametrize backbone−electrostatic terms and the multibody terms corresponding to the coupling between backbone−local and backbone−electrostatic interactions.

2. Methods

2.1. UNRES Force Field. In the UNRES model,[8−16] a polypeptide chain is represented as a sequence of α-carbon (C^α) atoms linked by virtual bonds with attached united side chains (SC) and united peptide groups (p). Each united peptide group is located in the middle of two consecutive α carbons. Only these united peptide groups and the united side chains serve as interaction sites, the α carbons serving only to define the chain geometry (Figure 1). All virtual bond lengths (i.e., $C^\alpha-C^\alpha$ and $C^\alpha-SC$) are fixed; the distance between neighboring C^α atoms is 3.8 Å, corresponding to trans peptide groups, and the side-chain angles (α_{SC} and β_{SC}) and virtual-bond (θ) and dihedral (γ) angles can vary. The energy of the virtual-bond chain is expressed by eq 1.

$$U = \sum_{i<j} U_{SC_iSC_j} + w_{SCp}\sum_{i\neq j} U_{SC_ip_j} + w_{el}\sum_{i<j-1} U_{p_ip_j} + w_{tor}\sum_i U_{tor}(\gamma_i) + w_{tord}\sum_i U_{tord}(\gamma_i, \gamma_{i+1}) + w_b\sum_i U_b(\theta_i) + w_{rot}\sum_i U_{rot}(\alpha_{SC_i}, \beta_{SC_i}) + \sum_{m=3}^{6} w_{loc-el}^{(m)} U_{loc-el}^{(m)} + w_{turn}^{(3)} U_{turn}^{(3)} + w_{turn}^{(4)} U_{turn}^{(4)} + w_{turn}^{(6)} U_{turn}^{(6)} \quad (1)$$

The term $U_{SC_iSC_j}$ represents the mean free energy of the hydrophobic (hydrophilic) interactions between the side chains, which implicitly contains the contributions from the interactions of the side chain with the solvent. The term $U_{SC_ip_j}$ denotes the excluded-volume potential of the side-chain−peptide-group interactions. The peptide-group interaction potential ($U_{p_ip_j}$) accounts mainly for the electrostatic interactions (i.e., the tendency to form backbone hydrogen bonds) between peptide groups p_i and p_j. U_{tor}, U_{tord}, U_b, and U_{rot} represent the energies of virtual-bond dihedral angle torsions, coupled (double) virtual-bond dihedral angle torsions, virtual-bond angle bending, and side-chain rotamers; these terms account for the local propensities of the polypeptide chain. Details of the parametrization of all of these terms are provided in earlier publications.[8,9] Finally, the terms $U_{loc-el}^{(m)}$, with $m = 3, 4, 5, 6$, and $U_{turn}^{(3)}$, $U_{turn}^{(4)}$, and $U_{turn}^{(6)}$ are the multibody contributions from a cumulant expansion[14] of the restricted free energy (RFE), and the w's are the weights of the energy terms. The terms $U_{loc-el}^{(m)}$, with $m = 3, 4, 5, 6$, refer to interactions between noncontiguous parts of the chain, and the terms $U_{turn}^{(m)}$ pertain to correlations within contiguous parts of the chain containing m peptide groups. Such correlations are important when chain reversal occurs in the respective fragment; therefore, they are referred to as the "turn" contributions. The correlation terms are illustrated in Figure 2.[14] The correlation terms of a given order correspond to the factors of the RFE of that order.[14] Therefore, there are no correlation terms of order 1 because they correspond to the free energies of interaction between isolated pairs of sites or within an isolated site.[14] In the case of backbone−local and backbone−electrostatic interactions, the only nonvanishing factors of order 2 correspond to coupling of the local interactions within two consecutive residues, which leads to torsional potentials and parts of virtual-bond-angle-bending (U_b) and rotamer (U_{rot}) potentials.[14] The term $U_{turn}^{(5)}$ is absent because a turn term similar to $U_{turn}^{(6)}$ involving two electrostatic interactions and three local interactions is identically equal to zero within the cumulant approximation.[14] The terms $U_{SC_iSC_j}$, U_b, and U_{rot} were parametrized in our earlier work[8,9] on the basis of the distribution functions calculated from a set of 195 nonhomologous high-resolution structures selected from the Protein Data Bank (PDB),[28] and the parameters of $U_{SC_ip_j}$ were determined to reproduce the local geometry of model α helices and β sheets.[10,18] It should be noted, however, that although the terms $U_{SC_iSC_j}$ (which are one of the major structure-determining factors) are formally knowledge-based, the side-chain hydrophobicities calculated from these terms correlate well with the hydrophobicity scale of Fauchere and Pliška,[29] which is based on the free energy of transfer from water to an *N*-octanol phase.[8] It can therefore be stated that $U_{SC_iSC_j}$ is physics-based at least as far as the relations between the values of the energies of interaction between side chains are concerned. The terms U_{tor} and U_{tord} were parametrized in their final forms in our most recent work[16] by fitting the RFE factors corresponding to local interactions within two or three consecutive peptide units to 1D and 2D Fourier series, respectively. The electrostatic and multibody terms are described in more detail below.

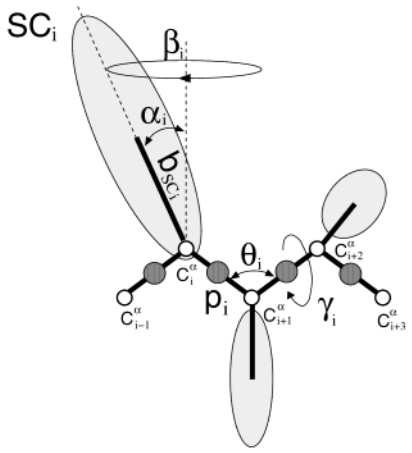


Figure 1. UNRES model of polypeptide chains. The interaction sites are side-chain centroids of different sizes (SC), and the peptide-bond centers (*p*) that are indicated by shaded circles, whereas the α-carbon atoms (small empty circles) are introduced only to assist in defining the geometry. The virtual $C^\alpha-C^\alpha$ bonds have a fixed length of 3.8 Å, corresponding to a trans peptide group; the virtual-bond (θ) and dihedral (γ) angles are variable. Each side chain is attached to the corresponding α carbon with a fixed bond length, b_{SC_i}, a variable bond angle, α_{SC_i}, formed by SC_i and the bisector of the angle defined by C_{i-1}^α, C_i^α, and C_{i+1}^α, and a variable dihedral angle, β_{SC_i}, of counterclockwise rotation about the bisector, starting from the right side of the C_{i-1}^α, C_i^α, C_{i+1}^α frame.

The peptide-group interaction contribution (electrostatic) is expressed by eq 2, which is eq 5 in ref 18. The first part of this equation (except for the Lennard-Jones term) was derived[17,18] by averaging the energy of interacting peptide-group dipoles

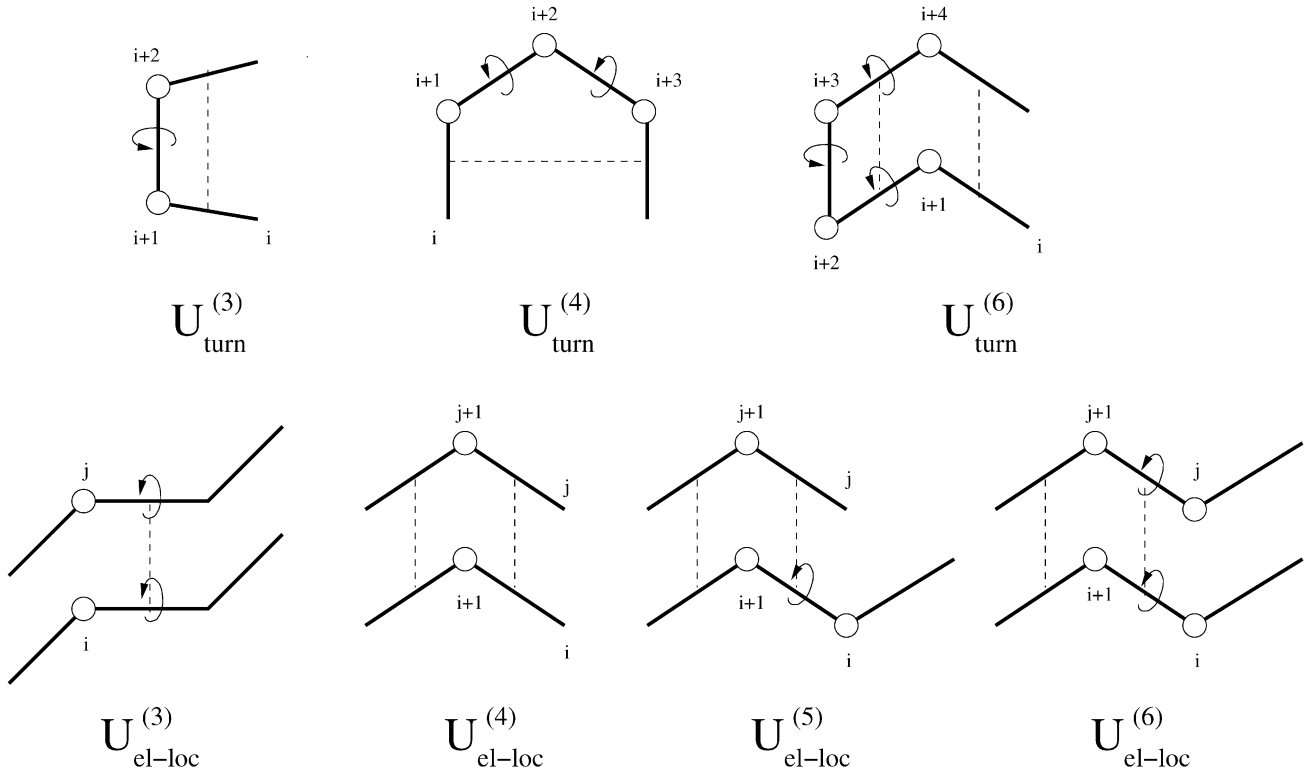


Figure 2. Graphical representation of the analytical formulation of the multibody terms in the cumulant expansion for the restricted free energy of the polypeptide chain. Circles represent local interactions, and broken lines represent the electrostatic interactions that enter into the multibody terms. The term local interaction (as used here) refers to the all-atom interactions of the peptide groups flanking the central C^{α} atom.

over the angles of rotation λ (Figure 3) about the $C^{\alpha}\cdots C^{\alpha}$ virtual-bond axes. In our general theory, this contribution is the second-order (the lowest nonvanishing) cumulant in the cumulant expansion of the RFE of two interacting peptide groups.[10,14] The Lennard-Jones term was added to account for the exchange-repulsion and dispersion interactions between the peptide groups.[18]

$$U_{p_ip_j} = \frac{A_{p_ip_j}}{r_{ij}^{3}}(\cos\alpha_{ij} - 3\cos\beta_{ij}\cos\gamma_{ij}) - \frac{B_{p_ip_j}}{r_{ij}^{6}}[4 + (\cos\alpha_{ij} - 3\cos\beta_{ij}\cos\gamma_{ij})^2 - 3(\cos^2\beta_{ij} + \cos^2\gamma_{ij})] + \epsilon_{p_ip_j}\left[f\left(\frac{r^{\circ}_{p_ip_j}}{r_{ij}}\right)^{12} - 2\left(\frac{r^{\circ}_{p_ip_j}}{r_{ij}}\right)^{6}\right] \quad (2)$$

with

$$\cos\alpha_{ij} = \mathbf{v}_i\cdot\mathbf{v}_j \qquad \cos\beta_{ij} = \mathbf{v}_i\cdot\mathbf{e}_{r_{ij}} \qquad \cos\gamma_{ij} = \mathbf{v}_j\cdot\mathbf{e}_{r_{ij}} \quad (3)$$

where $A_{p_ip_j}$, $B_{p_ip_j}$, and $\varepsilon_{p_ip_j}$ are constants that are characteristic of the kind of interacting peptide groups, r_{ij} is the distance between the peptide-group centers, $\mathbf{v_i}$ is the unit vector pointing from C_i^{α} to C_{i+1}^{α}, $\mathbf{v}_j$ is the unit vector pointing from C_j^{α} to C_{j+1}^{α}, and $\mathbf{e}_{r_{ij}}$ is the unit vector pointing from p_i to p_j. (See Figure 3 for illustration.) The factor $f = 0.5$ is introduced in this work to account for the fact that the r^{-12} term can result in too large a repulsion for the 1,3-adjacent peptide group. Similar scaling is used in all-atom and united-atom force fields.[27,30−32]

The four adjustable parameters of $U_{p_ip_j}$ ($A_{p_ip_j}$, $B_{p_ip_j}$, $\epsilon_{p_ip_j}$, and $r^{\circ}_{p_ip_j}$) depend on the type of peptide group: ordinary (ord) or proline (pro, referring to a peptide group preceding a proline residue). In the UNRES force field developed prior to the improvements described in this paper, we used the values obtained in our earlier work[18] by fitting the analytical expression to the numerically computed average ECEPP/3[27] energy surfaces of three possible pairs composed of ordinary or proline-type peptide groups.

