United-Residue Force Field for Off-Lattice Protein-Structure Simulations: III. Origin of Backbone Hydrogen-Bonding Cooperativity in United-Residue Potentials

A. LIWO,[1,2,3] R. KAŹMIERKIEWICZ,[1] C. CZAPLEWSKI,[1] M. GROTH,[1] S. OŁDZIEJ,[1] R. J. WAWAK,[2] S. RACKOVSKY,[3] M. R. PINCUS,[4] H. A. SCHERAGA[2]

[1]*Department of Chemistry, University of Gdańsk, Gdańsk, Poland*
[2]*Baker Laboratory of Chemistry, Cornell University, Ithaca, New York 14853-1301*
[3]*Department of Biomathematical Sciences, Mount Sinai School of Medicine, New York, New York*
[4]*Department of Pathology, Brooklyn Veterans Administration Medical Center and State University of New York, Health Science Center, Brooklyn, New York*

Received 6 June 1997; accepted 22 August 1997

ABSTRACT: Based on the dipole model of peptide groups developed in our earlier work [Liwo et al., *Prot. Sci.*, **2,** 1697 (1993)], a cumulant expansion of the average free energy of the system of freely rotating peptide-group dipoles tethered to a fixed α-carbon trace is derived. A graphical approach is presented to find all nonvanishing terms in the cumulants. In particular, analytical expressions for three- and four-body (correlation) terms in the averaged interaction potential of united peptide groups are derived. These expressions are similar to the cooperative forces in hydrogen bonding introduced by Koliński and Skolnick [*J. Chem. Phys.*, **97,** 9412 (1992)]. The cooperativity arises here naturally from the higher order terms in the power-series expansion (in the inverse of the temperature) for the average energy. Test calculations have shown that addition of the derived four-body term to the statistical united-residue

Correspondence to: H. A. Scheraga; e-mail: has5@cornell.edu
Contract/grant sponsor: Polish State Committee for Scientific Research; contract/grant number: PB 190/T09/96/10
Contract/grant sponsor: National Institute on Aging; contract/grant number: AG-00322
Contract/grant sponsor: National Institute of General Medical Sciences; contract/grant number: GM-14312
Contract/grant sponsor: National Science Foundation; contract/grant number: MCB95-13167
Contract/grant sponsor: National Cancer Institute; contract/grant number: CA 42500

CCC 0192-8651 / 98 / 030259-18

potential of our earlier work [Liwo et al., *J. Comput. Chem.*, **18,** 849, 874 (1997)] greatly improves its performance in folding poly-L-alanine into an α-helix.
 J Comput Chem 19: 259–276, 1998

Keywords: protein folding; multibody interactions; electrostatic interactions; cumulant expansion; potential of mean force

Introduction

United-residue representations of polypeptide chains, in which each amino acid residue is represented by one or a few interaction sites, have received much attention, because, in contrast to the all-atom representation, their simplicity enables large-scale simulations of protein folding to be carried out.[1–8] In our previous work,[9,10] we proposed a united-residue model with side-chain (SC) and peptide-group (p) centers as the interaction sites (Fig. 1). The energy of the simplified chain is expressed by eq. (1):

$$U = \sum_{i<j} U_{SC_iSC_j} + \sum_{i\neq j} U_{SC_ip_j} + w_{el} \sum_{i<j-1} U_{p_ip_j} + w_{tor} \sum_i U_{tor}(\gamma_i) + w_{loc} \sum_i \left[U_b(\theta_i) + U_{rot}(\alpha_{SC_i}, \beta_{SC_i}) \right] + w_{corr} U_{corr} \quad (1)$$

where $U_{SC_iSC_j}$, $U_{SC_ip_j}$, and $U_{p_ip_j}$ denote the energies of the interactions between side chains, between side chains and peptide groups, and between peptide groups, respectively; $U_{tor}(\gamma_i)$ denotes the energy of variation of the virtual-bond dihedral angle γ_i; $U_b(\theta_i)$ denotes the ''bending'' energy of the virtual-bond angle θ_i; $U_{rot}(\alpha_{SC_i}, \beta_{SC_i})$ is the local energy of side-chain i; U_{corr} includes cooperative terms, and the w values denote relative weights of the respective energy terms.

The two-body ($U_{SC,SC}$, $U_{SC,p}$, and U_{pp}) and local-interaction (U_{tor}, U_b, and U_{rot}) terms were parameterized by a statistical analysis of protein-crystal data.[9,10] We have shown[10] that the potential, containing only two-body and local-interaction terms, is capable of recognizing the native structures of some moderately sized proteins among the structural motifs from the Protein Data Bank (PDB).[11]

Earlier, we developed a simpler (first generation) model of a polypeptide chain and the associated force field[5,6] for *de novo* simulations; it consists of only local and pairwise terms. In contrast to the new (second generation) force field presented here and in refs. 9 and 10, it assumes fixed virtual-valence geometry (with constant virtual bond C^α—C^α—C^α angles of 90° and constant angles α_{SC} and β_{SC} that define the location of the side-chain centroids with respect to the C^α—C^α—C^α frame). Its parameters were derived from interresidue contact energies determined by Miyazawa and Jernigan,[12] the geometric parameters of the virtual polypeptide chain determined

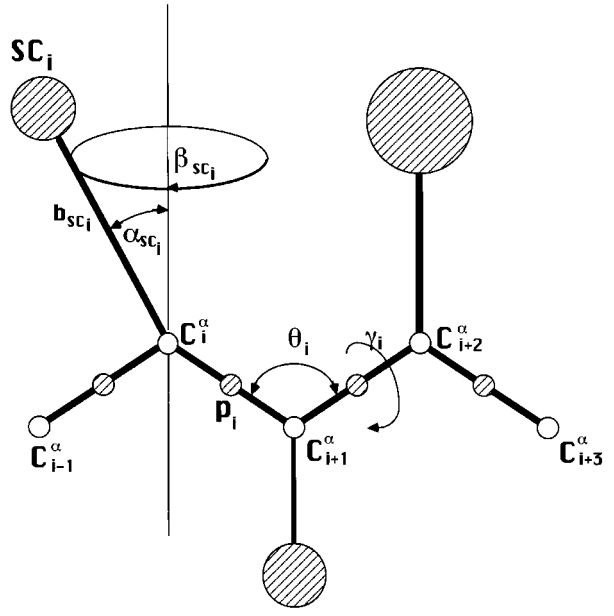

FIGURE 1. United-residue representation of a polypeptide chain. The interaction sites are side-chain centroids of different sizes (SC) and peptide-bond centers (p) indicated by dashed circles, where the α-carbon atoms (small empty circles) are introduced only to assist in defining the geometry. The virtual C^α—C^α 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}; variable ''bond angle,'' α_{SC_i}, formed by SC_i and the bisector of the angle defined by C^α_{i-1}, C^α_i, and C^α_{i+1}; and with a variable ''dihedral angle,'' β_{SC_i}, of counterclockwise rotation about the bisector, starting from the right side of the C^α_{i-1}, C^α_i, C^α_{i+1} frame.

by Nishikawa et al.[13] and Levitt,[1] and the torsional and electrostatic-interaction parameters were obtained by averaging the all-atom ECEPP/2[14,15] potential.[6] In spite of these shortcomings, the first-generation force field was able to predict successfully the three-dimensional structures of simple helical proteins, such as the avian pancreatic polypeptide (APP)[6] and galanin.[16] Thus, *de novo* prediction was achieved without the incorporation of correlation contributions to energy. However, as we will show in the Results section, one of the reasons for the success of the first generation force field in *de novo* prediction was that the C^{α}—C^{α}—C^{α} angles were fixed; this resulted in strengthening the electrostatic interactions in regular helical structures. Therefore, although it was able to reproduce the structures of the α-helical proteins, there were problems with β-sheet structures.

Although application of the first generation potential to *de novo* folding of proteins with more complicated structural motifs that included β-sheets distinguished the native-like structures from alternative structures as low-energy ones, the resulting best structures were too distorted with respect to the native structures and resembled molten globules rather than the native structures, even though they had essentially native-like packing of the side chains (A. Liwo, M. R. Pincus, R. J. Wawak, S. Rackovsky, and H. A. Scheraga, unpublished work). This deficiency of pairwise united-residue potentials was also found by Skolnick and coworkers,[3] who concluded that the deficiency of such united-residue potentials can be caused by insufficient stability of regular secondary structures.[3] Therefore, they introduced into their potential explicit terms that contribute extra negative energy for the presence of regular structures. Additional negative energy is contributed if, for example, residue i forms a hydrogen bond with residue j and, simultaneously, residue $i + 1$ forms a hydrogen bond with residue $j \pm 1$. Similarly, if side-chain i is at a contact distance with respect to side-chain j and side-chain $i + k$ is at a similar contact distance with respect to side-chain $j \pm k$, $1 \le k \le 3$, the system acquires an additional negative energy. Because these additional energy terms involve more than two interaction sites, they have been called *multibody* or *cooperative* terms. Such modification of the force field results in a markedly higher stability of regular structures and makes it useful for *de novo* folding simulations.[3]

In this article, we show that the multibody terms in united-residue potentials arise naturally from the fact that such potentials can be derived from all-atom potentials by averaging them over the ''less important'' degrees of freedom, *viz*., those that are associated with minor changes in conformation [e.g., the internal geometry of a side chain or the positioning of the peptide groups between consecutive α-carbons and those of the surrounding water molecules (the averaging over the corresponding water molecules is implicit, because of our use of data from the PDB)]. Furthermore, by taking advantage of our earlier developed dipole model of the peptide group,[5] we derive *analytical* expressions for the multibody terms in backbone hydrogen-bonding energy. Finally, we demonstrate the role of the multibody terms in the stabilization of the regular α-helical structure of terminally blocked nonadeca-alanine.

Theory

GENERAL EXPRESSION FOR AVERAGE FREE ENERGY

Let us consider a general physical system, whose energy, $E(\mathbf{x};\mathbf{y})$, depends on two kinds of variables: $\mathbf{x} = [x_1, x_2, \ldots, x_m]^T$ and $\mathbf{y} = [y_1, y_2, \ldots, x_n]^T$. The variables contained in the vector $\mathbf{x}$ will be termed *important* variables; they are associated with major changes in the conformation of the system (such as the changes of the global fold of the polypeptide chain). The variables contained in the vector $\mathbf{y}$ will be termed *less important*. We want to create a simplified representation of the system in which only the important variables are taken into account. This means that we have to average the energy of the system over the less important variables, $\mathbf{y}$. In general, this can be expressed by:

$$\mathscr{U}(\mathbf{x}) = \mathscr{F}^{-1}\left\{\frac{\int_{\mathbf{y}} w[E(\mathbf{x};\mathbf{y})]\mathscr{F}[E(\mathbf{x};\mathbf{y})]\,\mathrm{d}V_y}{\int_{\mathbf{y}} w[E(\mathbf{x};\mathbf{y})]\,\mathrm{d}V_y}\right\} \quad (2)$$

where $\mathscr{F}$ is a real monotonic function in one real variable, w is a weight function for the energy, and $\mathrm{d}V_y = \mathrm{d}y_1\mathrm{d}y_2 \cdots \mathrm{d}y_n$.