The analytical expressions for the correlation terms are the lowest-order Kubo generalized cumulants[25] corresponding to the coupling between a given number of dipole−dipole and backbone−local interactions; the interactions involved are illustrated in Figure 2. Except for $U^{(4)}_{\mathrm{loc-el}}$, the equations are taken from our recent paper;[14] for $U^{(4)}_{\mathrm{loc-el}}$, we use equations from our earlier paper (eqs 39 and 40 in ref 10). The reason for using the simpler equations is that they have fewer parameters than those derived in our recent work (eqs 49 and 50 in ref 14) and were proven to represent well the correlation of the electrostatic interactions of two neighboring hydrogen-bonded pairs of adjacent peptide groups, which leads to the promotion of α-helical structure.[10] The backbone−local interactions are defined as the energy maps of terminally blocked amino acid residues of the following three basic types: glycine (Gly), L-alanine (Ala), and L-proline (Pro), where Ala represents all amino acid residues except Gly and Pro. These energy surfaces are expanded into a Fourier series in the angles $\lambda^{(1)}$ and $\lambda^{(2)}$ for the rotation of the peptide groups of terminally blocked residues about the $C^{\alpha}\cdots C^{\alpha}$ virtual-bond axes defined in ref 33 and illustrated in Figure 4. The expansion is given by eq 4 (i is the residue number and e_i is the local-interaction energy surface of this residue), which is eq 23 in ref 14.

$$e_i \approx \mathbf{u}^{\mathrm{T}}(\lambda_i^{(1)})\mathbf{b}_{\mathbf{1},i} + \mathbf{u}^{\mathrm{T}}(\lambda_i^{(2)})\mathbf{b}_{\mathbf{2},i} + \mathbf{u}^{\mathrm{T}}(\lambda_i^{(1)})\,\mathbf{C}_i\mathbf{u}(\lambda_i^{(1)}) + \mathbf{u}^{\mathrm{T}}(\lambda_i^{(2)})\mathbf{D}_i\mathbf{u}(\lambda_i^{(2)}) + \mathbf{u}^{\mathrm{T}}(\lambda_i^{(1)})\mathbf{E}_i\mathbf{u}(\lambda_i^{(2)}) \quad (4)$$

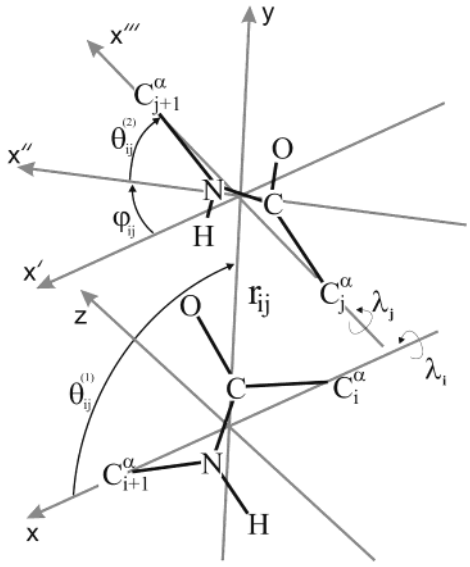


Figure 3. Definition of the variables defining the mutual position, orientation, and rotation about the virtual-bond axes of two interacting peptide groups p_i and p_j. The points C_i^{α} and C_{i+1}^{α} (C_j^{α} and C_{j+1}^{α}) denote the C^{α} atoms of the peptide groups p_i and p_j, respectively. The *x* axis of the coordinate system goes from C_i^{α} to C_{i+1}^{α}, and this group is rotated counterclockwise (toward the *z* axis) about the virtual-bond axis by dihedral angle λ_i from the initial position in which it lies in the *xy* plane with the carbonyl oxygen atoms having a positive *y* coordinate. To define the position, orientation, and rotation of peptide group p_j, we place it initially in the initial position of peptide group p_i and rotate by λ_j about the virtual-bond axis. Then this peptide group is moved in the *xy* plane to reach the distance r_{ij} between the centers of p_i and p_j and the angle $\theta_{ij}^{(1)}$ between the *x* axis and the vector r_{ij} connecting the centers of p_i and p_j. After this operation, the *x* axis of the coordinate system becomes the *x*′ axis. Next, the virtual-bond axis of p_j is rotated by the angle ϕ_{ij} about the *y* axis (after this operation, the *x* axis becomes the *x*″ axis) and finally by the angle $\theta_{ij}^{(2)}$ about the *z* axis (after this operation, the *x* axis becomes the *x*‴ axis).

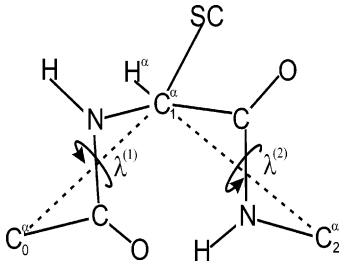


Figure 4. Definition of the dihedral angles $\lambda^{(1)}$ and $\lambda^{(2)}$ for the rotation of the peptide groups about the $C^{\alpha}-C^{\alpha}$ virtual bonds (dashed lines) of a peptide unit.

with

$$\mathbf{u}(\lambda) = \begin{pmatrix} \cos\lambda \\ \sin\lambda \end{pmatrix} \tag{5}$$

where the superscript T denotes the transpose of a matrix or a vector. The elements of vectors $\mathbf{b_1}$ and $\mathbf{b_2}$ and of matrices **C**, **D**, and **E** are constants of the Fourier expansion. There are 4 symmetry relations between the coefficients (eqs 6−9), which makes 12 independent coefficients for Ala and Pro; for Gly, because of the inversion symmetry of the energy map, all coefficients that are multiplied by a single sine term are zero. This means that both elements of vector $\mathbf{b_2}$ (b_{21} and b_{22}) as well as the off-diagonal elements of matrices **C**, **D**, and **E** are zero, which reduces the number of parameters for this type of residue to six.

$$c_{i;22} = -c_{i;11} \tag{6}$$

$$c_{i;21} = c_{i;12} \tag{7}$$

$$d_{i;22} = -d_{i;11} \tag{8}$$

$$d_{i;21} = d_{i;12} \tag{9}$$

As an example, we recall here the expressions for $U_{\mathrm{loc-el}}^{(3)}$ and $U_{\mathrm{turn}}^{(3)}$ (eqs 10 and 15, respectively, which are eqs 43 and 46 in ref 14). For expressions for other correlation terms as well as for the derivation of these expressions, the reader is referred to our earlier paper.[14]

$$\begin{aligned} U_{\mathrm{loc-el}}^{(3)} &= \sum_{i<j+2} u_{ij}^{(3)} = \sum_{i<j+2} \langle e_i E_{ij} e_j \rangle_{\mathrm{c}} \\ &\quad + \langle e_i E_{ij} e_{j+1} \rangle_{\mathrm{c}} + \langle e_{i+1} E_{ij} e_j \rangle_{\mathrm{c}} + \langle e_{i+1} E_{ij} e_{j+1} \rangle_{\mathrm{c}} \\ &= \frac{1}{4} \sum_{i<j+2} \boldsymbol{\mu}_i^{\mathrm{T}} \mathbf{A}'_{ij} \boldsymbol{\mu}_j \end{aligned} \tag{10}$$

where E_{ij} denotes the interaction energy between the dipoles of peptide groups p_i and p_j, $\langle \cdots \rangle_{\mathrm{c}}$ denotes the cumulant average of the quantity inside the brackets, and the dipole moments $\boldsymbol{\mu}_k$ ($k = 1, 2$) are

$$\boldsymbol{\mu}_k = \mathbf{U}(\gamma_k)\mathbf{b}_{\mathbf{2},k} + \mathbf{b}_{\mathbf{1},k+1} \tag{11}$$

$$\mathbf{U}(\gamma) = \begin{pmatrix} -\cos\gamma & -\sin\gamma \\ -\sin\gamma & \cos\gamma \end{pmatrix} = \begin{pmatrix} \cos\gamma & -\sin\gamma \\ \sin\gamma & \cos\gamma \end{pmatrix} \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} = \mathbf{R}(\gamma)\,\mathbf{S}(xz) \tag{12}$$

where $\mathbf{R}(\gamma)$ is the matrix for rotation about the $C^{\alpha}\cdots C^{\alpha}$ virtual-bond axis by angle γ and $\mathbf{S}(xz)$ is the matrix for reflection with respect to the plane perpendicular to that defined by three consecutive C^{α} atoms (referred to as the *xz* plane),

$$\mathbf{A}_{ij} = \frac{p_i p_j}{\epsilon r_{ij}^3} (\mathbf{I} - 3\mathbf{e}_{r_{ij}} \mathbf{e}_{r_{ij}}^{\mathrm{T}}) \mathbf{T}_{ij} \tag{13}$$

$$\mathbf{A}' = \begin{pmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{pmatrix} \tag{14}$$

where p_i and p_j are the magnitudes of the dipole moments of the *i*th and the *j*th peptide groups, respectively, ϵ is the dielectric constant, r_{ij} is the distance between peptide groups *i* and *j*, $\mathbf{e}_{r_{ij}}$ is the unit vector pointing from peptide group *i* to peptide group *j*, **I** is the 3×3 unit matrix, and $\mathbf{T}_{ij}$ is the transformation matrix from the reference system of peptide group *i* to that of peptide group *j*.

$$\begin{aligned} U_{\mathrm{turn}}^{(3)} &= \sum_i v_i^{(3)} = \sum_i \langle (e_i + e_{i+1}) E_{i,i+2} (e_{i+2} + e_{i+3}) \rangle_{\mathrm{c}} \\ &= \frac{1}{4} \sum_i \boldsymbol{\mu}_{\mathbf{i}}^{\mathrm{T}} \mathbf{A}_{i,i+2}{}' \mu_{i+2} + \\ &\quad \frac{1}{8} \sum_i \mathrm{tr}[\mathbf{U}(\gamma_{i+1}) \mathbf{E}_{i+1}^{\mathrm{T}} \mathbf{A}_{i,i+2}{}' \mathbf{U}(\gamma_{i+2}) \mathbf{E}_{i+2}^{\mathrm{T}}] \end{aligned} \tag{15}$$

It can be seen that the components of both $U_{\mathrm{loc-el}}^{(3)}$ and $U_{\mathrm{turn}}^{(3)}$ (and, likewise, those of $U_{\mathrm{loc-el}}^{(5)}$, $U_{\mathrm{loc-el}}^{(6)}$, $U_{\mathrm{turn}}^{(4)}$, and $U_{\mathrm{turn}}^{(6)}$) not only depend on the distance and orientation of the two interacting peptide groups but also depend explicitly on the

adjacent virtual-bond dihedral angles γ.[14] The latter feature results from coupling between backbone−electrostatic and backbone−local interactions in the RFE and is responsible for the stabilization of parallel and antiparallel β sheets.[14] Technically, the correlation terms can be regarded as additional torsional potentials that are dependent on the distance and orientation of the peptide groups involved.[14]

The constants p_i, p_j, and ϵ (for part of the backbone−electrostatic interactions) can be absorbed into the energy-term weights and the coefficients of the second-order Fourier expansion for e (eq 4). Therefore, the adjustable parameters in the expressions for the correlation terms for the backbone−local interactions and energy-term weights are the coefficients of the Fourier expansion, which makes a total of 30 parameters for all three types of amino acid residues (Gly, Ala, and Pro).

2.2. Parametrization of UNRES Energy Terms by Fitting the RFEs of Model Systems. The procedures for the parametrization of $U_{p_ip_j}$ and U_{corr} were developed in their basic forms in our earlier work.[14,16,18] The analytical expressions for these potentials are parametrized by fitting the respective analytical expressions to the RFEs or relevant parts (factors) of the RFEs of appropriate model systems, which were calculated numerically from the respective all-atom energy surfaces. In order not to distract the reader, we outline only the applied procedures in this section, referring the reader to more details given in the Appendix (section 5).

For $U_{p_ip_j}$, the all-atom energy surfaces of pairs of interacting peptide groups were determined in this work by calculating the respective ab initio energy surfaces on a coarse grid of distances and orientations of the interacting peptide groups and subsequently fitting a simplified energy function to the ab initio energies; this is described in the Appendix (section 5.3). For the correlation contributions, the local-interaction energy surfaces of terminally blocked Gly, Ala, and Pro residues were also required; these surfaces were calculated in our recent work.[16]

Fitting of the RFEs was accomplished in all cases by minimizing the target function defined by eq 16.[14,16,18]

$$\Phi(\mathbf{X}) = \sum_{s=1}^{S} \sum_{k=1}^{N_s} w_k^{(s)} [F_k^{(s)} - U^{(s)}(\mathbf{X}; \mathbf{R}_k^{(s)})]^2 \qquad (16)$$

where $\mathbf{X}$ is the vector of adjustable parameters, S is the number of model systems considered simultaneously in the fitting, N_s is the number of points at which the RFE of system s has been evaluated numerically, $F_k^{(s)}$ is the RFE of system s at the kth point, $U^{(s)}(\mathbf{X}; \mathbf{R}_k^{(s)})$ is the sum of the relevant UNRES terms at the kth point of the RFE of system s given the vector of parameters $\mathbf{X}$, $\mathbf{R}_k^{(s)}$ is the vector of variables defining the kth point of system s, and $w_k^{(s)}$ is the weight of the kth point of the RFE surface of system s. The assignment of weights was system-specific and is described in the Appendix (sections 5.1 and 5.2, respectively) for each individual system.

For practical reasons, we determined the parameters of the peptide-group interaction term, $U_{p_ip_j}$, first; then, taking the parameters of $U_{p_ip_j}$ as known, we evaluated the 30 Fourier coefficients for the backbone−local interactions by fitting them to reproduce the RFE surfaces of model terminally blocked tetrapeptides. (See section 5.2.)

To parametrize $U_{p_ip_j}$, we evaluated the RFEs of pairs composed of *N*-acetyl-*N*′-methylacetamide (AcNHMe, which represents ordinary peptide groups) and *N*-acetyl-*N*′,*N*′-dimethylacetamide ($AcNMe_2$, which represents the proline peptide group). Three pairs were considered, namely, AcNHMe−AcNHMe, AcNHMe−$AcNMe_2$, and $AcNMe_2$−$AcNMe_2$. Each system was considered separately, and for each of them, the set of 4 parameters ($A_{p_ip_j}$, $B_{p_ip_j}$, $\epsilon_{p_ip_j}$, and $r^{\circ}_{p_ip_j}$) was determined (a total of 12 parameters for all systems).

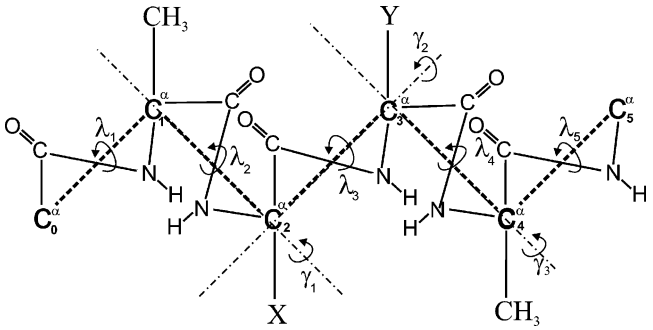


Figure 5. Definition of the variables of the model tetrapeptides, the RFE surfaces of which were evaluated to parametrize the multibody terms in the UNRES force field.

For each of the three systems, the variables for the fitting procedure were the distance between peptide-group centers r and the angles $\theta^{(1)}$, $\theta^{(2)}$, and ϕ defining the orientation of two peptide groups illustrated in Figure 3.

To determine the coefficients of the cumulant-based expressions for the correlation terms, we calculated the parts of the RFEs corresponding to the interactions within the backbones of the following model terminally blocked tetrapeptides: Ac-Ala-Ala-Ala-Ala-NHMe (to determine the coefficients of Ala), Ac-Ala-Gly-Gly-Ala-NHMe, Ac-Ala-Gly-Ala-Ala-NHMe, and Ac-Ala-Ala-Gly-Ala-NHMe (to determine the coefficients of Gly), and Ac-Ala-Ala-Pro-Ala-NHMe, Ac-Ala-Pro-Ala-Ala-NHMe, and Ac-Ala-Pro-Gly-Ala-NHMe (to determine the coefficients of Pro). The reason for choosing tetrapeptides was that they are the smallest fragments of polypeptide chains such that their UNRES energy contains all correlation components. Moreover, they can be folded into a minimum α-helical or antiparallel β sheet structure. The reason for choosing three tetrapeptides with different composition and placement of non-Ala residues to parametrize Gly and Pro was to cover frequent situations in which a hairpin has a turn at Gly-Gly, Gly-Ala, Ala-Gly, Pro-Ala, and Pro-Gly types of residues. A tetrapeptide forms the minimum β hairpin if the turn occurs in the middle of the chain. In addition to this, the tetrapeptide Ac-Ala-Ala-Pro-Ala-NHMe was added to provide information that forming β turns with Pro as the third residue of the turn is unfavorable. The RFEs were calculated as functions of the three virtual-bond dihedral angles γ_1, γ_2, and γ_3 (Figure 5) to obtain both the local and electrostatic correlation terms.

2.3. Determination of Energy-Term Weights and Refinement of the Parameters of the Correlation Terms. To determine the energy-term weights and to refine other energy parameters of eq 1 including the Fourier coefficients in the correlation terms, we used our hierarchical energy-landscape optimization procedure.[15] This method assumes a hierarchical structure of the energy landscape, which means that the energy decreases as the number of nativelike elements in a structure increases, being lowest for structures from the native family and highest for structures with no nativelike element. A level of the hierarchy is defined as a family of structures with the same number of nativelike elements (or the degree of nativelikeness). The optimization of a potential-energy function is aimed at achieving such a hierarchical structure of the energy landscape by forcing appropriate free-energy gaps between hierarchy levels to place their energies in ascending order. This is achieved by minimizing the following target function:

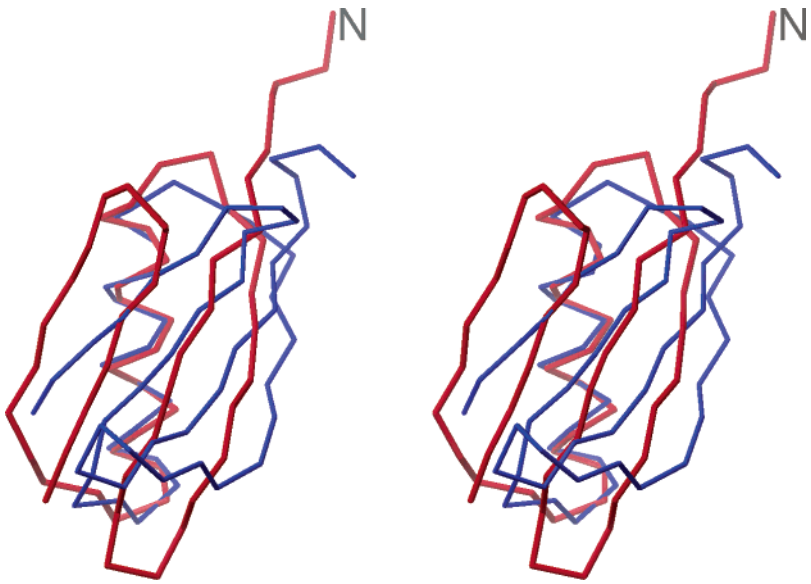


Figure 6. Stereoview of a superposition of the C^{α} trace of the crystal structure of 1IGD (red)[36] and the lowest-energy structure obtained with the UNRES force field parametrized in this work (blue). The rmsd is 5.3 Å. The N terminus is indicated.

TABLE 1: Values of the Constants in Equation 2 for the Peptide Group−Peptide Group Interaction Energy Obtained in This Work Compared with Those Obtained from ECEPP/3 Energy Surfaces[18]

	this work					ref 18			
pair	ε_{ij}°	r_{ij}°	A_{ij}	B_{ij}	σ_E [a]	ε_{ij}°	r_{ij}°	A_{ij}	B_{ij}
ordinary−ordinary	1.085	5.27	−3.29	−8797	0.47	0.305	4.51	3.73	−1306
ordinary−proline	0.554	5.46	0.00	−8167	0.42	0.365	4.54	0.00	−1129
proline−proline	0.962	5.23	1.02	−427	0.15	0.574	4.48	5.13	−335

[a] Standard deviation of the RFE.

$$\Phi = \sum_{i=0}^{n} w_i g(F_i - F_{i+1}; -\infty, -\Delta_i) + \sum_{I=1}^{n} \sum_{J=1}^{I-1} w_{IJ} g(\tilde{F}_I - \tilde{F}_J; -\Delta_{IJ}, \Delta_{IJ}) \quad (17)$$

with

$$g(x; x_{\min}, x_{\max}) = \begin{cases} \frac{1}{4}(x - x_{\min})^4 & x < x_{\min} \\ \frac{1}{4}(x - x_{\max})^4 & x > x_{\max} \\ 0 & \text{otherwise} \end{cases} \quad (18)$$

and the w's being the weights of the respective terms in the target function. (It should be noted that they are different from the energy-term weights of eq 1). F_i is the configurational free energy of structural level i, F_I is the configurational free energy of the set of structures with certain elements of secondary structure defined by eqs 19 and 20, respectively, Δ_i is the target free-energy gap between the ith and the $i + 1$ level of the hierarchy, and Δ_{IJ} is the maximum allowed free-energy difference between structures of type I and J, respectively, which contain the same number of secondary-structure elements.

$$F_i = -\frac{1}{\beta} \ln \sum_{k \in \{i\}} \exp(-\beta E_k) \quad (19)$$

$$\tilde{F}_I = -\frac{1}{\beta} \ln \sum_{k \in \{I\}} \exp(-\beta E_k) \quad (20)$$

with $\{i\}$ and $\{I\}$ denoting the set of conformations of the ith structural level and those containing nativelike fragment I, respectively. β can be identified with $1/RT$, T being the absolute temperature, or treated as a parameter of the method.

This hierarchical method is different from methods developed thus far, in which the energy gap and/or the Z score between the native structure and all non-native structures are maximized, regardless of the degree of nativelikeness of the non-native structures.[34,35]

In this work, we chose a 63-residue $\alpha + \beta$ protein, PDB designation 1IGD,[36] to optimize the force field. This protein consists of an N-terminal β hairpin, an α helix in the middle, and a C-terminal β hairpin, the hairpins being packed antiparallel to each other and also packed to the α helix (Figure 6). These segments were chosen as native ones, and the goal of the optimization was that all of them should be present in the lowest-energy structure. The packing of the segments or the root-mean-square deviation (rmsd) from the native structure was not taken into account in the hierarchical procedure.

As in our earlier work,[35] we imposed several restraints to maintain proper secondary structure, viz., (i) the energy difference between the left- and right-handed α helices was required to be at least 0.5 kcal/mol per residue, (ii) the energy difference between a 3_{10} helix and a right-handed α helix was required to be at least 0.5 kcal/mol per residue, and (iii) proper values of virtual-bond angles θ and virtual-bond dihedral angles γ were required to be present in right-handed α helices and in β sheets.

3. Results and Discussion

3.1. Parametrization of the Peptide-Group Interaction Energy $U_{p_ip_j}$. The parameters of $U_{p_ip_j}$ (eq 2), obtained by fitting the RFE surfaces calculated with the simplified energy function (eq A-7), are summarized in Table 1. The curves from the fitted

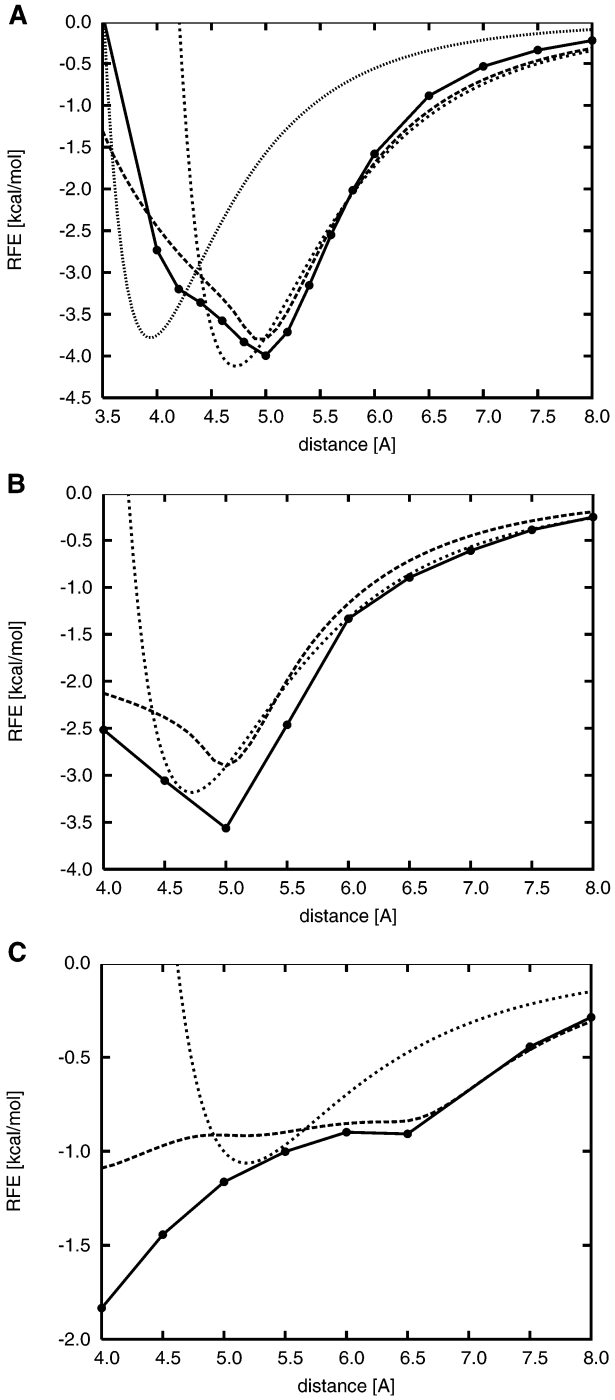


Figure 7. Comparison of the dependence of the RFE of (A) AcNHMe−AcNHMe, (B) AcNHMe−AcNMe$_2$, and (C) AcNMe$_2$−AcNMe$_2$ pairs on the distance (r) between peptide-group centers at parallel orientation ($\theta^{(1)} = 90°$, $\phi = \theta^{(2)} = 0°$) calculated directly from the ab initio energy surfaces (solid lines and filled circles), the energy surfaces obtained with a simplified all-atom energy function (eq A-7) (long-dashed lines), and the fitted $U_{p_ip_j}$ UNRES term (eq 2) (short-dashed lines). For the AcNHMe-AcNHMe pair, the distance curve of the $U_{p_ip_j}$ term parametrized using the ECEPP/3 energy surface[18] is also shown (dotted line); these values (dotted lines) are multiplied by a factor of 2.5 for better visualization. For the AcNHMe−AcNHMe pair, $U_{p_ip_j}$ was fitted beginning from $r = 4.0$ Å, and for the AcNHMe−AcNMe$_2$ and AcNMe$_2$−AcNMe$_2$ pairs, $U_{p_ip_j}$ was fitted beginning from $r = 4.6$ Å (see text). All orientations of the interacting molecules were considered in the fitting (see text for details).

$U_{p_ip_j}$ surface, calculated as functions of the distance between the centers of the peptide groups at $\theta^{(1)} = 90°$, $\phi = \theta^{(2)} = 0°$ (short-dashed lines), are compared with the corresponding ab initio RFE curves (solid lines) in Figure 7A−C. It should be noted that all orientations, and not only the special orientation of the peptide groups for which the plots in Figure 7A−C are shown, were used in fitting. It can be seen for the AcNHMe−AcNHMe and the AcNHMe−AcNMe$_2$ pairs that the distance curves from the fitted expression for $U_{p_ip_j}$ approximate the RFE curves well for distances from about 5 Å on. For the AcNMe$_2$−AcNMe$_2$ system, the fit is significantly worse but still reasonable at distances greater than 5.0 Å. For shorter distances, the curves corresponding to eq 2 deviate from the corresponding RFE curves, which means that a more complex distance dependence than a combination of the r^{-6} and r^{-12} terms would be required to fit the RFE in the short-distance region. However, we did not consider it worthwhile because, as mentioned earlier in this section, the distances between hydrogen-bonded peptide groups in regular protein structures (α helices and β sheets) are close to 4.8 Å on average and rarely decrease below 4.5 Å (Figure 8). For comparison, the $U_{p_ip_j}$ curve parametrized in our earlier work[18] based on the ECEPP/3 force field is also shown in Figure 7A (dotted line). It can be seen that the curve has a minimum at 4.0 Å, which is shorter by 1 Å than the minimum of the ab initio-based RFE curve. In agreement with this observation, the values of $r°$ in eq 2 determined in this work are significantly larger compared to those determined in our earlier paper from the ECEPP/3-based RFE surfaces[18] (Table 1). This is the reason that the α helices obtained with the preliminary version of the UNRES force field[18] were significantly collapsed with the virtual-bond dihedral angles γ close to 35° instead of about 45° as found in protein structures. In later versions of the UNRES force field, this undesirable feature was removed by increasing $U_{SC_ip_j}$.[10] It can also be noted that the absolute values of ϵ and B are larger than the ECEPP/3-based values (Table 1). In our earlier work,[16] we demonstrated that the ECEPP/3 local-interaction energy surfaces are also different from ab initio energy surfaces; the latter are also different from the surfces obtained from the AMBER[31,32,37] and CHARMM[38] force fields. In particular, the latter locate the C_7^{ax} conformation of L-amino acid residues too low in energy.[39,40] Conversely, the ECEPP/3 energy surfaces overestimate the energy differences between conformations, which results in exaggerated asymmetry of the torsional-energy curves.[16] Therefore, our present use of ab initio

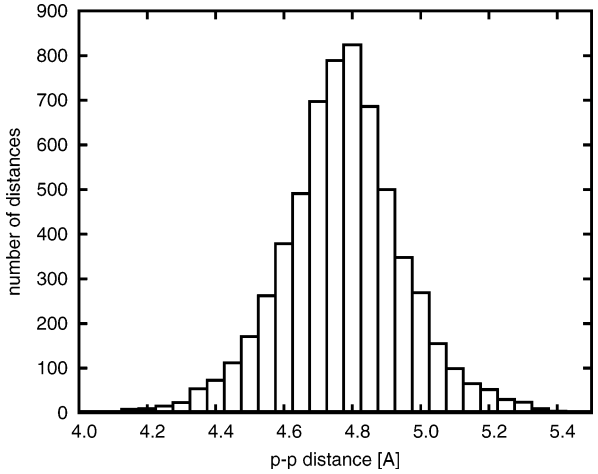


Figure 8. Distribution of distances between hydrogen-bonded peptide group centers calculated from the set of 195 high-resolution non-homologous protein structures selected in ref 8. The criteria for hydrogen-bond formation were (i) the N···O distance is less than 2.5 Å and (ii) the N−H···O angle is greater than 90°. The amide hydrogen coordinates were determined on the basis of standard valence geometry. The bin size is 0.05 Å.