The form of the averaged energy $\mathscr{U}$ will depend on the choice of the transformation $\mathscr{F}$ and the weight function w. Usually, the energy is Boltzmann-averaged,[1,6] which means that $\mathscr{F}(E) = E$ and $w(E) = \exp(-E/k_BT)$, T being the absolute temperature and k_B the Boltzmann constant. However, because the local and two-body terms in our

united-residue potential can be identified with free energies rather than with Boltzmann-averaged energies,[9,10] in this work we choose $\mathscr{F}(E) = \exp(-E/k_BT)$ and $w(E) = 1$. In this case, $\mathscr{U}(\mathbf{x})$ has the meaning of the average free energy associated with a set of fixed values of the important variables $\mathbf{x}$, as expressed by:

$$\mathscr{U}(\mathbf{x}) = F(\mathbf{x}) = -k_BT \ln\left\{\frac{1}{V_y}\int_{\mathbf{y}} \exp[-E(\mathbf{x};\mathbf{y})/k_BT]\, \mathrm{dV}_y\right\} \quad (3)$$

with:

$$V_y = \int_y \mathrm{dV}_y$$

Utilizing the cumulant expansion of the free energy,[17,18] eq. (3) can be expressed as a power series in $\beta = 1/k_BT$:

$$\begin{aligned} F(\mathbf{x}) &\equiv F(\beta,\mathbf{x}) \\ &= U_1 - \frac{1}{2}(U_2 - U_1^2)\beta \\ &\quad + \frac{1}{6}(U_3 - 3U_1U_2 + 2U_1^3)\beta^2 \\ &\quad - \frac{1}{24}(U_4 - 3U_2^2 - 4U_1U_3 \\ &\quad + 12U_1^2U_2 - 6U_1^4)\beta^3 + \cdots \\ &= -\sum_{k=1}^{\infty} \frac{(-1)^k}{k!} C_k(\mathbf{x})\beta^{k-1} \end{aligned} \quad (4)$$

where:

$$U_k = \frac{1}{V_y}\int_{\mathbf{y}} E(\mathbf{x};\mathbf{y})^k\, \mathrm{dV}_y \quad (5)$$

is the kth moment of the energy about $E = 0$ and C_k is the kth cumulant.

A general formula for the kth cumulant $C_k(\mathbf{x})$ derived in this work is given by eq. (A-14) of Appendix 1. In this Appendix, we also show that the expansion given by eq. (4) is guaranteed to converge if $E(\mathbf{x};\mathbf{y})$ is a bounded function of $\mathbf{y}$ for the considered range of $\mathbf{x}$ [this holds for the energy function considered in this article, defined by eq. (6)] and $\beta < 1/2M(\mathbf{x})$, where $M(\mathbf{x})$ is the boundary of the absolute value of $E(\mathbf{x};\mathbf{y})$ over the whole range of $\mathbf{y}$.

The presence of the integrals of various powers of the energy gives rise to the appearance of multibody terms in the average energy of the system in a simplified representation. Consider, for example, three side chains being in a close contact with each other. The forces acting between the individual atoms of the side chains are, to a good approximation, pairwise. If we consider only the side-chain centers and average the interaction energy over the *internal* degrees of freedom of each side chain according to eq. (4), U_1 will still include only the interactions of the pairs of the side-chain centroids. However, U_2 will consist of the integrals of the products of the interaction energies between two pairs of the side chains; that is, the result will, in general, depend on the positions of *three* centroids.

Following these general considerations and taking advantage of the dipole model of peptide groups developed in our earlier work,[5] in the next subsection we derive analytical expressions for the multibody contributions to the interaction energy of the peptide groups.

CASE OF BACKBONE PEPTIDE-GROUP INTERACTION

Energy of interaction of peptide groups. Assume that the polypeptide chain contains n peptide groups. The interaction energy E_{ij} between a pair of peptide groups i and j depends on the distance r_{ij} between their centers, the relative orientation of the virtual-bond vectors $\mathbf{v}_i$ and $\mathbf{v}_j$ (pointing from C_i^α to C_{i+1}^α and from C_j^α to C_{j+1}^α, respectively; Fig. 2), and the rotation angles λ_i and λ_j of the peptide-group planes about the C^α—C^α virtual-bond axes. In our earlier work,[5] we derived the following approximate formula for this interaction energy, based on the assumption that each peptide group is represented by a point dipole located in the middle between the corresponding C^αs (Fig. 2):

$$\begin{aligned} E_{ij} &= \frac{p_i p_j \cos\mu_i \cos\mu_j}{\epsilon r_{ij}^3} W_{ij} - \frac{p_i p_j \sin(\mu_i + \mu_j)}{2\epsilon r_{ij}^3} \\ &\quad \times \left\{ Y_{ij}^{(1)} \cos(\lambda_i - \lambda_i^0) + Y_{ij}^{(2)} \cos\left(\lambda_j - \lambda_j^0\right) \right\} \\ &\quad + \frac{p_i p_j \sin\mu_i \sin\mu_j}{\epsilon r_{ij}^3} \\ &\quad \times \left\{ Z_{ij}^{(1)} \cos\left(\lambda_i + \lambda_j - \Lambda_{ij}^{(1)}\right) \right. \\ &\quad \left. + Z_{ij}^{(2)} \cos\left(\lambda_i - \lambda_j - \Lambda_{ij}^{(2)}\right) \right\} \end{aligned} \quad (6)$$

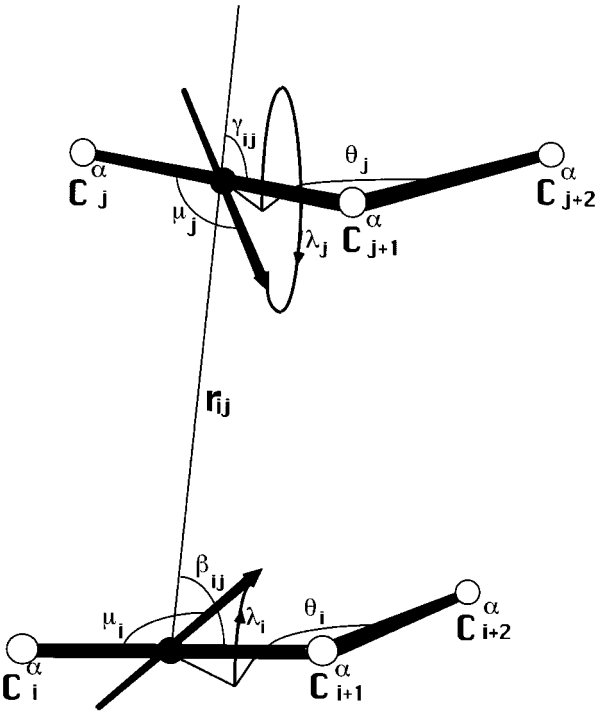

FIGURE 2. The relative orientation of the virtual bonds C_i^α—C_{i+1}^α and C_j^α—C_{j+1}^α is described by the angles α_{ij}, β_{ij}, and γ_{ij}, defined by eq. (8). The angle α_{ij} is not shown here because the two virtual bonds, C_i^α—C_{i+1}^α and C_j^α—C_{j+1}^α, are not necessarily coplanar. θ is the angle between two successive virtual bonds. The peptide-group dipole moments are represented by arrows (pointing from the carbonyl oxygen to the amide hydrogen of a peptide group), and the angles μ_i and μ_j between them and the virtual bonds are also shown, as well as the rotation angles λ_i and λ_j of the peptide-group dipoles. The two dipoles are separated by a distance r_{ij}.

where[5]:

$$
\begin{aligned}
W_{ij} &= \cos\alpha_{ij} - 3\cos\beta_{ij}\cos\gamma_{ij} \\
Y_{ij}^{(1)} &= \sqrt{1 + 3\cos^2\beta_{ij} - W_{ij}^2} \\
Y_{ij}^{(2)} &= \sqrt{1 + 3\cos^2\gamma_{ij} - W_{ij}^2} \\
Z_{ij}^{(1)} &= \frac{1}{2}\sqrt{4(1+\cos\alpha_{ij}) + W_{ij}^2 - 3(\cos\beta_{ij} + \cos\gamma_{ij})^2} \\
Z_{ij}^{(2)} &= \frac{1}{2}\sqrt{4(1-\cos\alpha_{ij}) + W_{ij}^2 - 3(\cos\beta_{ij} - \cos\gamma_{ij})^2}
\end{aligned}
\tag{7}
$$

$$
\begin{aligned}
\cos\alpha_{ij} &= \mathbf{v_i} \cdot \mathbf{v_j} \\
\cos\beta_{ij} &= \mathbf{v_i} \cdot \mathbf{e}_{\mathbf{r_{ij}}} \\
\cos\gamma_{ij} &= \mathbf{v_j} \cdot \mathbf{e}_{\mathbf{r_{ij}}}
\end{aligned}
\tag{8}
$$

$\mathbf{e}_{\mathbf{r_{ij}}}$ being the unit vector pointing from p_i to p_j (Fig. 2). The derivation of eq. (6) and the formulas for the constants λ_i^0, λ_j^0, $\Lambda_{ij}^{(1)}$, and $\Lambda_{ij}^{(2)}$ can be found in the Appendix of Liwo et al.[5]

The terms in eq. (6) correspond to the interaction between the components of the peptide-group dipoles parallel to the C^α—C^α virtual-bond axes, between the parallel and perpendicular components, and between the perpendicular components, respectively.