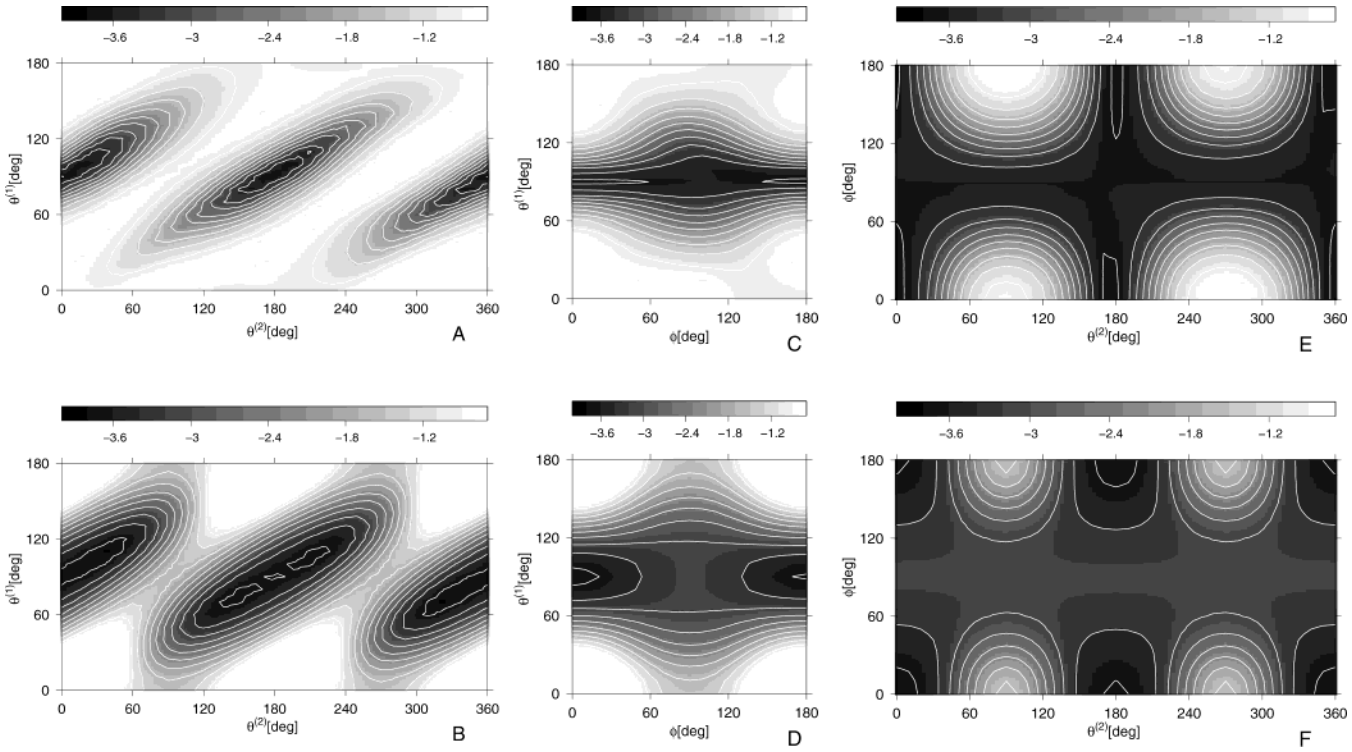


Figure 9. Comparison of the dependence of the simplified ab initio RFE (eq A-7) (in kcal/mol) of AcNHMe−AcNHMe at the 5-Å distance between peptide-group centers (A, C, and E) with that of the fitted $U_{p_ip_j}$ (eq 2) parametrized by fitting to the RFE (B, D, and F). (A and B): Section in $\theta^{(2)}$ and $\theta^{(1)}$ with $\phi = 0°$; (C and D): section in ϕ and $\theta^{(1)}$ with $\theta^{(2)} = 0°$; (E and F): section in $\theta^{(2)}$ and ϕ with $\theta^{(1)} = 90°$.

energy surfaces in parametrization is a step forward toward the derivation of a physics-based united-residue force field.

The angular dependence of the simplified ab initio RFE (eq A-7) is compared with that of the fitted $U_{p_ip_j}$ (eq 2) for the AcNHMe−AcNHMe system in Figure 9 calculated in the $\theta^{(2)} - \theta^{(1)}$, $\phi - \theta^{(1)}$, and $\theta^{(2)} - \phi$ planes for the distance between peptide-group centers of 5 Å. The minima shown in Figure 9 correspond to the parallel or antiparallel orientation of the peptide groups with centers positioned over one another ($\theta^{(1)} = 90°$, $\theta^{(2)} = 0$ or 180°, $\phi = 0$ or 180°), which enables hydrogen bonds to form between the peptide groups. It can be seen that the angular dependence of the fitted $U_{p_ip_j}$ (eq 2) with parameters obtained by fitting the simplified RFE surface (eq A-7) agrees well with the angular dependence of the simplified RFE. This result is very important because it demonstrates the validity of the dipole model of a peptide group, which was the basis of both the analytical derivation of the simplified potential defined by eq 2[17,18] and the derivation of the analytical equations for the correlation terms involving the backbone−electrostatic interactions,[10,14,19] at distances corresponding to hydrogen-bonded peptide groups. However, the fact that the curves in Figure 7 that show the distance dependence of $U_{p_ip_j}$ deviate from the corresponding RFE curves at short distances indicates that the dipole approximation is probably not good at shorter distances (below 4.5 Å). However, the peptide groups are not very close in protein structure; therefore, the dipole approximation seems to be a good basis for the derivation of the parts of the UNRES potential that involve the interactions between the backbone peptide groups.

3.2. Parametrization of the Correlation Terms and Determination of Energy-Term Weights. Irreducible coefficients of the second-order Fourier expansion of the backbone−local interaction energy (eq 4), which are parameters of the correlation terms, for the three basic types of amino acid residues obtained by fitting the ab initio-based RFE surfaces of model tetrapeptides as well as their final values obtained by hierarchical optimization[15] using the 1IGD protein are summarized in Table 2. The details of the hierarchical optimization procedure and the discussion of the hierarchical optimization of the system studied will be reported in our forthcoming paper;[41] here we present only the resulting parameters of the energy function. For comparison, values obtained in our earlier work by fitting the ECEPP-based RFE surfaces[14] are also given in this Table. The values of the energy-term weights determined by hierarchical optimization are given in Table 3. The values of the weights indicate that the most important correlation components are $U^{(3)}_{\mathrm{loc-el}}$, $U^{(4)}_{\mathrm{loc-el}}$, and $U^{(3)}_{\mathrm{turn}}$, but the higher-order correlation terms seem to be less significant. It can also be seen that $w^{(4)}_{\mathrm{loc-el}}$ is about 5 times greater than w_{el}, which means that $U^{(4)}_{\mathrm{loc-el}}$ dominates $U_{p_ip_j}$.

With the final refined parameters, the lowest-energy structure of 1IGD obtained by a global conformational search with the conformational space annealing (CSA) search method[42] had a 5.3-Å rmsd from the native structure (Figure 6). It differs from the native structure by a flip of the C-terminal hairpin. This feature is not unexpected because proper packing of secondary-structure elements was not included in the hierarchical optimization. The proper packing of secondary-structure elements of the 1IGD protein was achieved in our later work after the hierarchical optimization procedure had been developed further.[41,43]

As an illustration of the optimization procedure, the correlation contribution to the ab initio RFE profile of the Ac-$(Ala)_4$-NHMe peptide calculated along the line defined by $\gamma_1 = \gamma_2 = \gamma_3$ obtained by numerical integration is compared in Figure 10 with those resulting from fitting the UNRES energy expression to the RFE surface and from the final refinement of the parameters of the UNRES correlation-energy contributions by hierarchical optimization using the 1IGD protein. The reaction coordinate of choice corresponds to squeezing an extended

TABLE 2: Fourier Coefficients (kcal/mol) of Approximate RFE Expressions Pertaining to the Correlation between Local and Electrostatic Interactions up to Sixth Order

coefficient (kcal/mol)	this work (fitted)[a]			this work (final)[b]			ECEPP/3[14]		
	Gly	Ala	Pro	Gly	Ala	Pro	Gly	Ala	Pro
b_{11}	0.340274	0.051794	0.031809[c]	0.164513	−0.499500	0.031809[c]	−0.899672	−0.068616	−0.194383
b_{12}	0.000000	0.799109	1.015390	0.000000	1.044194	1.015390	0.000000	1.160200	0.780293
b_{21}	−0.219031	−0.197181	−1.286480	0.887388	0.224897	−1.286480	0.212255	0.230066	−0.575677
b_{22}	0.000000	−0.548889	−0.906628	0.000000	−0.469470	−0.906628	0.000000	−0.563423	−0.067433
c_{11}	−1.142510	2.854970	1.914370	−1.175519	2.606108	1.914370	−0.772650	1.419590	−1.170890
c_{12}	0.000000	−1.072890	−0.454839	0.000000	−1.348282	−0.454839	0.000000	0.247397	−2.331910
d_{11}	2.384190	−2.166280	−1.548870	2.351567	−2.169758	−1.548870	0.027307	−0.128765	0.586987
d_{12}	0.000000	−0.367110	0.664143	0.000000	−0.021834	0.664143	0.000000	0.221955	−1.860730
e_{11}	0.657192	0.530804	0.051888	1.700467	1.136668	0.051888	1.118324	−1.738019	−0.303282
e_{12}	0.000000	0.003727	0.361644	0.000000	−0.340904	0.361644	0.000000	0.059297	−0.008513
e_{21}	0.000000	0.029784	0.271090	0.000000	−0.234006	0.271090	0.000000	−0.305605	−0.805879
e_{22}	−0.696582	−0.525930	0.154470	−1.055494	−1.134231	0.154470	0.022810	0.389101	−0.153466

[a] Obtained by fitting to ab initio-based RFE surfaces of model tetrapeptides. [b] Final values obtained by hierarchical optimization using the 1IGD protein. [c] Coefficients corresponding to proline were not included in the parameter set for the hierarchical optimization of the energy landscape of 1IGD because this protein contains only one proline residue in the N-terminal part; therefore, the presence of this residue does not determine the fold of 1IGD.

TABLE 3: Energy-Term Weights Obtained by Hierarchical Optimization Using the 1IGD Protein

weight	value (dimensionless)
w_{SCp}	1.54864
w_{el}	0.20016
w_{tor}	1.70537
w_{tord}	1.24442
w_{b}	1.00572
w_{rot}	0.06764
$w^{(3)}_{loc-el}$	1.51083
$w^{(4)}_{loc-el}$	0.91583
$w^{(5)}_{loc-el}$	0.00607
$w^{(6)}_{loc-el}$	0.02316
$w^{(3)}_{turn}$	2.00764
$w^{(4)}_{turn}$	0.05345
$w^{(6)}_{turn}$	0.05282

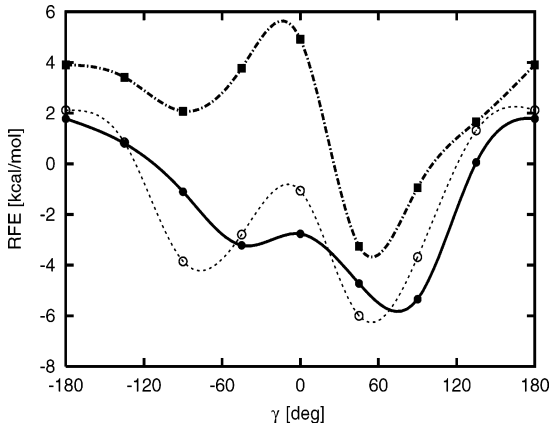


Figure 10. Comparison of the plot of the correlation contribution to the ab initio-based RFE of the Ac-$(Ala)_4$-NHMe model tetrapeptide (solid line and filled circles) calculated as a function of the virtual-bond dihedral angles γ_1, γ_2, and γ_3 (defined in Figure 5) along the line defined by $\gamma_1 = \gamma_2 = \gamma_3$ (which corresponds to uniform compression of the chain) with the curve resulting from fitting the sum of the UNRES correlation terms to the RFE (dashed line and empty circles) and the sum of the UNRES correlation terms resulting from the optimization of energy-term weights and the final refinement of the parameters of the correlation terms by hierarchical optimization using the 1IGD protein (dot−dashed line and filled squares).

polypeptide chain into a right- or left-handed α helix. As can be seen from Figure 10, the correlation contribution to the ab initio RFE has its global minimum close to the right-handed helix ($\gamma = 75°$) and the second minimum with a higher RFE value close to the left-handed helix ($\gamma = -45°$). It should be noted that the ab initio-based UNRES torsional potential corresponding to the Ala-Ala type virtual-bond dihedral angles derived in our earlier work[16] is, in contrast to the corresponding ECEPP/3-based torsional potential,[14] nearly symmetric and therefore does not differentiate the left- from the right-handed helix. The double-torsional potentials[16] also are unable to differentiate between these two structures. It can therefore be concluded that, in the case of the ab initio-based RFE, the correlation terms not only act as β-sheet formers but also determine the proper chirality of α helices.

As can be seen from Figure 10, the correlation component of the ab initio RFE surface for the alanine-type tetrapeptide is fitted reasonably well with the sum of the UNRES correlation terms; in particular, the energy relation between the left- and the right-handed helices is preserved. After hierarchical optimization, the energy difference between the right- and the left-handed helices increased to favor the first structure. However, the latter fact is a consequence of the hierarchical optimization procedure used here. (See section 2.3.)

Another illustration of the importance of the correlation terms resulting from the present work is provided by Figure 11. In this Figure, the correlation contribution to the RFE of two interacting peptide groups with adjacent peptide units at a configuration that is the same as that in "ideally flat" parallel β sheets (Figure 11D) is compared with $U^{(3)}_{loc-el}$ of that system parametrized by fitting to the RFE and, subsequently, by hierarchical optimization using the 1IGD protein. The surfaces are drawn as functions of the virtual-bond dihedral angles γ centered at the virtual-bond axes belonging to the peptide groups. It can be seen that the RFE possesses a broad region of minima near $\gamma_1 = \pm 180°$, $\gamma_2 = \pm 180°$ (Figure 11A). There are four minima in these regions with ($\gamma_1 = \gamma_2 = 150°$), ($\gamma_1 = \gamma_2 = -120°$), ($\gamma_1 = 120°$, $\gamma_2 = -120°$), and ($\gamma_1 = -120°$, $\gamma_2 = 120°$), respectively. There is also a shallow minimum with $\gamma_1 = \gamma_2 = 60°$, which corresponds to the right-handed helix, and a maximum at $\gamma_1 = \gamma_2 = -45°$, which corresponds to the left-handed helix. Thus, the third-order correlation term acts as a former of β sheets (extended structure) and favors the right-handed helices over the left-handed helices. The surface of $U^{(3)}_{loc-el}$ itself (Figure 11B) looks like a smoothed RFE surface: the multiple minima in the extended-structure region are converted to a single minimum close to $\gamma_1 = \gamma_2 = 180°$, and

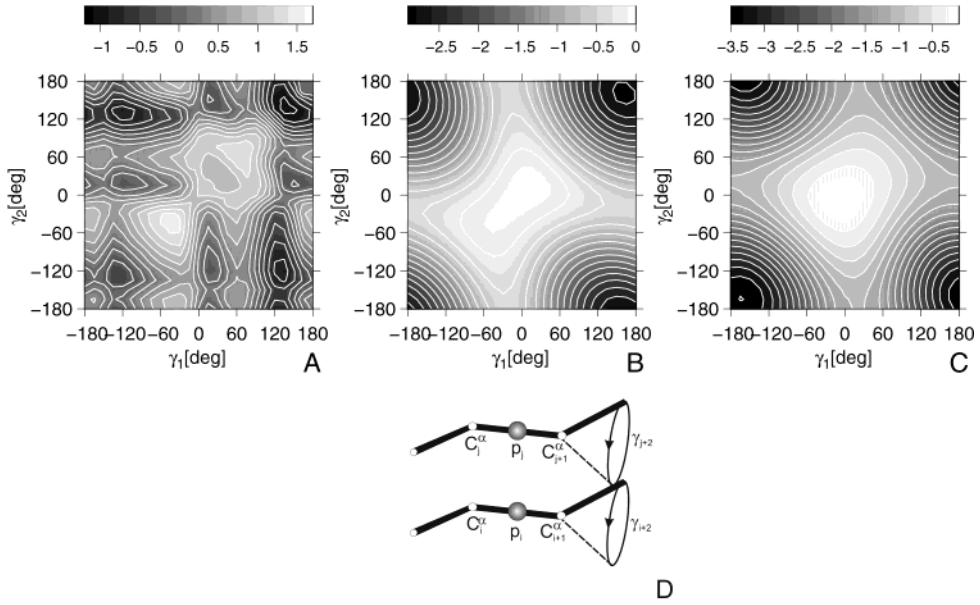


Figure 11. Comparison of the correlation contribution to the ab initio-based RFE (in kcal/mol) of two interacting ordinary-type peptide groups p_i and p_j with local interactions within the adjacent Ala-type peptide units ($F^{(3)}_{\mathrm{loc-el}}$) taken into account as functions of virtual-bond dihedral angles γ_i and γ_j (A), with $U^{(3)}_{\mathrm{loc-el}}$ itself with parameters obtained by fitting the UNRES energy to the ab initio-based RFE (B), and with $U^{(3)}_{\mathrm{loc-el}}$ itself with parameters finally refined by hierarchical optimization using the 1IGD protein (C). (D) Illustration of the system considered.

the shallow minimum corresponding to the right-handed helix disappears. Qualitatively similar conclusions result from the analysis of the ECEPP/3-based RFEs and $U^{(3)}_{\mathrm{loc-el}}$ calculated in our earlier work.[14]

An analysis of the Fourier coefficients collected in Table 2 leads to the observation that, though many of them vary depending on the method of optimization and the all-atom energy surface used to calculate the RFEs (ab initio or ECEPP/3), the coefficients with the largest absolute values are often quite well conserved. In particular, this applies to the coefficients corresponding to the most frequently occurring residue type (i.e., Ala). The coefficient b_{12}, which is the coefficient at sin $\lambda^{(1)}$ (eq 4), is close to 1.0 kcal/mol regardless of the parametrization method, and b_{22} (the coefficient at sin $\lambda^{(2)}$) is close to −0.5 kcal/mol. These coefficients can therefore be considered to encode the core features of the local-interaction energy surfaces of terminally blocked Ala-type residues. In fact, the value and sign of b_{12} indicate that the region of the energy map with $\lambda^{(1)} < 0$ is strongly preferred over the region with $\lambda^{(1)} > 0$. Because $\lambda^{(1)}$ is close to the backbone angle ϕ and $\lambda^{(2)}$ is close to the backbone angle ψ,[33] this shows that conformations of L-amino acid residues with negative ϕ-angle values are favored; this is exactly what is observed in protein structures.[44] Likewise, the fact that b_{22} is negative points to structures with positive $\lambda^{(2)}$ (and, consequently, ψ) angles as favorable ones. This, together with the preferences with regard to $\lambda^{(1)}$ to a first-order approximation of the local-interaction energy surface, points to the C region[44] as the most favorable region of the conformational space of a single terminally blocked L-amino acid residue, in agreement with experiment.[44]

The above analysis of the first-order coefficients of the Fourier expansion of the local-interaction energy surfaces enables us to understand the physical meaning of the third-order correlation terms $u^{(3)}_{ij}$ expressed by eq 10. It should be noted that this contribution to the total energy is very important because, first, its weight is large (Table 3) and, second, it varies with the distance as r^{-3} (eqs 10 and 13), whereas the other nonlocal terms that involve the interactions between the peptide groups ($U_{p_ip_j}$, $U^{4}_{\mathrm{loc-el}}$, $U^{5}_{\mathrm{loc-el}}$, and $U^{6}_{\mathrm{loc-el}}$) vary with the distance[14] as

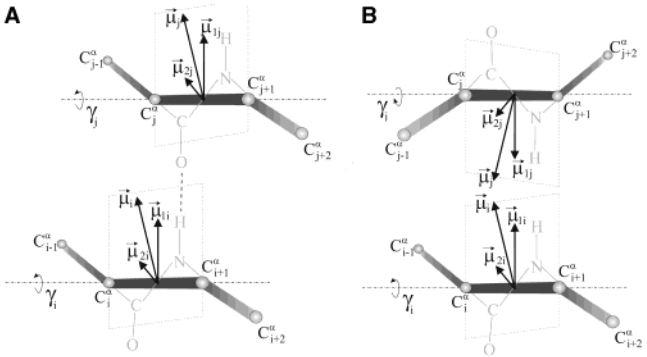


Figure 12. Illustration of the physical meaning of the $U^{(3)}_{\mathrm{loc-el}}$ correlation terms. The arrows represent the components of the dipole moments (which represent the peptide groups between the C^{α}_i and C^{α}_{i+1} (C^{α}_j and C^{α}_{j+1}) atoms, respectively; these peptide groups are represented by light-grey letters and lines) from the cumulant-based expression (eq 11) for the third-order correlation contribution to the RFE of two interacting peptide groups p_i and p_j considered together with the local interactions within the adjacent amino acid residues and the vector components of these dipole moments defined by eq 11. The positions of atoms C^{α}_i, C^{α}_{i+1}, and C^{α}_{i+2} (C^{α}_j, C^{α}_{j+1}, and C^{α}_{j+2}) are fixed, and so are the vector components $\boldsymbol{\mu}_{1i}$ and $\boldsymbol{\mu}_{1j}$ of the dipole moments, whereas the virtual bonds $C^{\alpha}_{i-1}\cdots C^{\alpha}_i$ ($C^{\alpha}_{j-1}\cdots C^{\alpha}_j$) and the components of $\boldsymbol{\mu}_{2i}$ ($\boldsymbol{\mu}_{2j}$) can rotate about the $C^{\alpha}_i\cdots C^{\alpha}_{i+1}$ ($C^{\alpha}_j\cdots C^{\alpha}_{j+1}$) axis by the angle γ_i (γ_j). (A) Parallel orientation of the planes defined by the C^{α}_i, C^{α}_{i+1}, C^{α}_{i+2} and the C^{α}_j, C^{α}_{j+1}, C^{α}_{j+2} atoms, as found in parallel β sheets. (B) Opposite orientation of these planes, which are not found in protein structures. The dotted rectangles mark the planes containing the $C^{\alpha}_i\cdots C^{\alpha}_{i+1}$ ($C^{\alpha}_j\cdots C^{\alpha}_{j+1}$) virtual bonds and the vectors $\boldsymbol{\mu}_i$ ($\boldsymbol{\mu}_j$).

r^{-6} and are, therefore, short-range interaction terms. Let us consider a pair of interacting peptide groups p_i and p_j together with the local interactions within the four adjacent peptide groups, as shown in Figure 12. By virtue of eq 10, the term $u^{(3)}_{ij}$ in the cumulant approximation to the RFE is the energy of interaction of two point dipoles μ_i and μ_j located in the centers of the peptide groups. Each of these dipoles is a sum of two vector components (eq 11): $\boldsymbol{\mu}_{1i}$ ($\boldsymbol{\mu}_{1j}$), which is equal to the vector $\mathbf{b}_{1i}$ ($\mathbf{b}_{1j}$) in the second-order Fourier expansion of the

local-interaction energy surface of a terminally blocked amino acid residue (eq 4) (this vector consists of the coefficients of $\cos\lambda^{(1)}$ and $\sin\lambda^{(1)}$, respectively) and $\boldsymbol{\mu}_{2i}$ ($\boldsymbol{\mu}_{2j}$), which is obtained by a reflection of the vector $\mathbf{b}_{2i}$ ($\mathbf{b}_{2j}$) in the plane perpendicular to that defined by atoms C_i^α, C_{i+1}^α, and C_{i+2}^α (C_j^α, C_{j+1}^α, and C_{j+2}^α) and subsequent rotation about the $C_i^\alpha\cdots C_{i+1}^\alpha$ ($C_j^\alpha\cdots C_{j+1}^\alpha$) virtual-bond axis by the angle γ_i (γ_j). The component $\boldsymbol{\mu}_1$ (fixed) is the component determined by the local interactions within residue $i+1$ ($j+1$), and the component $\boldsymbol{\mu}_2$ is determined by local interactions within residue i (j). Therefore, $u_{ij}^{(3)}$ can be understood as the energy of the interactions of peptide-group dipoles, with the orientation of the dipoles determined by the local interactions within the adjacent peptide groups.

Let us now consider the case of Ala-type residues. Because the second component (b_{12}) of vector $\mathbf{b_1}$ is dominant and positive (Table 2), the fixed components of the dipole moments ($\boldsymbol{\mu}_{1i}$ and $\boldsymbol{\mu}_{1j}$) point perpendicularly to the planes defined by the C_i^α, C_{i+1}^α, and C_{i+2}^α (C_j^α, C_{j+1}^α, and C_{j+2}^α) atoms in the direction shown in Figure 12. Because the second component of vector $\mathbf{b_2}$ (b_{22}) is dominant and negative (Table 2), the rotatable components of the dipole moments ($\boldsymbol{\mu}_{2i}$ and $\boldsymbol{\mu}_{2j}$) point in a direction nearly opposite to that of $\boldsymbol{\mu}_{1i}$ and $\boldsymbol{\mu}_{1j}$ if both angles γ in Figure 12 are zero and are aligned with $\boldsymbol{\mu}_{1i}$ and $\boldsymbol{\mu}_{1j}$, respectively, and if both angles γ are close to 180°. Therefore, for the $C_j^\alpha\cdots C_{j+1}^\alpha\cdots C_{j+2}^\alpha$ frame positioned exactly above the $C_i^\alpha\cdots C_{i+1}^\alpha\cdots C_{i+2}^\alpha$ frame (an orientation close to that found in parallel β sheets) as shown in Figure 12A, the extended configuration of the strands will result in the most negative value of the dipole−dipole interaction energy or, in other words, in the most favorable hydrogen-bonding interactions between the peptide groups (Figure 12). The same analysis can be applied to antiparallel β sheets. In the case of α helices, the alignment of $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ is not perfect; therefore, the α helices are less favorable than β sheets with regard to the coupling between the local and the electrostatic interactions. However, there are more hydrogen-bonded contacts in the α helices compared to the number in the β sheets involving chain segments with the same length. Therefore, these two types of secondary structures have comparable stability, and other interactions, presumably those involving the side chains, are responsible for the appearance of one or the other type of secondary structure. It should be noted that the fact that the right-handed helices are favored over the left-handed helices can also be explained in terms of the alignment of the $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ dipoles.

If the planes defined by the C_i^α, C_{i+1}^α and C_{i+2}^α (C_j^α, C_{j+1}^α, and C_{j+2}^α) atoms are oriented oppositely as in Figure 12, then the fixed parts of the dipole moments point in opposite directions, and this misalignment cannot be improved by rotating the second component of each dipole moments. Therefore, such configurations do not occur in proteins.

Because the purpose of the present study was to parametrize the UNRES energy terms and not to identify the physical origin of the energy components, which would require another study to draw solid conclusions, we will refrain from further analysis of the correlation terms. Nevertheless, because of the fact that some of these coefficients seem to be quite transferable and their values and sign can be connected with the core features of the conformational preferences of amino acid residues, with only some of them seeming to contribute to the approximate expression for the simplified energy (eq 4), a conclusion can be drawn that, by an analysis of the components of physics-based simplified energy functions, it is possible to determine quite accurately and understand the most important features and, in turn, the molecular forces responsible for the formation of structure of biomolecules.