In earlier work,[5] we derived analytical expressions for the average energy of the peptide-group interaction, averaged over the rotation angles λ_i and λ_j, and identified with the average hydrogen-bond energy per pair of interacting peptide groups. The purpose of the present work is to derive analytical expressions for the higher order terms in the expression for the average free energy of a system of n interacting peptide groups; these terms can be identified with the cooperative hydrogen-bonding terms implemented by Skolnick and coworkers in their simulation of protein folding.[3]

To simplify further considerations, we will treat only the terms in eq. (6) corresponding to the interaction between the perpendicular components of the peptide-group dipoles, and neglect the others. We have shown that this part of the interaction energy is dominant.[5] We also assume that the distances between the peptide groups and the directions of the C^α—C^α vectors do not change with the λ values; for a polypeptide chain, the latter assumption is only an approximation. In this approximation, each E_{ij} depends only on λ_i and λ_j. Furthermore, based on eq. (A4) of the Appendix of Liwo et al.,[5] it can be shown that:

$$
\begin{aligned}
\Lambda_{ij}^{(1)} &= \lambda_i^0 + \lambda_j^0 \\
\Lambda_{ij}^{(2)} &= \lambda_i^0 - \lambda_j^0 + \pi
\end{aligned}
\tag{9}
$$

Thus, the interaction energy can be expressed approximately by:

$$
\begin{aligned}
E_{ij} &\approx e_{ij}(\lambda_i, \lambda_j) \\
&= \zeta_{ij}\cos(\lambda_i' + \lambda_j') - \tilde{\zeta}_{ij}\cos(\lambda_i' - \lambda_j')
\end{aligned}
\tag{10}
$$

with:

$$
\begin{aligned}
\lambda_i' &= \lambda_i - \lambda_i^0 \\
\lambda_j' &= \lambda_j - \lambda_j^0 \\
\zeta_{ij} &= \frac{\sin\mu_i \sin\mu_j}{\epsilon r_{ij}^3} Z_{ij}^{(1)} \\
\tilde{\zeta}_{ij} &= \frac{\sin\mu_i \sin\mu_j}{\epsilon r_{ij}^3} Z_{ij}^{(2)}
\end{aligned}
\tag{11}
$$

In further considerations, for the sake of clarity, we drop the ''prime'' superscripts from the λ values.

Cumulant expansion for the average free energy of interacting peptide-group dipoles. Substituting eq. (10) into eq. (4) we obtain eq. (12). We take advantage of the fact that, because the integral of each e_{ij} of eq. (10) over λ_i and λ_j will be identically zero (because $\int_0^{2\pi} \cos\alpha \, d\alpha = 0$), the first moment of the energy [U_1 in eq. (4)] is equal to zero.

$$\begin{aligned} F &= -\frac{1}{\beta}\ln\left\{\frac{1}{(2\pi)^n}\int_0^{2\pi}\cdots \right. \\ &\quad \left.\int_0^{2\pi}\exp\left[-\beta\sum_{j<i}^{n} e_{ij}(\lambda_i, \lambda_j)\right]\right\} d\lambda_1 \cdots d\lambda_n \\ &= -\frac{1}{\beta}\ln\left\{\frac{1}{(2\pi)^n}\right. \\ &\quad \left.\times\int_{[0,2\pi]^n}\exp\left[-\beta\sum_{j<i}^{n} e_{ij}(\lambda_i, \lambda_j)\right]\right\} d\Lambda \\ &= -\frac{1}{2}U_2\,\beta + \frac{1}{6}U_3\,\beta^2 - \left(\frac{1}{24}U_4 - \frac{1}{8}U_2^2\right)\beta^3 + \cdots \end{aligned} \tag{12}$$

where:

$$d\Lambda = d\lambda_1 \cdots d\lambda_n \tag{13}$$

and:

$$U_k = \frac{1}{(2\pi)^n}\int_{[0,2\pi]^n}\left[\sum_{j<i}^{n} e_{ij}(\lambda_i, \lambda_j)\right]^k d\Lambda. \tag{14}$$

By ordering the quantities e_{ij} in the following way:

$$e_{21}, e_{31}, \ldots, e_{n1}, e_{32}, e_{42}, \ldots, e_{n2}, \ldots, e_{n,n-1} \tag{15}$$

and by defining auxiliary quantities a_I, for $I = 1, 2, \ldots, N$, where $N = n(n-1)/2$, as the Ith function e in the sequence (15) (I is the so-called Cantor index[19] corresponding to the pair of integers ij, such that $j < i$), we can represent the formula in eq. (14) as:

$$\begin{aligned} U_k &= \frac{1}{(2\pi)^n}\int_{[0,2\pi]^n}\left(\sum_{I=1}^{N} a_I\right)^k d\Lambda \\ &= \frac{1}{(2\pi)^n}\int_{[0,2\pi]^n}\sum_{p_1+\cdots+p_N=k}\frac{k!}{p_1!\cdots p_n!} \\ &\quad \times a_1^{p_1}\cdots a_N^{p_N}\, d\Lambda \\ &= k!\sum_{p_1+\cdots+p_N=k}\frac{1}{(2\pi)^n}\int_{[0,2\pi]^n}\prod_{l=1}^{N}\frac{a_l^{p_l}}{p_l!}\, d\Lambda \\ &= k!\sum_{p_1+\cdots+p_N=k}\left(\prod_{l=1}^{N}\frac{1}{p_l!}\right)\frac{1}{(2\pi)^{n'}} \\ &\quad \times\int_{[0,2\pi]^{n'}}\prod_{\substack{l=1\\ p_l>0}}^{N} a_l^{p_l}\, d\Lambda'. \end{aligned} \tag{16}$$

The ''prime'' over n and over Λ denotes that the integration is carried out only over those λ's that occur in the function under the integral sign, for a given term in the summation over indices $p_1, \ldots, p_N$.

Substituting $e_{i(l)j(l)}(\lambda_{i(l)}, \lambda_{j(l)})$[$\{i(l), j(l)\}$ is a pair of indices corresponding to the Cantor index l] back for a_l in eq. (16), we obtain:

$$\begin{aligned} U_k &= k!\sum_{p_1+\cdots+p_N=k}\left(\prod_{l=1}^{N}\frac{1}{p_l!}\right)\frac{1}{(2\pi)^{n'}} \\ &\quad \times\int_{[0,2\pi]^{n'}}\prod_{\substack{l=1\\ p_l>0}}^{N} e_{i(l),j(l)}(\lambda_{i(l)}, \lambda_{j(l)})^{p_l}\, d\Lambda' \\ &= k!\sum_{p_1+\cdots+p_N=k}\left(\prod_{l=1}^{N}\frac{1}{p_l!}\right)\mathcal{I}_k(p_1, p_2, \ldots, p_N) \end{aligned} \tag{17}$$

where

$$\begin{aligned} &\mathcal{I}_k(p_1, p_2, \ldots, p_N) \\ &\quad = \frac{1}{(2\pi)^{n'}}\int_{[0,2\pi]^{n'}}\prod_{\substack{l=1\\ p_l>0}}^{N} e_{i(l),j(l)}(\lambda_{i(l)}, \lambda_{j(l)})^{p_l}\, d\Lambda'. \end{aligned} \tag{18}$$

The role of the terms in eq. (17) that are identical with cooperativity is discussed after eq. (36).

Evaluation of the integrals in the expression for the kth moment of the energy. Let us consider one of the integrals in eq. (17) that corresponds to a given set of exponents $p_1, \ldots, p_N$ [*viz.*, eq. (18)]. For clarity, we write $\mathcal{I}$ instead of $\mathcal{I}_k(p_1, p_2, \ldots, p_N)$ throughout the rest of this subsection.

By expanding the product in eq. (18) by using only the first powers of the function $e_{i(l)j(l)}$, a product of exactly k functions e is obtained [because $p_1 + \cdots + p_N = k$ in eq. (17)]. Obviously, now some of the e values will occur multiply in the expression:

$$\mathcal{I} = \frac{1}{(2\pi)^{n'}}\int_{[0,2\pi]^{n'}}\prod_{m=1}^{k} e_{i_m,j_m}(\lambda_{i_m}, \lambda_{j_m})\, d\Lambda'. \tag{19}$$

Based on eq. (10), the energy of interaction between peptide groups i_m and j_m can be expressed as follows:

$$\begin{aligned} & e_{i_m j_m}(\lambda_{i_m}, \lambda_{j_m}) \\ & \quad = \zeta_{i_m j_m} \cos(\lambda_{i_m} + \lambda_{j_m}) - \tilde{\zeta}_{i_m j_m} \cos(\lambda_{i_m} - \lambda_{j_m}) \\ & \quad = \frac{1}{2} \zeta_{i_m j_m} \Big\{ \exp\big[i(\lambda_{i_m} + \lambda_{j_m}) \big] \\ & \quad + \exp\big[-i(\lambda_{i_m} + \lambda_{j_m}) \big] \Big\} \\ & \qquad - \frac{1}{2} \tilde{\zeta}_{i_m j_m} \Big\{ \exp\big[i(\lambda_{i_m} - \lambda_{j_m}) \big] \\ & \quad + \exp\big[-i(\lambda_{i_m} - \lambda_{j_m}) \big] \Big\} \\ & \quad = \frac{1}{2} \sum_{s_m = \pm 1} \sum_{t_m = \pm 1} t_m \zeta_{i_m j_m; t_m} \\ & \qquad \times \exp\big[i s_m (\lambda_{i_m} + t_m \lambda_{j_m}) \big] \end{aligned} \tag{20}$$

where $\zeta_{i_m j_m; 1} = \zeta_{i_m j_m}$ and $\zeta_{i_m j_m; -1} = \tilde{\zeta}_{i_m j_m}$.