TABLE 4: Results of the Test of the Force Field Obtained by Global Optimization

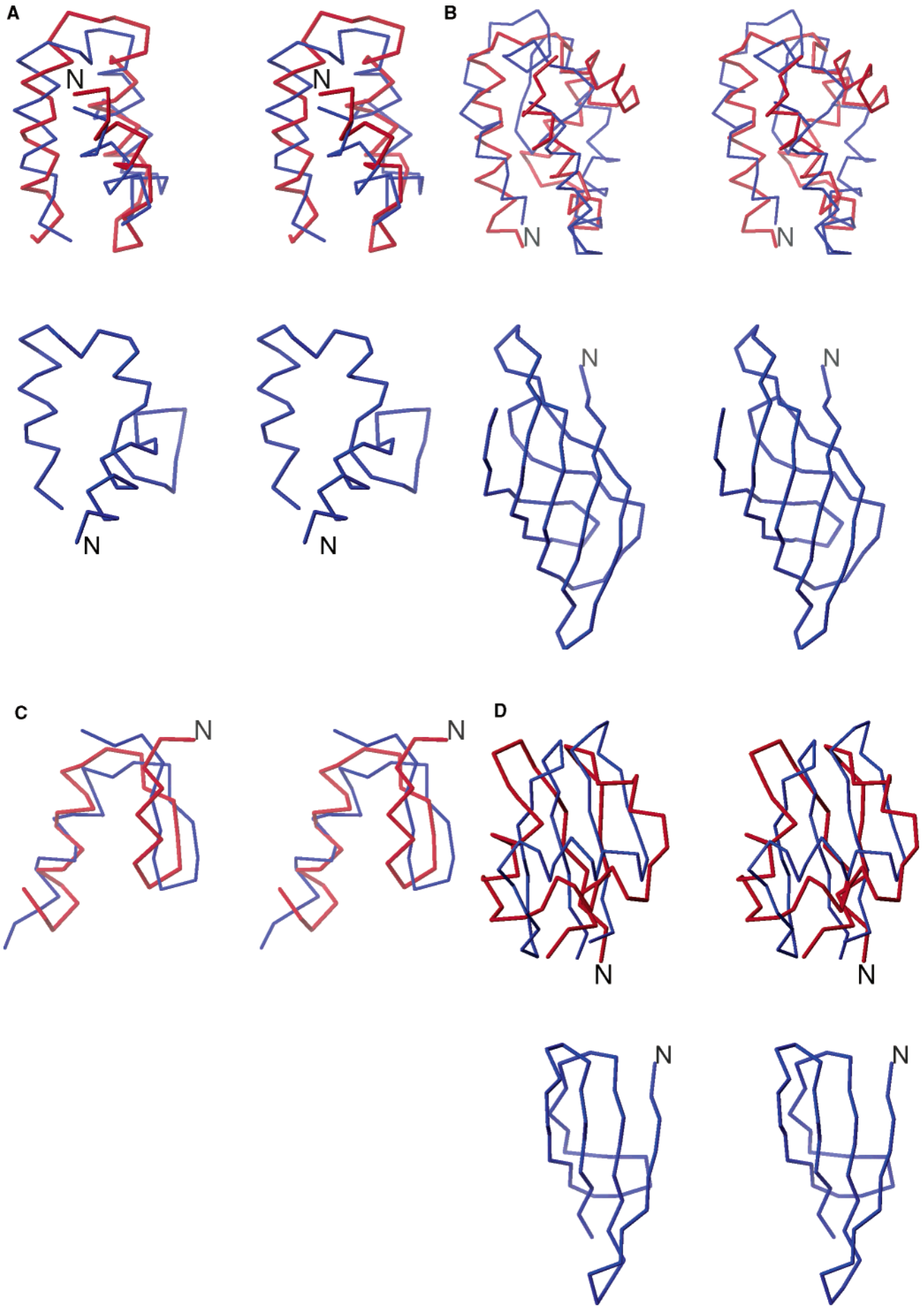
protein[a]	type	nres[b]	ΔE (kcal/mol)[c]	rmsd [Å][d]	ss[e]
1BDD	α	46	0.0	8.9	α
			4.4	3.7	α
1GAB	α	47	0.0	10.0	β
1CLB	α	75	0.0	10.4	α
			25.1	6.0	α
1POU	α	76	0.0	12.8	β
			9.9	8.4	α
1FSD	$\alpha+\beta$	28	0.0	3.2	$\alpha+\beta$
2PTL	$\alpha+\beta$	61	0.0	11.0	β
			20.8	8.9	$\alpha+\beta$
1UBQ	$\alpha+\beta$	75	0.0	13.0	β
1E0G	$\alpha+\beta$	48	0.0	10.7	β
			10.6	5.7	$\alpha+\beta$
1QHK	$\alpha+\beta$	45	0.0	10.3	$\alpha+\beta$
			33.5	6.1	$\alpha+\beta$
1E0L	β	28	0.0	3.9	β
1ED7	β	45	0.0	8.3	$\beta+\alpha$
			4.1	5.9	β
1BK2	β	57	0.0	8.4	β
1FYN	β	61	0.0	11.1	β
			21.8	7.4	β

[a] The upper row corresponds to the lowest-energy structure, and the lower row, to the lowest-energy structure with nativelike topology. For 1FSD and 1E0L, the nativelike structures had the lowest energies, but for 1GAB, 1UBQ, and 1BK2, no structure with nativelike topology was located. [b] Number of residues. [c] Relative energy with respect to the lowest-energy structure. [d] Root-mean-square deviation from the native structure. [e] Secondary structure of the calculated conformation.

3.3. Tests of the Force Field. We tested the optimized force field on several proteins with α, β, and mixed structure. These proteins were not included in the force-field optimization with 1IGD.

It should be stressed that the force field reported here is only the first version obtained with the parameters derived from ab initio energy surfaces and requires further optimization by our hierarchical procedure.[15] Further improvements of the force field will be reported in our forthcoming papers.[41,43]

The results of a global conformational search with the CSA method[11] are summarized in Table 4. The overlap of the native and the lowest-energy structures with nativelike topology and the lowest-energy structures (if different from the lowest-energy nativelike structures) of six selected examples [i.e., the 46-residue fragment of staphylococcal protein A (1BDD; α-helical structure),[45] the 28-residue 1FSD designed $\alpha+\beta$ peptide,[46] 1POU (α-helical structure),[47] 1E0G ($\alpha+\beta$ structure),[48] 1E0L (a three-stranded antiparallel β sheet),[49] and 1ED7 (mostly β structure)[50]] are shown in Figure 13. It can be seen that, for most of the test cases, a nativelike structure was obtained as a low-energy structure. It can therefore be stated that the force field is generally able to locate the structures with nativelike topology (in the sense of the gross correctness of the secondary structure and spatial arrangements of secondary-structure elements manifested by a reasonably low rmsd of about 0.1 Å/residue) of small proteins (with chain lengths of up to 80 amino acid residues) as the lowest- or at least the low-energy structures in global conformational searches with the CSA


A
N
N
B
N
N
N
N
N
N
C
N
N
D
N
N
N
N

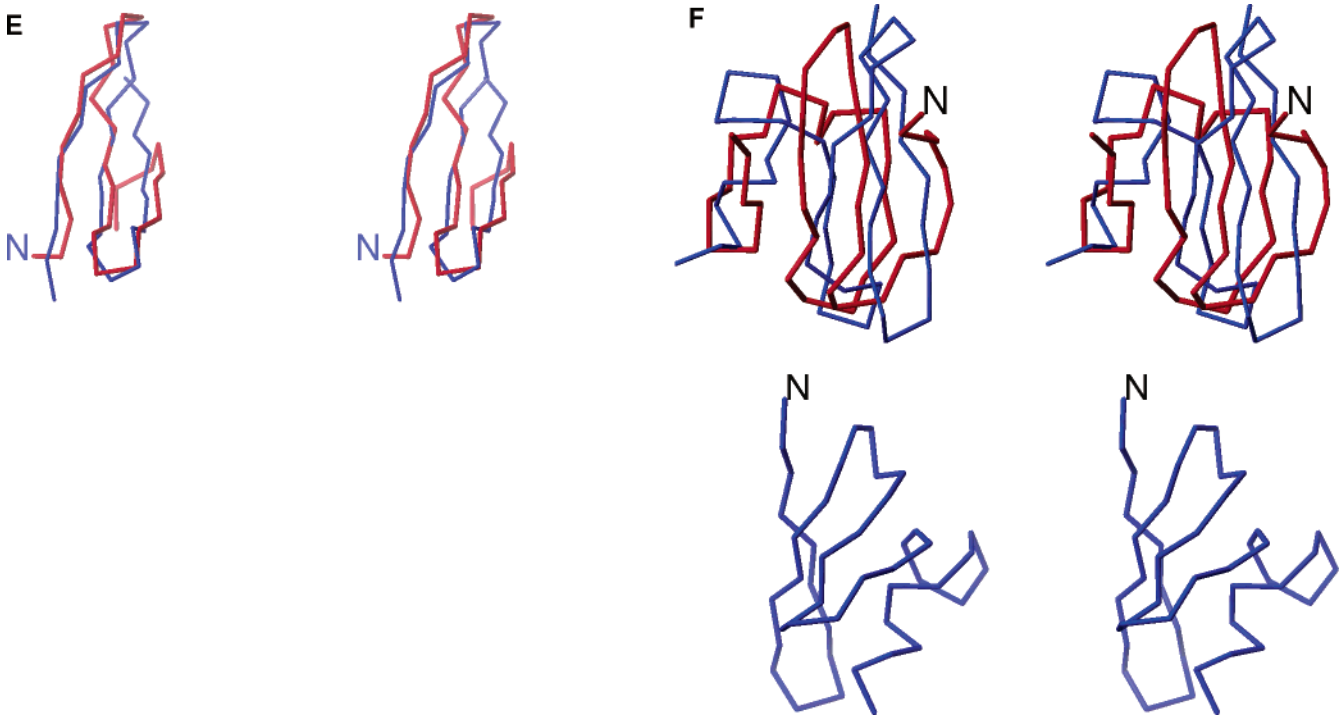


Figure 13. Stereoviews of the superpositions of the experimental structures (red) and most nativelike low-energy structures obtained with the UNRES force field parametrized in this work (blue) of protein A[45] (A; top, rmsd = 4.2 Å, 4.4 kcal/mol above the global minimum; bottom, lowest-energy non-native-like structure with the present force-field), 1POU[47] (B; top, rmsd = 8.8 Å; 13.2 kcal/mol above the global minimum; bottom, lowest-energy non-native-like structure with the present force-field), 1FSD[46] (C; rmsd = 3.2 Å; lowest-energy structure), 1E0G[48] (D; top, rmsd = 5.6 Å; 10.6 kcal/mol above the global minimum; bottom, lowest-energy non-native-like structure with the present force-field), 1E0L (E; rmsd = 3.9 Å; lowest-energy structure), and 1ED7[50] (F; top, rmsd = 5.9 Å, 4.1 kcal/mol above the global minimum; bottom, lowest-energy non-native-like structure with the present force-field). The N terminus is indicated.

method.[42] However, the predicted structures (including 1IGD, which was used to parametrize the force field; Figure 6) are characterized by significant rmsd values from the native structures, which shows that the force field is only moderately accurate. In particular, for the lowest-energy nativelike structure of the largest of the proteins analyzed, 1POU, the rmsd is quite large (8.4 Å); however, the topology of the native fold is preserved. A large rmsd value in this case arises from the fact that there is a coil fragment instead of the second helix in the calculated structure. Moreover, the force field apparently has a bias toward β structure because for two α-helical proteins (1GAB and 1POU) the lowest-energy structures are β structures, although α-helical structures were also obtained as low-energy structures. Nevertheless, the results obtained here are encouraging because the force field is able to find the native structures of proteins of diverse secondary structures and tertiary folds. It should also be kept in mind that only one protein was used to parametrize the force field and hierarchical optimization was aimed only at obtaining structures with all nativelike elements of the secondary structure (the N- and the C-terminal hairpins and the middle α helix), without considering their proper packing or the rmsd from the native structure. These features have recently been included[41,43] and gave significantly better results. These further improvements along with more extensive tests on a greater number of proteins will be reported in our forthcoming papers.[41,43]

The force field derived in this work was also tested in the Fifth Community Wide Experiment on the Critical Assessment of Techniques for Protein Structure Prediction (CASP5).[51] We submitted structures for 27 targets with lengths varying from 53 to 237 amino acid residues. For most of the targets, we predicted the structure of 60−80 residue segments with a 6.0-Å rmsd or lower, irrespective of structure type. This is a significant improvement with respect to our performance in the CASP4 exercise, in which most of our good predictions were proteins with only α-helical structure.[26] These improvements were achieved because of the careful parametrization of individual interactions based on the results of both ab initio calculations and our new hierarchical procedure that optimizes the energy landscape to a caldera-like form with the bottom region corresponding to the nativelike structure and energy decreasing with increasing degree of nativelikeness, which in turn ensures a quick search for the native structure.[15] Details of the results obtained in the CASP5 exercise will be published elsewhere.[52]

4. Conclusions

We parametrized the backbone−electrostatic and backbone−correlation terms in the UNRES force field by using ab initio energy surfaces. Such an approach is free from the imperfections of the all-atom force fields that we used previously to derive the parameters of these terms.[14,18] Together with the results of our recent parametrization of the torsional and double-torsional terms,[16] this work concludes the essential step of the parametrization of those UNRES energy terms that account for those interactions within the polypeptide backbone that stabilize regular secondary structures. The remaining terms for ab initio-based parametrization are the U_b (virtual-bond bending) and U_{rot} (side-chain rotamer) potentials, for which we now use the knowledge-based potentials derived in our earlier work.[9] Work on the ab initio-based parametrization of these contributions to the energy function is already in progress in our laboratory. It should be noted, however, that these terms seem to be less essential than the correlation, the torsional, and the double-torsional potentials parametrized in this work and in our most recent work,[16] respectively. More important is the replacement

of the statistics-based $U_{SC_iSC_j}$ by entirely physics-based potentials determined as potentials of mean force from molecular dynamics simulations on systems modeling interacting side chains; these potentials will be further calibrated to reproduce relevant experimental quantities (e.g., the association constants of compounds modeling amino acid side chains). We are already carrying out such simulations to treat the interactions of hydrophobic[53−55] and hydrophilic side chains.[56] The immediate goal of these studies is to determine physically justifiable functional forms of these potentials, which are largely arbitrary in the present version of UNRES.[8] Our final goal is to remove all knowledge-based energy terms from UNRES.

After the optimization of the energy-term weights of eq 1 and the simultaneous refinement of the coefficients of the cumulant-based expressions for the correlation terms directed at achieving a partially hierarchical structure of the energy landscape of the 1IGD ($\alpha + \beta$) protein, we obtained a 4−6-Å resolution force field that is capable of finding the nativelike structures of proteins with different types of secondary structure and different folds. It should be noted, however, that the complete parametrization of the energy-term weights and the refinement of parameters of the correlation terms will require the simultaneous hierarchical optimization of more proteins and the development of the hierarchical procedure itself. Recent developments in this field will be reported in our forthcoming papers.[41,43]

Acknowledgment. This work was supported by grants from the National Institutes of Health (GM-14312), the National Science Foundation (MCB00-03722), the Fogarty Foundation (TW1064), and grant DS 8372-4-0138-4 from the Polish State Committee for Scientific Research (KBN). Support was also received from the National Foundation for Cancer Research. This research was conducted by using the resources of (a) our 392-processor Beowulf cluster at the Baker Laboratory of Chemistry and Chemical Biology, Cornell University, (b) our 45-processor Beowulf cluster at the Faculty of Chemistry, University of Gdańsk, (c) the Informatics Center of the Metropolitan Academic Network (IC MAN) in Gdańsk, (d) the National Science Foundation Terascale Computing System at the Pittsburgh Supercomputer Center, and (e) the Interdisciplinary Center of Mathematical and Computer Modeling (ICM) at the University of Warsaw.

5. Appendix

5.1. Parametrization of $U_{p_ip_j}$. As mentioned in Methods, for each of the three systems used to parametrize $U_{p_ip_j}$, namely, AcNHMe−AcNHMe, AcNHMe−$AcNMe_2$, and $AcNMe_2$−$AcNMe_2$, the RFE surface was evaluated as a function of the distance between peptide-group centers r and the angles $\theta^{(1)}$, $\theta^{(2)}$, and ϕ defining the orientation of two peptide groups (Figure 3). For the AcNHMe−AcNHMe system, the following grid was used: r from 4.0 to 7.8 Å with a step size 0.2 Å, $\theta^{(1)}$ from 0 to 135° with a step size of 45°, $\theta^{(2)}$ from 0 to 180° with a step size of 45°, and ϕ from 0 to 180° with a step size of 45°. For the AcNHMe−$AcNMe_2$ and the $AcNMe_2$−$AcNMe_2$ systems, r varied from 4.6 to 8.0 Å with a step size of 0.2 Å, and the grid in $\theta^{(1)}$, $\theta^{(2)}$, and ϕ was the same as for the AcNHMe−AcNHMe system.