Substituting eq. (20) into eq. (19), we obtain:

$$\begin{aligned} \mathscr{I} &= \frac{1}{2^k} \frac{1}{(2\pi)^{n'}} \\ &\quad \times \int_{[0, 2\pi]^{n'}} \prod_{m=1}^{k} \Bigg\{ \sum_{s_m = \pm 1} \sum_{t_m = \pm 1} t_m \zeta_{i_m j_m; t_m} \\ &\quad \times \exp\big[i s_m (\lambda_{i_m} + t_m \lambda_{j_m}) \big] \Bigg\} \, \mathrm{d}\Lambda' \\ &= \frac{1}{2^k} \frac{1}{(2\pi)^{n'}} \\ &\quad \times \int_{[0, 2\pi]^{n'}} \sum_{s_1 = \pm 1} \sum_{t_1 = \pm 1} \cdots \sum_{s_k = \pm 1} \sum_{t_k = \pm 1} \\ &\quad \times \prod_{m=1}^{k} \Big\{ t_m \zeta_{i_m j_m; t_m} \exp\big[i s_m (\lambda_{i_m} + t_m \lambda_{j_m}) \big] \Big\} \, \mathrm{d}\Lambda' \\ &= \frac{1}{2^k} \frac{1}{(2\pi)^{n'}} \sum_{s_1 = \pm 1} \sum_{t_1 = \pm 1} \cdots \\ &\qquad \sum_{s_k = \pm 1} \sum_{t_k = \pm 1} \prod_{m=1}^{k} t_m \zeta_{i_m j_m; t_m} \\ &\quad \times \int_{[0, 2\pi]^{n'}} \exp\left[i \sum_{m=1}^{k} s_m (\lambda_{i_m} + t_m \lambda_{j_m}) \right] \mathrm{d}\Lambda' \end{aligned} \tag{21}$$

The integration in eq. (21) is carried out over n' of the variables $\lambda_1, \ldots, \lambda_n$ (see the definition of n' and $\mathrm{d}\Lambda'$ in the preceding subsection); these variables will be denoted as $\lambda_{h_1}, \ldots, \lambda_{h_{n'}}$. Hence:

$$\sum_{m=1}^{k} s_m (\lambda_{i_m} + t_m \lambda_{j_m}) = \sum_{\nu=1}^{n'} \lambda_{h_\nu} (A_\nu + B_\nu) \tag{22}$$

where:

$$A_\nu = \sum_{\substack{m=1 \\ i_m = \nu}}^{k} s_m, \quad B_\nu = \sum_{\substack{m=1 \\ j_m = \nu}}^{k} s_m t_m \quad \text{for } \nu = 1, 2, \ldots, n' \tag{23}$$

Consequently, the integrals in eq. (21) can be represented as the following products of one-dimensional integrals:

$$\prod_{\nu=1}^{n'} \int_0^{2\pi} \exp\big[i \lambda_{h_\nu} (A_\nu + B_\nu) \big] \, d\lambda_{h_\nu} \tag{24}$$

which is equal to $(2\pi)^{n'}$, if the following system of equations holds:

$$A_\nu + B_\nu = 0 \quad \text{for } \nu = 1, 2, \ldots, n' \tag{25}$$

and zero otherwise.

Hence, the integral $\mathscr{I}$ of eq. (18) is expressed by:

$$\begin{aligned} \mathscr{I} = \frac{1}{2^k} \sum_{s_1 = \pm 1} \sum_{t_1 = \pm 1} \cdots \sum_{s_k = \pm 1} \sum_{t_k = \pm 1} \Upsilon(i_1, j_1, \ldots, \\ i_k, j_k, s_1, t_1, \ldots, s_k, t_k) \end{aligned} \tag{26}$$

where:

$$\begin{aligned} & \Upsilon(i_1, j_1, \ldots, i_k, j_k, s_1, t_1, \ldots, s_k, t_k) \\ & \quad = \begin{cases} \prod_{m=1}^{k} t_m \zeta_{i_m j_m; t_m} & \text{if eqs. (25) hold,} \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{27}$$

From eq. (27) it is immediately obvious that the values of the nonvanishing integrals Υ depend only on $t_1, \ldots, t_k$, and not on $s_1, \ldots, s_k$. Given $t_1, \ldots, t_k$ ($t_m = \pm 1$), the number of equal integrals Υ is the same as the number of sequences $s_1, \ldots, s_k$ ($s_m = \pm 1$) that, together with $t_1, \ldots, t_k$, satisfy eqs. (25). We denote such integrals by $\tilde{\Upsilon}(i_1, j_1, \ldots, i_k, j_k, t_1, \ldots, t_k)$. Thus, eq. (26) becomes:

$$\begin{aligned} \mathscr{I} = \frac{1}{2^k} \sum_{t_1 = \pm 1} \cdots \sum_{t_k = \pm 1} \gamma(i_1, j_1, \ldots, i_k, j_k, t_1, \ldots, t_k) \\ \times \tilde{\Upsilon}(i_1, j_1, \ldots, i_k, j_k, t_1, \ldots, t_k) \end{aligned} \tag{28}$$

where $\gamma(i_1, j_1, \ldots, i_k, j_k, t_1, \ldots, t_k)$ is the number of solutions of eqs. (25), given the numbers $t_1, \ldots, t_k$ (this number is equal to zero, if there are no solutions for a particular set of numbers $t_1, \ldots, t_k$).

The problem of the computation of the integrals $\mathcal{I}$ of eq. (17) has therefore been reduced to the problem of finding all possible solutions of eqs. (25), given a set of exponents $p_1, p_2, \ldots, p_N$. It can be seen immediately that, in eq. (23), the number of terms in the sum to calculate A_ν equals the number of occurrences of the variable λ_{h_ν} as the first variable in the functions e of eq. (19); similarly, the number of terms in the sum to calculate B_ν equals the number of occurrences of the variable λ_{h_ν} as the second variable. Because the terms in both summations of eq. (23) equal ± 1, it is necessary that the total number of occurrences of λ_{h_ν} as a variable (first or second) in functions e of eq. (19) be even, to satisfy the appropriate equation of the system of equations of eq. (25) for this value of ν. If this condition is not fulfilled, there are no solutions of eqs. (25) and, therefore, all of the integrals $\mathcal{I}$ corresponding to the set of exponents under consideration are identically equal to zero. In particular, this implies that $U_1 = 0$, as mentioned in the preceding subsection.

The integrals $\mathcal{I}$ can be represented as graphs, in a similar way to the Mayer theory of imperfect gases.[20,21] Any choice of the nonnegative integer numbers $p_1, \ldots, p_N$ of eq. (17) defines a set of peptide groups which appear in the function at the integration symbol; these peptide groups have the indices $i(l)$ or $j(l)$, where $l = 1, 2, \ldots, N$, and $p_l > 0$. The peptide groups can be represented as vertices of a graph, and each factor of the product at the integration symbol can be represented as an edge connecting the vertices $i(l)$ and $j(l)$. One edge connecting two vertices i and j indicates that e_{ij} occurs only once as a factor in eq. (19). The number of edges connected to vertex i indicates the number of factors of eq. (19) in which λ_i occurs. For the integrals corresponding to the second (U_2), third (U_3), and fourth (U_4) moment of the energy, this is illustrated in Figure 3. From the above considerations it follows that any graph in which there is even just one vertex with an odd number of connections (edges) corresponds to a zero integral $\mathcal{I}$. Similarly, the integrals $\tilde{\Upsilon}$ of eq. (28) can be represented as graphs with edges labeled differently, depending on whether they represent a ζ or a $\tilde{\zeta}$ (see Fig. 4 for illustration). The graphical representation of the integrals $\mathcal{I}$ can also be used to find all nonvanishing integrals $\tilde{\Upsilon}$ and the numbers of their occurrences, γ, in the expression for a given integral $\mathcal{I}$ [eq. (28)]. The corresponding algorithm is presented in Appendix 2.

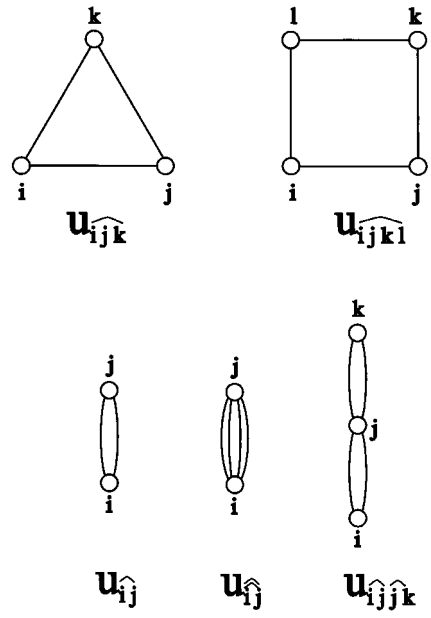

FIGURE 3. Graphical representations of the integrals of eq. (29). Circles represent interacting peptide-group centers. If peptide groups i and j are connected by a single line, the term e_{ij} [defined by eq. (10)] occurs once in eq. (30) for the respective u; if they are connected by two lines, e_{ij} occurs twice, and so on. Owing to the form of e_{ij}, only graphs in which each vertex has an even number of edges correspond to nonzero values of the integrals. As an example, $u_{\widehat{ij}\widehat{jk}}$ and $u_{\widehat{ijkl}}$ represent the correlation energy in a hydrogen-bonded chain and in a cluster, respectively.

Interpretation of the constituents of the three- and four-body terms in the cumulant expansion for the average free energy. For the energy moments U_2, U_3, and U_4, eqs. (17) and (28), together with the graphical approach presented in Appendix 2, lead to the following expressions*:

$$U_2 = \sum_{j<i} u_{\widehat{ij}}$$

$$U_3 = 6 \sum_{k<j<i} u_{\widehat{ijk}}$$

$$U_4 = \sum_{j<i} u_{\widehat{\widehat{ij}}} + 6 \sum_{l<k<j<i} (u_{\widehat{ij}}u_{\widehat{kl}} + u_{\widehat{ik}}u_{\widehat{jl}} + u_{\widehat{il}}u_{\widehat{jk}}) + 6 \sum_{k<j<i} (u_{\widehat{\widehat{ijik}}} + u_{\widehat{\widehat{ijjk}}} + u_{\widehat{\widehat{ikjk}}}) + 24 \sum_{l<k<j<i} (u_{\widehat{ijkl}} + u_{\widehat{ikjl}} + u_{\widehat{ijlk}}) \quad (29)$$

*The derivation of eqs. (29)–(36) was also carried out with the aid of the *Maple* symbolic algebra package.[22]

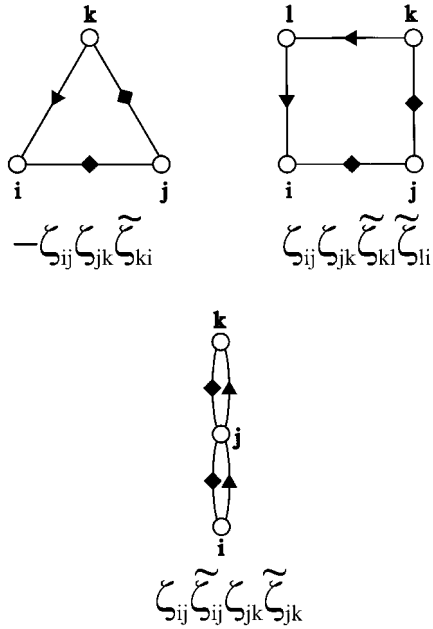

FIGURE 4. Graphical representation of the integrals $\tilde{\Upsilon}$ of the expressions for $u_{\widehat{ijk}}$ and $u_{\widehat{ijkl}}$ (with all four indices different), and $u_{\widehat{ijjk}}$. Circles (vertices of the graphs) represent peptide-group centers. A diamond on edge *ij* indicates that its contribution to the product is ζ_{ij}, whereas a triangle indicates that its contribution is $-\tilde{\zeta}_{ij}$.