For each point, the RFE of the two interacting peptide groups, F_i^{el}, was evaluated numerically by integrating the electrostatic energy surfaces, E_{ij}, over the angles λ, as given by eq A-1, which was adapted from eq 7 of ref 14.

$$F_k^{el} = -RT \ln\left\{\frac{1}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \times \exp\left[\frac{-E_{ij}(\lambda_i, \lambda_j; r_{ij;k}, \theta_{ij;k}^{(1)}, \theta_{ij;k}^{(2)}, \phi_{ij;k})}{RT}\right] d\lambda_1\, d\lambda_2\right\}$$
$$\approx -RT \ln\left\{\frac{\Delta^2}{(2\pi)^2}\sum_{l=-\pi/\Delta}^{\pi/\Delta}\sum_{m=-\pi/\Delta}^{\pi/\Delta} \times \exp\left[\frac{-E_{ij}(l\Delta, m\Delta; r_{ij;k}, \theta_{ij;k}^{(1)}, \theta_{ij;k}^{(2)}, \phi_{ij;k})}{RT}\right]\right\} \quad \text{(A-1)}$$

where k is the index of the point of the energy surface and i and j are the indices of the first and the second peptide groups in the pair, respectively. We assumed that $T = 298$ K. The increment Δ was assigned the value of 5°. The parameters of $U_{p_ip_j}$ were fitted to the calculated RFE surfaces by using a linear least-squares method. Each "data point" in the general expression of the minimized target function given in eq 16 was weighted by the respective Boltzmann factor, as given by eq A-2.

$$w_i = \exp\left[-\frac{F_i^{el} - F_{min}^{el}}{RT}\right] \quad \text{(A-2)}$$

where F_{min}^{el} is the minimum value of the RFE over all distances and orientations considered. The calculations of the all-atom energy surfaces E_{ij} are described in section 5.3.

5.2. Parametrization of the Correlation Terms. As mentioned in Methods, the RFEs of terminally blocked tetrapeptides used to determine the parameters of the cumulant-based expressions, namely, Ac-Ala-Ala-Ala-Ala-NHMe, Ac-Ala-Gly-Gly-Ala-NHMe, Ac-Ala-Gly-Ala-Ala-NHMe, Ac-Ala-Ala-Gly-Ala-NHMe, Ac-Ala-Ala-Pro-Ala-NHMe, Ac-Ala-Pro-Ala-Ala-NHMe, and Ac-Ala-Pro-Gly-Ala-NHMe, were calculated as functions of the three virtual-bond dihedral angles γ_1, γ_2, and γ_3 (Figure 5). Each angle was varied from −180 to 180° with a step size of 45°. Integration was carried out over the five dihedral angles λ shown in Figure 5 with a step size of $\Delta = 15°$. A general expression for that part of the RFE that needs to be considered for the model systems described above at the ith point is given by eq A-3.

$$F_i^{tetr} = -RT \ln\left\{\frac{1}{(2\pi)^5}\int_{-2\pi}^{2\pi}\cdots\int_{-2\pi}^{2\pi} \times \exp\left[-\frac{\sum_{j=1}^{3} e_j(\lambda_j, \pi - \gamma_{ij} - \lambda_{j+1}) + e_4(\lambda_4, \lambda_5)}{RT} + \frac{\sum_{j=1}^{4}\sum_{k=0}^{j-1} E_{jk}(\lambda_j, \lambda_k; \gamma_{i1}, \gamma_{i2}, \gamma_{i3})}{RT}\right] d\lambda_1 \ldots d\lambda_5\right\} \quad \text{(A-3)}$$

where e_j denotes the all-atom local-interaction energy surface of residue j (which depends on λ_j and λ_{j+1}) and E_{jk} denotes the all-atom energy surface for the interaction of peptide groups j and k (which depends on λ_j and λ_k and on the distance and orientation of the two peptide groups, which is in turn determined by λ_j, λ_k, and the angles γ_1, γ_2, and γ_3 because for every peptide unit the virtual-bond angle θ is unequivocally determined given the angles $\lambda^{(1)}$ and $\lambda^{(2)}$, as given by eq A16

in ref 17; the angle $\lambda_i^{(2)}$ is, in turn, calculated as a function of the respective virtual-bond dihedral angle γ_i, as given by eq A-5). The second argument in all but the last expression for e results from the fact that e is a function of the dihedral angles $\lambda_i^{(1)}$ and $\lambda_i^{(2)}$ defined by Nishikawa et al.,[33] and we have

$$\lambda_i^{(1)} = \lambda_i \tag{A-4}$$

$$\lambda_i^{(2)} = \pi - \gamma_i - \lambda_{i+1} \tag{A-5}$$

Equation A-5 is derived from eq 10 in ref 33. The local-interaction energy surfaces were determined by us recently from ab initio MP2/6-31G(d,p) calculations.[16] The determination of the all-atom peptide-group interaction energy surfaces, E_{jk}, is described in section 5.3.

To extract the correlation terms, the quantities $U_{p_ip_j}$, U_{tor}, and U_{tord} (calculated with the use of parameters determined in this and in our earlier work[16]) were subtracted from F^{tetr}, as given by eq A-6.

$$F'^{\text{tetr}}_i = F^{\text{tetr}}_i - [U_{\text{tor}}(\gamma_1) + U_{\text{tor}}(\gamma_2) + U_{\text{tor}}(\gamma_3) + U_{\text{tord}}(\gamma_1, \gamma_2) + U_{\text{tord}}(\gamma_2, \gamma_3) + \sum_{j=1}^{4}\sum_{k=0}^{j-1} U_{p_jp_k}] \tag{A-6}$$

The F'^{tetr} surfaces were fitted directly by sums of UNRES correlation terms. Because this is a nonlinear least-squares problem, Marquardt's iterative method was used.[57] The weight of each energy term was assumed to be 1 because we found that varying the weights does not improve the fit and the minimization problem became underdetermined. Each term in the expression for the minimized function (eq 16) was assigned a weight of 1 because the span of F'^{tetr} was only several kcal/mol and we wanted to obtain a function that not only can reproduce the RFE of low-energy structures but also can distinguish the secondary structure found in real proteins (e.g., right-handed α helices) from unlikely or disordered structures (e.g., left-handed helices). Fitting was carried out in three stages. First, the parameters of Ala were determined. Subsequently, taking the Ala parameters as known, we determined the parameters of Gly, and in the final stage, taking the parameters of Ala and Gly as known, we determined the parameters of Pro.

5.3. Determination of the All-Atom Energy Surfaces of Interacting Peptide Groups. As mentioned in section 2.2, AcNHMe and $AcNMe_2$ were used as models of peptide groups. The molecules were assumed to be rigid with valence geometry taken from the ECEPP/3 database.[27,30,58] By contrast with the integration of the local-interaction energy surfaces e, in which case there are only two variables for each surface ($\lambda^{(1)}$ and $\lambda^{(2)}$), the energy for the interaction of two peptide groups, E_{ij}, also depends on their intermolecular distance and orientation. It would therefore be impractical to calculate E_{ij} by an expensive ab initio method at each point considered in the integration (eqs A-1 and A-3). We therefore decided to evaluate E_{ij} by ab initio methods at a coarser grid and subsequently fit an analytical expression typical of an all-atom force field to the ab initio energy surfaces.

Ab initio calculations were carried out in the MP2/6-31G(d,p) scheme on the following three pairs: AcNHMe−AcNHMe, AcNHMe−$AcNMe_2$, and $AcNMe_2$−$AcNMe_2$, respectively. The basis set superposition error (BSSE) was estimated by using the Boys−Bernardi method.[59] The following ranges of variables

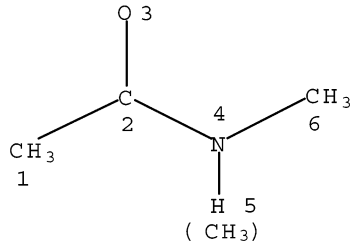


Figure 14. Atom numbering scheme for the AcNHMe and $AcNMe_2$ molecules.

and grid sizes were used (see Figure 3 for the definition of variables): 4.0 Å ≤ r ≤ 8.0 Å with a 0.5-Å grid size; 0° ≤ $\theta^{(1)}$ ≤ 90° with a 45° grid size; 0° ≤ ϕ ≤ 180° with a 45° grid size; 0° ≤ $\theta^{(2)}$ < 180° with a 180°/[max(4| cos ϕ| + 0.5, 1)] grid size; and 0 ≤ λ_1 < 360° and 0 ≤ λ_2 < 360° with a 15° grid size. The nonuniform grid size in $\theta^{(2)}$ provided a uniform distribution of points in the Cartesian space. For each system, the first series of runs was carried out having assumed that $\theta^{(1)} = 90°$ and $\phi = \theta^{(2)} = 0°$ and varying r, λ_1, and λ_2, whereas the second series of runs was carried out with $\theta^{(1)} = 90°$, $\phi = 0°$, and $\theta^{(2)} = 180°$ and varying r, λ_1, and λ_2. In the third series of runs, r was fixed at 5.0 Å, which is close to the energy minimum for all systems considered, and $\theta^{(1)}$, $\theta^{(2)}$, ϕ, $\lambda^{(1)}$, and $\lambda^{(2)}$ were varied over the ranges cited above.

The approximate all-atom energy function to which the ab initio surfaces for $U_{p_ip_j}$ were fitted is given by eq A-7.

$$E_k = \sum_{i=1}^{6}\sum_{j=1}^{6}\epsilon_{ij}\left[\left(\frac{\rho_{ij}^{\circ}}{r_{ijk}}\right)^{12} - 2\left(\frac{\rho_{ij}^{\circ}}{r_{ijk}}\right)^{6}\right] + \frac{q_iq_j}{r_{ijk}} \tag{A-7}$$

where the index i numbers the atoms in the first molecule, the index j numbers those in the second molecule, k is the number of the point of the energy surface, the Lennard-Jones potential parameters ϵ_{ij} and ρ_{ij}° are characteristic of a given pair of atoms in a given system, and the charges q_i and q_j are also characteristic of a system. The atom-numbering system is given in Figure 14. The terminal methyl groups were considered to be united atoms, and the charges on them were fixed at zero. All parameters of eq A-7, including the partial atomic charges, were determined by a nonlinear least-squares fitting of eq A-7 for a given system to the respective ab initio energy surface by using the Marquardt method.[57] The minimized sum of squares is given by eq A-8.

$$\Phi = \sum_{k=1}^{N} w_k(E_k^{\text{ab initio}} - E_k)^2 \tag{A-8}$$

with

$$w_k = \exp\left[-\frac{E_k^{\text{ab initio}} - E_{\text{min}}^{\text{ab initio}}}{RT}\right] \tag{A-9}$$

where E_{min} denotes the minimum value of the energy over all points considered in the fitting. Such weighting emphasizes points with low energy that, in turn, make the most significant contribution to the RFE. We assumed that T = 298 K.

It should be noted that the aim of this approach was to derive a function to approximate the particular ab initio energy surfaces of models of specific interacting peptide groups. Therefore, the parameters derived in this way cannot be regarded as transferable to other systems.

TABLE 5: Parameters of Simplified All-Atom Energy Functions for $U_{p_i p_j}$ (Eq A-7) of the Pairs of Peptide-Group Models[a]

(a) AcNHMe−AcNHMe system

	atom number[b]					
atom number	1	2	3	4	5	6
			ϵ(kcal/mol)			
1	0.01000					
2	0.01000	0.01000				
3	0.01000	0.01000	3.48215			
4	0.01000	0.01001	0.01000	0.01000		
5	0.74185	0.01000	1.54294	0.01000	3.91968	
6	0.01000	0.01000	0.18617	0.01000	0.01000	0.01000
			ρ° (Å)[b]			
1	3.11287					
2	2.60061	2.78743				
3	4.77948	2.44439	2.35217			
4	2.53155	5.06993	4.66335	2.52529		
5	3.18212	3.44372	1.71221	2.12346	1.88376	
6	2.91083	2.41847	2.99154	2.24209	2.09067	2.76540
			charge (electron charge unit)			
	0.00000	0.11554	−0.32465	−0.11611	0.32522	0.00000

weighted standard deviation of the energy: 0.14 kcal/mol

(b) AcNHMe−$AcNMe_2$ system

atom number	atom number ($AcNMe_2$)[b]					
(AcNHMe)	1	2	3	4	5	6
			ϵ(kcal/mol)			
1	0.04222	0.01000	0.01000	0.01002	0.01472	0.01006
2	0.01000	0.01000	0.01000	0.01178	0.02882	0.03576
3	1.03706	0.05858	0.09898	0.07533	0.01000	1.02973
4	0.06136	0.01000	0.01001	0.01000	0.01000	0.01000
5	0.59503	0.33676	1.00243	0.01000	1.42227	0.03281
6	0.01439	0.01000	0.55270	0.01000	0.01000	0.01000
			ρ° (Å)			
1	5.09151	3.87040	3.47622	3.58478	2.89112	3.70601
2	3.39418	3.31097	3.39841	2.80773	4.67582	3.06566
3	3.44356	2.57731	3.34657	4.08844	2.73053	3.54501
4	1.92567	2.78396	4.70091	3.93998	3.66589	3.41878
5	3.23907	2.59116	1.86664	3.31329	2.80570	3.66102
6	4.91793	4.84502	3.21520	3.10835	3.41774	3.42595
			charge (electron charge unit)			
AcNHMe	0.00000	−0.00531	−0.21075	−0.16321	0.37927	0.00000
$AcNMe_2$	0.00000	0.15481	−0.32055	0.03422	0.13151	0.00000

weighted standard deviation of the energy: 0.13 kcal/mol

(c) $AcNMe_2$−$AcNMe_2$ system

	atom number[b]					
atom number	1	2	3	4	5	6
			ϵ(kcal/mol)			
1	0.01000					
2	0.01000	0.01001				
3	0.07298	0.21499	0.01333			
4	0.01000	0.01050	1.18875	0.01000		
5	0.01000	0.01000	0.01000	0.01012	1.07417	
6	0.01052	0.07963	0.01000	0.34584	0.01000	0.03466
			ρ° (Å)			
1	3.10332					
2	3.80283	3.64190				
3	3.13615	4.19851	3.11444			
4	3.09198	4.74660	2.82373	3.44730		
5	2.50973	3.75653	4.58483	4.74314	3.75441	
6	1.17934	4.33703	4.56681	3.57805	4.54830	4.12684
			charge (electron charge unit)			
	0.00000	0.19437	−0.24519	−0.27488	0.32570	0.00000

weighted standard deviation of the energy: 0.34 kcal/mol

[a] Rows represent the first and columns the second molecule in a pair. [b] See Figure 14 for atom numbering.