The various u values, which are equal to the appropriate integrals $\mathcal{I}$ of eq. (17), are defined by the following equations:

$$u_{\widehat{ij}} = \frac{1}{(2\pi)^2}\int_0^{2\pi}\int_0^{2\pi} e_{ij}^2 \,\mathrm{d}\lambda_i \,\mathrm{d}\lambda_j$$

$$u_{\widehat{\widehat{ij}}} = \frac{1}{(2\pi)^2}\int_0^{2\pi}\int_0^{2\pi} e_{ij}^4 \,\mathrm{d}\lambda_i \,\mathrm{d}\lambda_j$$

$$u_{\underbrace{\widehat{ijk\cdots pq}}_{m \text{ indices}}} = \frac{1}{(2\pi)^m}\int_0^{2\pi}\cdots\int_0^{2\pi} e_{ij}e_{jk}\cdots e_{pq} \times e_{qi}\,\mathrm{d}\lambda_i\cdots\mathrm{d}\lambda_q$$

$$u_{\underbrace{\widehat{ij\cdots pqqr\cdots st}}_{m \text{ indices}}} = \frac{1}{(2\pi)^m}\int_0^{2\pi}\cdots\int_0^{2\pi} e_{ij}\cdots e_{pq} \times e_{qi}e_{qr}\cdots e_{st}e_{tr}\,\mathrm{d}\lambda_i\cdots\mathrm{d}\lambda_t \quad (30)$$

In particular, the constituents of the average energy of peptide-group interaction, u_{ij}, are expressed by:

$$u_{\widehat{ij}} = \frac{1}{2}\left[(\zeta_{ij})^2 + \left(\tilde{\zeta}_{ij}\right)^2\right] \quad (31)$$

where ζ_{ij} and $\tilde{\zeta}_{ij}$ are defined by eq. (11). Eq. (31) is equivalent to the second term in Eq. (A11) of Liwo et al.[5] for the average energy of a peptide-group interaction.

The other integrals that occur in eq. (29) are expressed as follows:

$$u_{\widehat{ijk}} = -\frac{1}{4}\Big(\tilde{\zeta}_{ij}\tilde{\zeta}_{ik}\tilde{\zeta}_{jk} + \tilde{\zeta}_{ij}\zeta_{ik}\zeta_{jk} + \zeta_{ij}\tilde{\zeta}_{ik}\zeta_{jk} + \zeta_{ij}\zeta_{ik}\tilde{\zeta}_{jk}\Big) \quad (32)$$

$$u_{\widehat{\widehat{ij}}} = \frac{3}{8}\left[(\zeta_{ij})^4 + \left(\tilde{\zeta}_{ij}\right)^4 + 4\left(\zeta_{ij}\tilde{\zeta}_{ij}\right)^2\right] = \frac{3}{2}u_{\widehat{ij}}^2 + \frac{3}{4}\left(\zeta_{ij}\tilde{\zeta}_{ij}\right)^2 \quad (33)$$

$$u_{\widehat{ijjk}} = \frac{1}{4}\left[(\zeta_{ij}\zeta_{jk})^2 + \left(\tilde{\zeta}_{ij}\tilde{\zeta}_{jk}\right)^2 + \left(\zeta_{ij}\tilde{\zeta}_{jk}\right)^2 + \left(\tilde{\zeta}_{ij}\zeta_{jk}\right)^2 + \left(\tilde{\zeta}_{ij}\tilde{\zeta}_{jk}\right)^2 + 2\zeta_{ij}\tilde{\zeta}_{ij}\zeta_{jk}\tilde{\zeta}_{jk}\right] = u_{\widehat{ij}}u_{\widehat{jk}} + \frac{1}{2}\zeta_{ij}\tilde{\zeta}_{ij}\zeta_{jk}\tilde{\zeta}_{jk} \quad (34)$$

$$u_{\widehat{ijkl}} = \frac{1}{8}\Big(\zeta_{ij}\zeta_{jk}\zeta_{kl}\zeta_{li} + \zeta_{ij}\zeta_{jk}\tilde{\zeta}_{kl}\tilde{\zeta}_{li} + \zeta_{ij}\tilde{\zeta}_{jk}\zeta_{kl}\tilde{\zeta}_{li} + \zeta_{ij}\tilde{\zeta}_{jk}\tilde{\zeta}_{kl}\zeta_{li} + \tilde{\zeta}_{ij}\zeta_{jk}\zeta_{kl}\tilde{\zeta}_{li} + \tilde{\zeta}_{ij}\zeta_{jk}\tilde{\zeta}_{kl}\zeta_{li} + \tilde{\zeta}_{ij}\tilde{\zeta}_{jk}\zeta_{kl}\zeta_{li} + \tilde{\zeta}_{ij}\tilde{\zeta}_{jk}\tilde{\zeta}_{kl}\tilde{\zeta}_{li}\Big) \quad (35)$$

Finally, the cumulant expansion for the average free energy [eq. (12)] up to fourth order can be expressed by:

$$F = -\frac{1}{2}\beta\sum_{j<i}u_{\widehat{ij}} + \beta^2\sum_{k<j<i}u_{\widehat{ijk}} - \beta^3\left[-\frac{1}{32}\sum_{j<i}v_{ij} + \frac{1}{8}\sum_{k<j<i}(w_{ijik} + w_{ijjk} + w_{ikjk}) + \sum_{l<k<j<i}(u_{\widehat{ijkl}} + u_{\widehat{ikjl}} + u_{\widehat{ijlk}})\right] + \cdots \quad (36)$$

with:

$$v_{ij} = (\zeta_{ij})^4 + \left(\tilde{\zeta}_{ij}\right)^4 \quad (37)$$

$$w_{ijkl} = \zeta_{ij}\tilde{\zeta}_{ij}\zeta_{kl}\tilde{\zeta}_{kl} \quad (38)$$

The components of the third-order terms $u_{\widehat{ijk}}$ can be interpreted as three-body contributions to the average free energy of interaction of the system. The fourth-order terms consist of both two-body terms, v_{ij}, three-body correlation terms, w_{ijjk}, w_{ijik}, and w_{ikjk}, as well as four-body correlation terms contained in $u_{\widehat{ijkl}}$. The terms v_{ij} are only pairwise terms, such as the terms $u_{\widehat{ij}}$ that constitute the mean-field potential of peptide-group interaction. The terms inherent in $u_{\widehat{ijk}}$ and $u_{\widehat{ijkl}}$ correspond to three- and four-body interactions within clusters of peptide groups close in space. They can be regarded as correlation contributions to the average free energy pertinent to clusters of three or four simultaneous hydrogen bonds (Figs. 3 and 4). The terms such as w_{ijjk} can be represented as linear graphs of three vertices linked sequentially by double edges (Fig. 4). These contributions to the interaction energy should be responsible for propagation of a chain of neighboring hydrogen bonds —such as those that link the peptide groups in α-helices or the peptide groups of the neighboring strands of β-sheets. There is an analogy between these correlation interactions and the concept of dipole paths that link chains of peptide groups close in space (but **not** sequential in the polypeptide chain).

We now take a closer look at the constituents of the four-body correlation terms. Clearly, these terms will rarely be appreciable if all four peptide groups belong to different parts of the chain, because clusters of four peptide groups at hydrogen-bonding contacts rarely occur in protein structures. However, if we assume that two pairs of peptide groups are adjacent, for example, $j = i - 1$ and $l = k - 1$, peptide group i is always in contact with peptide group j and peptide group k is always in contact with peptide group l. Because neither the distance nor the orientation of the adjacent peptide groups changes substantially, the factors $\zeta_{i,i-1}$, $\tilde{\zeta}_{i,i-1}$, $\zeta_{j,j-1}$, and $\tilde{\zeta}_{j,j-1}$ can be assumed to be constant. Taking the approximate mean value of the virtual-bond–valence angle between three consecutive C^{α}'s as $\theta = 90°$, we find that $\zeta_{i,i-1} = 5\tilde{\zeta}_{i,i-1}$. Using this relation, we can express the four-body correlation terms $u_{\widehat{ijkl}}$ and $u_{\widehat{ijlk}}$ of eq. (29) by eqs. (39) and (40), respectively:

$$\begin{aligned} u_{\widehat{ijkl}} &= u_{\widehat{i,i-1,k,k-1}} \\ &= \frac{z^2}{8}\Big[26\big(\zeta_{i-1,k}\zeta_{i,k-1} + \tilde{\zeta}_{i-1,k}\tilde{\zeta}_{i,k-1}\big) \\ &\quad + 10\big(\zeta_{i-1,k}\tilde{\zeta}_{i,k-1} + \tilde{\zeta}_{i-1,k}\zeta_{i,k-1}\big)\Big] \\ &= \frac{z^2}{4}\Big[9\big(\zeta_{i-1,k} + \tilde{\zeta}_{i-1,k}\big)\big(\zeta_{i,k-1} + \tilde{\zeta}_{i,k-1}\big) \\ &\quad + 4\big(\zeta_{i-1,k} - \tilde{\zeta}_{i,k-1}\big)\big(\zeta_{i-1,k} - \tilde{\zeta}_{i,k-1}\big)\Big] \quad (39) \end{aligned}$$

$$\begin{aligned} u_{\widehat{ijlk}} &= u_{\widehat{i,i-1,k-1,k}} \\ &= \frac{z^2}{4}\Big[9\big(\zeta_{i-1,k-1} + \tilde{\zeta}_{i-1,k-1}\big)\big(\zeta_{ik} + \tilde{\zeta}_{ik}\big) \\ &\quad + 4\big(\zeta_{i-1,k-1} - \tilde{\zeta}_{i-1,k-1}\big)\big(\zeta_{i,k} - \tilde{\zeta}_{i,k}\big)\Big] \quad (40) \end{aligned}$$

with $z = \tilde{\zeta}_{i,i-1}$ (within the above approximation this quantity depends only on the kind of peptide group, i.e., proline or nonproline[6]).

Eqs. (39) and (40) are equivalent to the cooperative terms in hydrogen-bonding interactions implemented by Skolnick and coworkers.[3] Eq. (39) describes the hydrogen-bonding correlation terms, pertinent to antiparallel β-sheets, whereas eq. (40) describes those pertinent to parallel β-sheets or helices. The expressions $(\zeta_{\iota\kappa} + \tilde{\zeta}_{\iota\kappa})$ and $(\zeta_{\iota\kappa} - \tilde{\zeta}_{\iota\kappa})$, respectively (where $\iota = i$ or $i - 1$ and $\kappa = k$ or $k - 1$), correspond to the maximum and minimum absolute values of the electrostatic energy of interaction between the rotatable parts of the dipoles of the peptide groups ι and κ. Therefore, $u_{\widehat{i,i-1,k,k-1}}$ and $u_{\widehat{i,i-1,k-1,k}}$ can be considered as the weighted average of the products of the electrostatic (and thereby hydrogen-bonding) energy between peptide-group pairs $i \cdots k$ and $i - 1 \cdots k - 1$ or $i \cdots k - 1$ and $i - 1 \cdots k$, where the three dots indicate a hydrogen bond; the energy will become more negative when *both* pairs are at hydrogen-bonding contact and orientation. The two terms of the sum in eq. (39) and (40) are illustrated in Figure 5. The first component of the sum corresponds to simultaneously aligned dipoles of the peptide groups (Fig. 5a and c). The second term corresponds to antiparallel location of the dipoles in each pair (Fig. 5b and d).

The three-body correlation terms $u_{\widehat{ijk}}$ can be treated likewise, assuming that two of the three peptide groups are consecutive in the polypeptide chain. These terms describe the cooperativity between the hydrogen-bonded pairs that share a

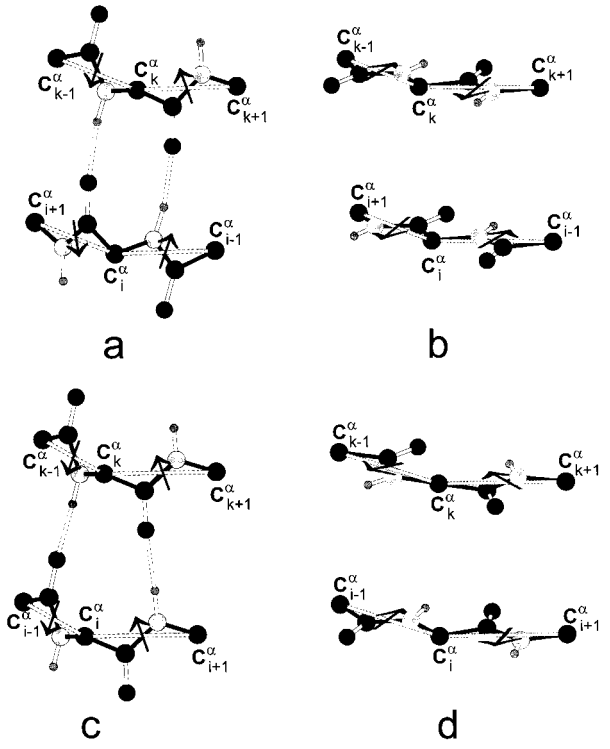

FIGURE 5. Illustration of the representative two terms of the four-body contribution to the average free energy of hydrogen-bonding interaction in antiparallel β-sheets [(a) and (b); eq. (39)] and parallel β-sheets [(c) and (d); eq. (40)].

common peptide group:

$$u_{\widehat{i,i-1,k}} = -\frac{z}{4}\left[3\left(\zeta_{ik} + \tilde{\zeta}_{ik}\right)\left(\zeta_{i-1,k} + \tilde{\zeta}_{i-1,k}\right) - 2\left(\zeta_{ik} - \tilde{\zeta}_{ik}\right)\left(\zeta_{i-1,k} - \tilde{\zeta}_{i-1,k}\right)\right] \quad (41)$$

Tests of Multibody Terms on Ac—(Ala)$_{19}$—NHMe

To demonstrate the role of the correlation in the modified potential, we carried out a search of the conformational space of Ac—(Ala)$_{19}$—NHMe using the united-residue model. The lowest energy conformation of this system in the all-atom ECEPP/3 force field[14,15,23] is a full α-helix, as found by using many global-optimization methods; for example, the self-consistent electrostatic field (SCEF) method,[24] the electrostatically driven Monte Carlo (EDMC) method,[25] and the self-consistent mean field theory (SCMFT) method.[26] These all-atom studies, however, were carried out without including hydration, whereas our united-residue force field implicitly includes hydration in the long-range SC—SC potential. Therefore, to identify the lowest energy conformation of Ac—(Ala)$_{19}$—NHMe in the ECEPP/3 force field including hydration, we have carried out several EDMC simulations of this polypeptide using the ECEPP/3 force field,[23] supplemented with the hydration contribution, calculated by using a solvent-accessible surface area model SRFOPT.[27] The resulting lowest-energy conformation consists of two antiparallel α-helical sections packed against each other, which gives rise to its stabilization, because of favorable hydrophobic interactions. Qualitatively, the same conclusion about the stabilization of a packed α-helical arrangement was drawn by Silverman and Scheraga[28] in their systematic study of the energetics of helix-packing interactions of polyalanine chains. A full α-helical conformation has an energy that is 3.2 kcal/mol higher than the lowest energy conformation. A full α-helix was the most frequently occurring, but not the lowest energy conformation during the EDMC runs. The lowest energy and the full α-helical conformation are shown in Figure 6. Given the small energy difference between the lowest energy and full-helical conformations and the uncertainty inherent in any empirical force field, we cannot draw a definite conclusion regarding the stability of a full-helical versus the ''two-helix-bundle'' conformation. Nevertheless, it can be stated that the Ala$_{19}$ chain in water consists of well-defined α-helical sections.

In the united-residue simulations with the potential function derived here and in refs. 9 and 10, we used the Monte Carlo minimization (MCM) method[29,30] to search the conformational space at a ''temperature''† of 500 K. We used only the four-body correlation contributions $u_{\widehat{i,i-1,k,k\pm1}}$ of eqs. (39) and (40), respectively, which are similar to the hydrogen-bonding correlation energy introduced by Skolnick et al.[3] Each calculation was terminated after 500 MCM iterations. The parameters and the weights of the energy terms of eq. (1) were taken as determined in our earlier work[9,10] by using the results of *inverse-folding* calculations, except for the weight of the electrostatic term, which was assigned an excessive value of $w_{el} = 1.5$.

†As opposed to canonical Monte Carlo sampling of the energy landscape, in the MCM method the temperature is a purely abstract parameter which only defines the probability of accepting energy minima with a higher energy than the current one.

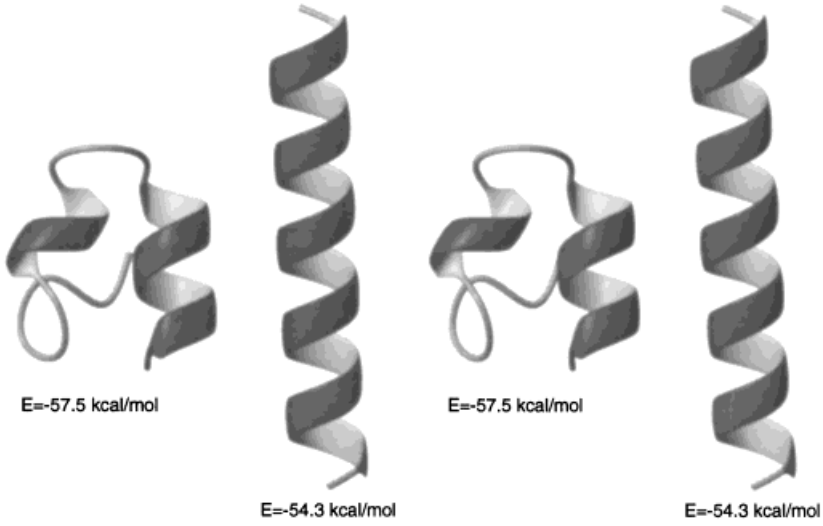

FIGURE 6. Stereoviews of the lowest energy conformation of the Ac—Ala_{19}—NHMe chain obtained in all-atom EDMC simulations, with the ECEPP / 3 force field[23] with an SRFOPT solvation contribution,[27] and the corresponding all-helical conformation of the all-atom chain.

All calculations were started from random conformations generated subject to the condition of no steric overlaps. To speed up the calculations, a cut-off value of 10 Å was assigned to the distances between the peptide groups in the two correlated pairs in eqs. (39) and (40). To assure a smooth transition between contact distances and distances greater than 10 Å, the correlation energy was multiplied by a quintic spline of the form given by eq. (42). With this cut-off, the computation time increased by about 30%, compared to that with the force field with only pair interaction-energy terms:

$$f(x) = \begin{cases} 1 & \text{for } x < -1 \\ 0.5 - 0.9375x + 0.625x^3 - 0.1875x^5 & \\ \quad \text{for } -1 \le x \le 1 & \\ 0 & \text{for } x > 1 \end{cases} \tag{42}$$

with:

$$x = \frac{r - r_{cut}}{\delta}, \quad \delta = 0.1 \text{ Å}$$

The results of the MCM simulations with different weights of the correlation term are presented in Figure 7. Similar results were obtained with the use of combined three- and four-body correlation contributions, respectively. As shown, without the hydrogen-bonding correlation terms, the lowest energy conformations are far from helical [with large RMS deviations from the helical structure and a small fraction of helical contacts (Fig. 7a); see also Fig. 8 for the lowest energy conformation]. From our all-atom simulations of the polyalanine chain of this size, it follows that the structure should contain well-defined helical sections. The fraction of helical contacts in the lowest energy conformations increases after introducing some cooperativity (Fig. 7b) and, with a greater weight of the cooperative term, the full α-helical conformation is the dominant one throughout the MCM run (Fig. 7c).

Figure 8 illustrates the change in the shape of the lowest energy conformation from a conformation almost devoid of secondary structure ($w_{corr} = 0$) through two hydrophobically packed helices ($w_{corr} = 0.3$ and $w_{corr} = 0.4$) to a full α-helix ($w_{corr} > 0.5$). The conformations obtained with $w_{corr} = 0.3$ and $w_{corr} = 0.4$ are qualitatively similar to the two-helix-bundle conformation obtained in the all-atom EDMC simulations shown in Figure 6.