We also tried more complicated expressions than that given by eq A-7 (e.g., those that contained the 6-9 instead of 6-12 potential as well as the r^{-4} and r^{-3} terms to mimic the polarization and dipole−dipole interactions, respectively). However, these expressions did not give a remarkably better fit to the ab initio energy surfaces but did pose additional problems

due to their weak ability to determine the parameters. We therefore decided to use eq A-7 to represent the all-atom energy surfaces of interacting peptide groups.

The parameters of the all-atom force fields obtained by fitting expressions for the peptide-group interaction energy given by eq A-7 for each of the three systems considered to the respective MP2/6-31G(d,p) energy surfaces are summarized in Table 5. As stressed above, eq A-7 together with parameters from Table 5 for a particular system constitutes an approximation to the respective ab initio energy surface and is not a transferable force field. The RFE curves calculated as functions of the distances r between the peptide-group centers at $\theta^{(1)} = 90°$, $\phi = \theta^{(2)} = 0°$ from the ab initio energy surfaces and using the approximate expression are shown in Figure 7 (solid lines). It can be seen for the systems that can form a hydrogen bond (AcNHMe−AcNHMe and AcNHMe−$AcNMe_2$) that the RFE has a minimum at $r = 5.0$ Å (Figure 7A and B); this distance is close to the most frequently occurring distances (about 4.8 Å, Figure 8) between the centers of hydrogen-bonded peptide groups found in protein structures (α helices, β sheets, and turns). For $AcNMe_2$−$AcNMe_2$ there is no minimum close to $r = 5.0$ Å, and the RFE increases monotonically with the distance (Figure 7C). There is only an inflection point at $r = 6.0$ Å. The reason for this is that, for the first two systems, hydrogen-bonded configurations correspond to the lowest energy; these configurations are characterized by r close to 5.0 Å. Conversely, no hydrogen bond can be formed in the case of $AcNMe_2$−$AcNMe_2$; therefore, the lowest-energy configurations are characterized by the "stacking" of the peptide groups with the peptide-group dipoles antiparallel to each other. In this case, the minimum distance is close to the sum of the largest VDW radii of the atoms of the *N*-methylated peptide group (i.e., about 4.0 Å). The stacking configurations also make a remarkable contribution to the RFE of regular peptide groups, which is manifested in Figure 7A as a moderate slope of the RFE for distances from 4.0 to 5.0 Å and the appearance of an inflection point at $r =$ 4.5 Å.

It can be seen that, for distances larger than 4.5 Å (Figure 7A−C, long-dashed lines), the simplified expression (eq A-7) well fits the ab initio energy surfaces of all systems considered except for the $AcNMe_2$−$AcNMe_2$ system. A more complex approximation than the simple expression given by eq A-7 is required to fit the ab initio energy surface of this system at short distances between peptide-group centers. However, because proline residues at a short distance are not frequently found in protein structures, we did not think it worthwhile to pursue this subject and simply discarded intermolecular distances of less than 5.0 Å between the centers of proline-type peptide groups from further considerations.

It is interesting that the fitted charges on the atoms of AcNHMe and $AcNMe_2$ vary depending on the pair considered. This suggests that polarization plays an important role in peptide-group interactions. (If it were not so, the charges would vary little.) This is also supported by the fact that the charges are about 2 times smaller in absolute value than those used in the AMBER force field,[31,32] which were obtained by fitting the molecular electrostatic potential calculated from the ab initio wave function and subsequent refinement to reproduce the energetics and dipole moments of small systems. This in turn indicates that the electrostatic-potential-based charges used in empirical force fields should probably be scaled down by a factor of 2 or, alternatively, a dielectric constant of 4 should be used to compute the effective electrostatic energy (i.e., by taking the polarization interactions implicitly into account) even if explicit solvent is included. However, because the charges derived here are parameters of very special nontransferable force fields, further studies would be required to justify this conclusion.

References and Notes

(1) Scheraga, H. A. Prediction of Protein Conformation. In *Current Topics in Biochemistry*; Anfinsen, C. B., Schechter, A. N., Eds.; Academic Press: New York, 1974; pp 1−42.

(2) Scheraga, H. A. *Biopolymers* **1983**, *22*, 1−14.

(3) Scheraga, H. A. *Int. J. Quantum Chem.* **1992**, *42*, 1529−1536.

(4) Vásquez, M.; Némethy, G.; Scheraga, H. A. *Chem. Rev.* **1994**, *94*, 2183−2239.

(5) Scheraga, H. A. *Biophys. Chem.* **1996**, *59*, 329−339.

(6) Scheraga, H. A.; Lee, J.; Pillardy, J.; Ye, Y.-J.; Liwo, A.; Ripoll, D. R. *J. Global Optimization* **1999**, *15*, 235−260.

(7) Scheraga, H. A.; Pillardy, J.; Liwo, A.; Lee, J.; Czaplewski, C.; Ripoll, D. R.; Wedemeyer, W. J.; Arnautova, Y. A. *J. Comput. Chem.* **2002**, *23*, 28−34.

(8) Liwo, A.; Ołdziej, S.; Pincus, M. R.; Wawak, R. J.; Rackovsky, S.; Scheraga, H. A. *J. Comput. Chem.* **1997**, *18*, 849−873.

(9) Liwo, A.; Pincus, M. R.; Wawak, R. J.; Rackovsky, S.; Ołdziej, S.; Scheraga, H. A. *J. Comput. Chem.* **1997**, *18*, 874−887.

(10) Liwo, A.; Kaźmierkiewicz, R.; Czaplewski, C.; Groth, M.; Ołdziej, S.; Wawak, R. J.; Rackovsky, S.; Pincus, M. R.; Scheraga, H. A. *J. Comput. Chem.* **1998**, *19*, 259−276.

(11) Lee, J.; Liwo, A.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 2025−2030.

(12) Liwo, A.; Pillardy, J.; Kaźmierkiewicz, R.; Wawak, R. J.; Groth, M.; Czaplewski, C.; Ołdziej, S.; Scheraga, H. A. *Theor. Chem. Acc.* **1999**, *101*, 16−20.

(13) Liwo, A.; Lee, J.; Ripoll, D. R.; Pillardy, J.; Scheraga, H. A. *Proc. Natl. Acad. Sci., U.S.A.* **1999**, *96*, 5482−5485.

(14) Liwo, A.; Czaplewski, C.; Pillardy, J.; Scheraga, H. A. *J. Chem. Phys.* **2001**, *115*, 2323−2347.

(15) Liwo, A.; Arłukowicz, P.; Czaplewski, C.; Ołdziej, S.; Pillardy, J.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 1937−1942.

(16) Ołdziej, S.; Kozłowska, U.; Liwo, A.; Scheraga, H. A. *J. Phys. Chem. B* **2003**, *107*, 8035−8046.

(17) Liwo, A.; Pincus, M. R.; Wawak, R. J.; Rackovsky, S.; Scheraga, H. A. *Protein Sci.* **1993**, *2*, 1697−1714.

(18) Liwo, A.; Pincus, M. R.; Wawak, R. J.; Rackovsky, S.; Scheraga, H. A. *Protein Sci.* **1993**, *2*, 1715−1731.

(19) Liwo, A.; Pillardy, J.; Czaplewski, C.; Lee, J.; Ripoll, D. R.; Groth, M.; Rodziewicz-Motowidło, S.; Kaźmierkiewicz, R.; Wawak, R. J.; Ołdziej, S.; Scheraga, H. A. UNRES−A United-Residue Force Field for Energy-Based Prediction of Protein Structure: Origin and Significance of Multibody Terms. In *RECOMB 2000*, Proceedings of the Fourth Annual International Conference on Computational Molecular Biology; Shamir, R., Miyano, S., Istrail, S., Pevzner, P., Waterman, M., Eds.; ACM: New York, 2000; pp 193−200.

(20) Skolnick, J.; Kolinski, A. *Science* **1990**, *250*, 1121−1125.

(21) Skolnick, J.; Kolinski, A.; Brooks, C. L., III; Godzik, A.; Rey, A. *Curr. Biol.* **1993**, *3*, 414−423.

(22) Godzik, A.; Kolinski, A.; Skolnick, J. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 397−438.

(23) Third Community Wide Experiment on the Critical Assessment of Techniques for Protein Structure Prediction; Asilomar Conference Center, December 13−17, 1998; http://predictioncenter.llnl.gov/casp3/Casp3.html.

(24) Orengo, C. A.; Bray, J. E.; Hubbard, T.; LoConte, L.; Sillitoe, I. *Proteins: Struct. Funct. Genet.* **1999**, *Suppl. 3*, 149−170.

(25) Kubo, R. *J. Phys. Soc. Jpn.* **1962**, *17*, 1100−1120.

(26) Pillardy, J.; Czaplewski, C.; Liwo, A.; Lee, J.; Ripoll, D. R.; Kaźmierkiewicz, R.; Ołdziej, S.; Wedemeyer, W. J.; Gibson, K. D.; Arnautova, Y. A.; Saunders, J.; Ye, Y.-J.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 2329−2333.

(27) Némethy, G.; Gibson, K. D.; Palmer, K. A.; Yoon, C. N.; Paterlini, G.; Zagari, A.; Rumsey, S.; Scheraga, H. A. *J. Phys. Chem.* **1992**, *96*, 6472−6484.

(28) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F., Jr.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J. Mol. Biol.* **1977**, *112*, 535−542.

(29) Fauchere, J.-L.; Pliška, V. *Eur. J. Med. Chem.* **1983**, *18*, 369−375.

(30) Momany, F. A.; McGuire, R. F.; Burgess, A. W.; Scheraga, H. A. *J. Phys. Chem.* **1975**, *79*, 2361−2381.

(31) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S., Jr.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765−784.

(32) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J. Comput. Chem.* **1986**, *7*, 230−252.

(33) Nishikawa, K.; Momany, F. A.; Scheraga, H. A. *Macromolecules* **1974**, *7*, 797−806.
(34) Lee, J.; Ripoll, D. R.; Czaplewski, C.; Pillardy, J.; Wedemeyer, W. J.; Scheraga, H. A. *J. Phys. Chem. B* **2001**, *105*, 7291−7298.
(35) Pillardy, J.; Czaplewski, C.; Liwo, A.; Wedemeyer, W. J.; Lee, J.; Ripoll, D. R.; Arłukowicz, P.; Ołdziej, S.; Arnautova, Y. A.; Scheraga, H. A. *J. Phys. Chem. B* **2001**, *105*, 7299−7311.
(36) Derrick, J. P.; Wigley, D. B. *J. Mol. Biol.* **1994**, *243*, 906−918.
(37) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E., III; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. *Comput. Phys. Commun.* **1995**, *91*, 1−41.
(38) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.
(39) Roterman, I. K.; Gibson, K. D.; Scheraga, H. A. *J. Biomol. Struct. Dyn.* **1989**, *7*, 391−419.
(40) Roterman, I. K.; Lambert, M. H.; Gibson, K. D.; Scheraga, H. A. *J. Biomol. Struct. Dyn.* **1989**, *7*, 421−453.
(41) Ołdziej, S.; Liwo, A.; Czaplewski, C.; Pillardy, J.; Scheraga, H. A. *J. Phys. Chem. B*, submitted for publication, 2004.
(42) Lee, J.; Scheraga, H. A. *Int. J. Quantum Chem.* **1999**, *75*, 255−265.
(43) Ołdziej, S.; Łągiewka, J.; Liwo, A.; Czaplewski, C.; Scheraga, H. A. *J. Phys. Chem. B*, submitted for publication, 2004.
(44) Zimmerman, S. S.; Pottle, M. S.; Némethy, G.; Scheraga, H. A. *Macromolecules* **1977**, *10*, 1−9.
(45) Gouda, H.; Torigoe, H.; Saito, A.; Sato, M.; Arata, Y.; Shimada, I. *Biochemistry* **1992**, *31*, 9665−9672.
(46) Dahiyat, B. I.; Mayo, S. L. *Science* **1997**, *278*, 82−87.
(47) Assa-Munt, N.; Mortishire-Smith, R. J.; Aurora, R.; Herr, W.; Wright, P. E. *Cell* **1993**, *73*, 193−205.
(48) Bateman, A.; Bycroft, M. *J. Mol. Biol.* **2000**, *299*, 1113−1119.
(49) Macias, M. J.; Gervais, V.; Civera, C.; Oschkinat, H. *Nat. Struct. Biol.* **2000**, *7*, 375−379.
(50) Ikegami, T.; Okada, T.; Hashimoto, M.; Seino, S.; Watanabe, T.; Shirakawa, M. *J. Biol. Chem.* **2000**, *275*, 13654−13661.
(51) Czaplewski, C.; Ripoll, D. R.; Ołdziej, S.; Kazmierkiewicz, R.; Vila, J. A.; Liwo, A.; Pillardy, J.; Saunders, J. A.; Chinchio, M.; Nanias, M.; Khalili, M.; Arnautova, Y. A.; Jagielska, A.; Kang, Y. K.; Gibson, K. D.; Scheraga, H. A. Physics-Based Protein-Structure Prediction Using the UNRES and ECEPP/3 Force Fields: Tests on CASP5 Targets, In *Fifth Community Wide Experiment on the Critical Assessment of Techniques for Protein Structure Prediction*, 2002. http://predictioncenter.llnl.gov/casp5/Casp5.html; p. A−142−143.
(52) Czaplewski, C.; Ripoll, D. R.; Ołdziej, S.; Kaźmierkiewicz, R.; Vila, J. A.; Liwo, A.; Pillardy, J.; Saunders, J. A.; Chinchio, M.; Nanias, M.; Khalili, M.; Arnautova, Y. A.; Jagielska, A.; Kang, Y. K.; Gibson, K. D.; Scheraga, H. A., to be submitted for publication.
(53) Czaplewski, C.; Rodziewicz-Motowidło, S.; Liwo, A.; Ripoll, D. R.; Wawak, R. J.; Scheraga, H. A. *Protein Sci.* **2000**, *9*, 1235−1245.
(54) Czaplewski, C.; Ripoll, D. R.; Liwo, A.; Rodziewicz-Motowidło, S.; Wawak, R. J.; Scheraga, H. A. *Int. J. Quantum Chem.* **2002**, *88*, 41−55.
(55) Czaplewski, C.; Rodziewicz-Motowidło, S.; Dabal, M.; Liwo, A.; Ripoll, D. R.; Scheraga, H. A. *Biophys. Chem.* **2003**, *105*, 339−359.
(56) Maksimiak, K.; Rodziewicz-Motowidło, S.; Czaplewski, C.; Liwo, A.; Scheraga, H. A. *J. Phys. Chem. B* **2003**, *107*, 13496−13504.
(57) Marquardt, D. W. *J. Soc. Ind. Appl. Math.* **1963**, *11*, 431−441.
(58) Ripoll, D. R.; Liwo, A.; Czaplewski, C. *TASK Q.* **1999**, *3*, 313−331.
(59) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553−566.