Figure 7d shows that increasing the weight of the electrostatic term (keeping $w_{corr} = 0$) can also produce helical conformations as the lowest energy ones. However, in this case, a π-helical conformation (with virtual-bond angles $\theta = 100°$ and virtual-bond-dihedral angles $\gamma = 25°$) appears, with almost the same energy as the α-helical conformations. (The difference in the ECEPP/3 energy between the α-helical and π-helical Ac—Ala_{19}—NHMe is about 20 kcal/mol and π-helices are only very rarely encountered in pro-

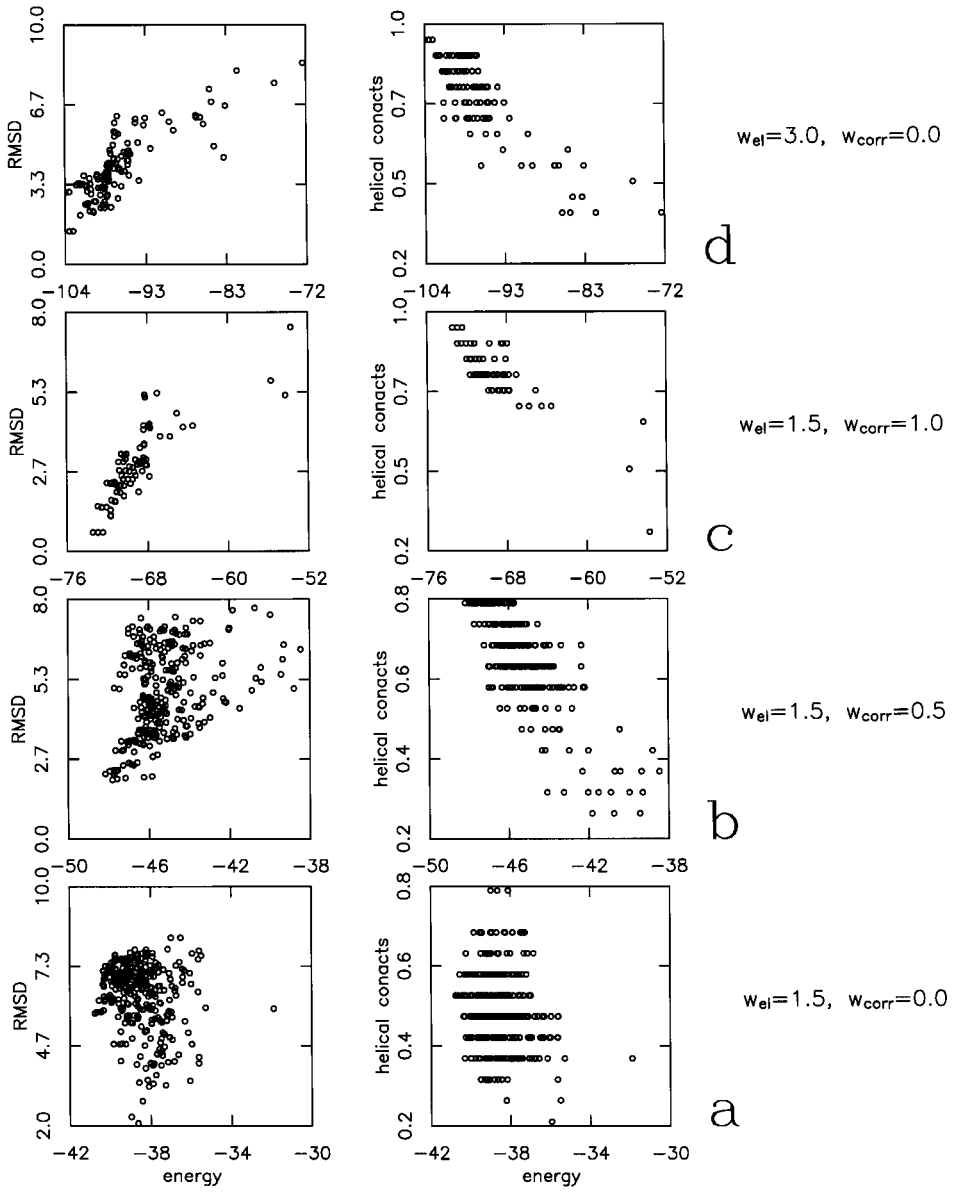

FIGURE 7. Diagrams of RMS deviation (Å) from the α-carbon trace of an ECEPP / 3-minimized Ac—Ala_{19}—NHMe chain, and the number of helical contacts, versus the total energy (kcal / mol) for the conformations obtained in the MCM runs with different weights of the correlation and electrostatic term.

teins[31]). In addition, overweighting the electrostatic term largely biases the resulting lowest energy structures toward helices.

It should be noted that, with our earlier force field,[5,6] which contains only pairwise terms, we were able to obtain partially or wholly α-helical structures as global-minimum conformations for helical proteins, such as the avian pancreatic polypeptide,[6] galanin,[16] and polyalanine (A. Liwo, M. R. Pincus, R. J. Wawak, S. Rackovsky, H. A. Scheraga, unpublished results). By contrast, our new force field apparently requires hydrogen-bonding cooperativity for the helix to form, even though the electrostatic-term weight was exaggerated with respect to the value determined using the threading-with-minimization calculations.[10] The reason for this is that, in our older force field,[5,6] the virtual-bond angles θ (Fig. 1) are fixed at 90°, which is approximately the value characteristic of α-helices. This causes the average distance

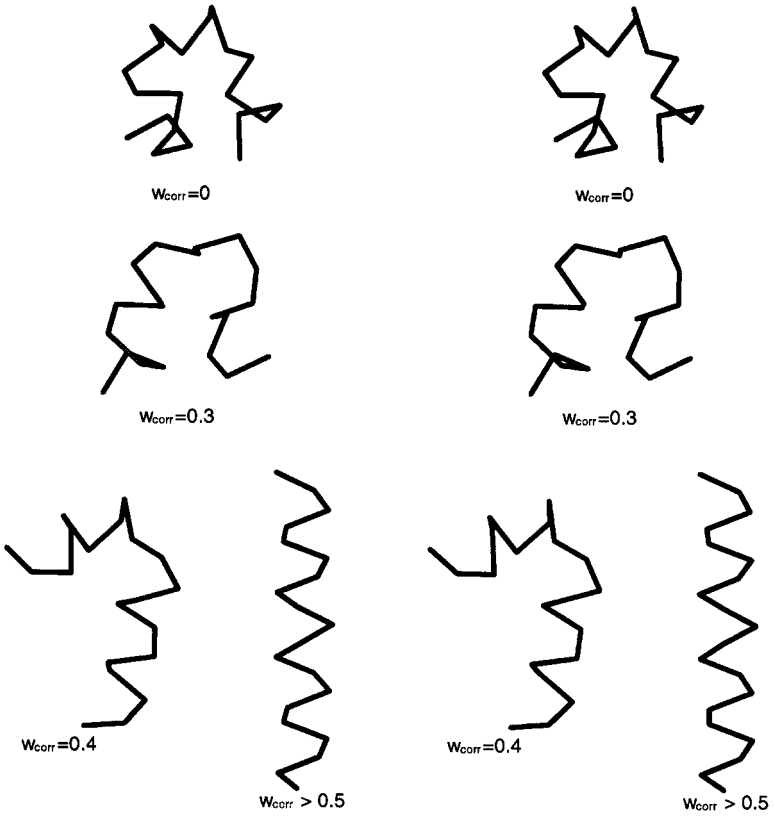

FIGURE 8. Effect of increasing the weight of the correlation term on the lowest energy conformation of Ac—Ala_{19}—NHMe in the united-residue representation. The weight of the average electrostatic contribution to the energy is $w_{el} = 1.5$ in all cases.

between the second- and third-neighbor peptide groups to be smaller than if the virtual-bond dihedral angles are variable; hence, the relative strength of the electrostatic interactions between the second- and third-neighbor peptide groups (as compared to that of the hydrophobic interactions between the side chains) increases. This, in turn, causes these peptide groups to assume a nearly parallel orientation forced by the directional electrostatic term in eq. (1) (see also the discussion of this term in ref. 6), which is most effectively realized in helices. When variable virtual-bond valence angles are used with our older force field, polyalanine does not fold to an α-helix, as occurs with the present force field *without* cooperative terms.

Despite the fact that regular helical structures can be obtained without cooperative terms, simply by fixing the values of the virtual-bond angles, it is certainly much more realistic to let these angles vary because, for real chains, they occur[13] in the range of $-47°$ to 145°. With fixed virtual-bond angles, the resulting united-residue structures are too distorted compared to real structures, unless the helix content in a protein is high (A. Liwo, M. R. Pincus, R. J. Wawak, S. Rackovsky, and H. A. Scheraga, unpublished work).

Implementation of Derived Correlation Terms in Force Field for *De Novo* Folding

The illustrated polyalanine calculations show that the hydrogen-bonding correlation terms derived in this work can be used directly in *de novo* folding with the united-residue force field described in earlier studies of this series.[9,10] Apart from the terms $u_{\widehat{i,i-1,k,k\pm1}}$ [eqs. (39) and (40)] implemented in this work, it might appear advantageous to consider also the correlation forces in ''chains'' of hydrogen bonds, which are expressed by contributions contained in $u_{\widehat{ijjk}}$ [eq. (34) and Fig. 5], although it seems that the force field can work without the latter contributions.

It is still necessary to obtain expressions for the correlation terms in side-chain (hydrophobic–hydrophilic) interactions and to determine the weights of the energy terms in eq. (1). Work on these problems is now underway in our laboratory.

Acknowledgments

The computations were carried out with one processor of the IBM-SP2 computer at the Cornell National Supercomputer Facility, a resource of the Center for Theory and Simulation in Science and Engineering at Cornell University, which is funded by the National Science Foundation, New York State, the IBM Corporation, and members of its Corporate Research Institute, with additional Research Resource funds from the National Institutes of Health. Work was also done using the IBM-SP2 computer at the Informatics Center of the Metropolitan Academic Network (IC MAN) at the Technical University of Gdańsk.

Appendix 1: Derivation of Formula for *n*th Order Cumulant and Convergence of Cumulant Expansion

Consider the following cumulant-generating function:

$$f(\beta) = -\beta F(\beta) = \ln\left\{\frac{1}{V_y}\int\cdots\int \exp[-\beta E(\mathbf{y})]\,\mathrm{d}V_y\right\} \quad \text{(A-1)}$$

where F is defined by eq. (4), and the variables contained in $\mathbf{x}$ are omitted for brevity.

We want to obtain the formulas for the coefficients c_k of the expansion of $f(\beta)$ in the McLaurin series:

$$f(\beta) = \sum_{k=1}^{\infty} c_k \beta^k \quad \text{(A-2)}$$

with the following relationship between the coefficients c_k and cumulants C_k:

$$c_k = \frac{(-1)^k}{k!} C_k \quad \text{(A-3)}$$

To obtain the formulas for c_k and, thereby, C_k, we proceed as follows. By expanding the argument of the logarithm in eq. (A-1) in the McLaurin series in powers of β, we obtain:

$$f(\beta) = \ln\left(\sum_{k=0}^{\infty} \epsilon_k \beta^k\right) \quad \text{(A-4)}$$

with:

$$\epsilon_k = \frac{(-1)^k}{k!}\frac{1}{V_y}\int\cdots\int E^k\,\mathrm{d}V_y = \frac{(-1)^k}{k!}U_k \quad \text{(A-5)}$$

Differentiation of eq. (A-2) with respect to β yields eq. (A-6):

$$\frac{df(\beta)}{d\beta} = \sum_{k=0}^{\infty}(k+1)c_{k+1}\beta^k = \sum_{k=0}^{\infty}\tilde{c}_k\beta^k \quad \text{(A-6)}$$

with:

$$c_k = \frac{\tilde{c}_{k-1}}{k} \quad \text{(A-7)}$$

To obtain an equivalent expression for $df(\beta)/d\beta$, but with known coefficients of a series expansion, we differentiate eq. (A-4) with respect to β, and obtain:

$$\frac{df(\beta)}{d\beta} = \frac{\sum_{k=0}^{\infty}(k+1)\epsilon_{k+1}\beta^k}{\sum_{k=0}^{\infty}\epsilon_k\beta^k} \quad \text{(A-8)}$$

Equating the right sides of eqs. (A-6) and (A-8), multiplying both sides by the denominator of eq. (A-8), and grouping the terms with the same powers of β on the left side, we obtain:

$$\sum_{k=0}^{\infty}\left(\sum_{l=0}^{k}\epsilon_{k-l}\tilde{c}_l\right)\beta^k = \sum_{k=0}^{\infty}(k+1)\epsilon_{k+1}\beta^k \quad \text{(A-9)}$$

which yields:

$$\sum_{l=0}^{m}\epsilon_{m-l}\tilde{c}_l = (m+1)\epsilon_{m+1}, \quad m = 0,1,2,\ldots \quad \text{(A-10)}$$

which can be written as the following system of linear equations:

$$\begin{array}{ccccccccccc}
\epsilon_0\tilde{c}_0 & & & & & & & & = & \epsilon_1 \\
\epsilon_1\tilde{c}_0 & + & \epsilon_2\tilde{c}_1 & & & & & & = & 2\epsilon_2 \\
\epsilon_2\tilde{c}_0 & + & \epsilon_1\tilde{c}_1 & + & \epsilon_0\tilde{c}_2 & & & & = & 3\epsilon_3 \\
\vdots & & \vdots & & \vdots & & & & \vdots & \vdots \\
\epsilon_m\tilde{c}_0 & + & \epsilon_{m-1}\tilde{c}_1 & + & \epsilon_{m-2}\tilde{c}_2 & + & \cdots & + & \epsilon_0\tilde{c}_m & = & (m+1)\epsilon_{m+1} \\
\vdots & & \vdots & & \vdots & & \vdots & & \vdots & \vdots & \vdots
\end{array} \tag{A-11}$$

Because only the lower triangle of the matrix of the system of equations of eq. (A-11) contains nonzero elements, the equations can be solved step by step starting from the first one. Therefore, to obtain an expression for $\tilde{c}_{m-1}$ [which is related to c_m by eq. (A-7)], we need to take only the first m equations. Using the well-known determinant formula for solving a system of linear equations, we obtain:

$$\tilde{c}_{m-1} = \frac{(-1)^{m-1}}{\epsilon_0^m} \times \begin{vmatrix} \epsilon_1 & \epsilon_0 & 0 & \cdots & 0 \\ 2\epsilon_2 & \epsilon_1 & \epsilon_0 & \cdots & 0 \\ 3\epsilon_3 & \epsilon_2 & \epsilon_1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ (m-1)\epsilon_{m-1} & \epsilon_{m-2} & \epsilon_{m-3} & \cdots & \epsilon_0 \\ m\epsilon_m & \epsilon_{m-1} & \epsilon_{m-2} & \cdots & \epsilon_1 \end{vmatrix} \tag{A-12}$$

Substituting eq. (A-12) into eq. (A-7), and recalling that $\epsilon_0 = U_0 = 1$ [cf. eq. (A-5)], we obtain:

$$c_m = \frac{(-1)^{m-1}}{m} \times \begin{vmatrix} \epsilon_1 & 1 & 0 & \cdots & 0 \\ 2\epsilon_2 & \epsilon_1 & 1 & \cdots & 0 \\ 3\epsilon_3 & \epsilon_2 & \epsilon_1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ (m-1)\epsilon_{m-1} & \epsilon_{m-2} & \epsilon_{m-3} & \cdots & 1 \\ m\epsilon_m & \epsilon_{m-1} & \epsilon_{m-2} & \cdots & \epsilon_1 \end{vmatrix} \tag{A-13}$$

and:

$$C_m = -(m-1)! \times \begin{vmatrix} \epsilon_1 & 1 & 0 & \cdots & 0 \\ 2\epsilon_2 & \epsilon_1 & 1 & \cdots & 0 \\ 3\epsilon_3 & \epsilon_2 & \epsilon_1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ (m-1)\epsilon_{m-1} & \epsilon_{m-2} & \epsilon_{m-3} & \cdots & 1 \\ m\epsilon_m & \epsilon_{m-1} & \epsilon_{m-2} & \cdots & \epsilon_1 \end{vmatrix} \tag{A-14}$$

A general formula for cumulants that does not involve determinants [but much more complicated than eq. (A-14)] and the formulas for the first few cumulants can be found in ref. 18.

To find the conditions for the convergence of the series given by eq. (A-2), we first note that expanding the determinant in eq. (A-13) will result in exactly 2^m products of the ϵ values, and the sum of the products of the orders and the powers of all ϵ values in each such terms will be equal to m. The first fact can be proved by the observation that expanding the determinant with respect to the first row results in two minors which have exactly the same structure (i.e., the first two elements of the first rows are nonzero elements and the remaining ones are zero). Thus, expanding each of these minors will again result in two minors of the same structure, and so on. The second fact can easily be proved by mathematical induction, taking advantage of eq. (A-10). Therefore, if the absolute value of the energy $|E|$ is bounded by M, and, consequently [by use of eq. (A-5)]:

$$|\epsilon_k| \le \frac{M^k}{k!} \le M, \quad k = 1, 2, \ldots \tag{A-15}$$

we can express [with the use of eqs. (A-13) and (A-15)] the upper bound of $|c_m|$ as:

$$|c_m| \le \frac{2^m}{m} m M^m = (2M)^m, \quad m = 1, 2, \ldots \quad \text{(A-16)}$$

The extra m in the first power appears in eq. (A-16) because of the coefficients $1, 2, \ldots, m$ that appear in the first column of the determinant of eq. (A-13).

Therefore, we can estimate the series of eq. (A-2) by:

$$\left| \sum_{k=1}^{\infty} c_k \beta^k \right| \le \sum_{k=1}^{\infty} (2M\beta)^k \quad \text{(A-17)}$$

which is a geometric series with a ratio equal to $2M\beta$, which is convergent for:

$$\beta < \frac{1}{2M} \quad \text{(A-18)}$$

Appendix 2: Graphical Approach to Evaluation of Integrals $\tilde{\Upsilon}$

Consider the graph representing $u_{\widehat{ijk}}$ as an example, as shown in Figure 9. The total graph is split into the sum of directed graphs, each of which represents one integral $\Upsilon(i_1, j_1, \ldots, i_k, j_k; s_1, t_1, \ldots, s_k, t_k)$ of the form given by eq. (21) given earlier. The total number of such integrals equals 2^{2k} (where k denotes the number of edges). An arrow pointing *from* vertex k in a directed graph indicates that $\exp(i\lambda_k)$ enters into the respective term in eq. (21), while an arrow pointing *towards* a vertex indicates that $\exp(-i\lambda_k)$ enters. The sign of λ_k is, on the other hand, equal to s_m or $s_m t_m$ in eq. (23) given earlier. For nonvanishing integrals Υ, eqs. (25) must hold; therefore, the number of contributions with a "+" sign must be equal to the number of contributions with a "−" sign at each

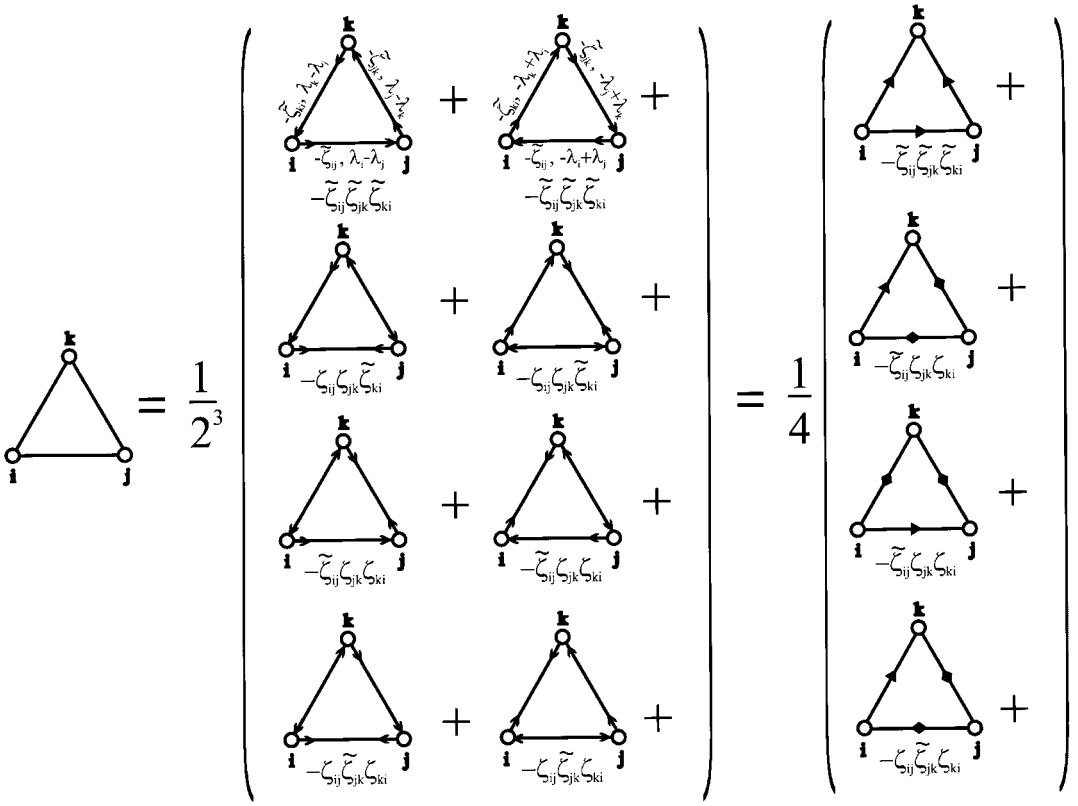

FIGURE 9. Graphical approach to finding of all nonvanishing terms in eq. (21) illustrated by the example of three-body correlation contributions, $u_{\widehat{ijk}}$. See Appendix 2 for explanation.

vertex (this will cancel the imaginary exponentials). This implies that the number of arrows going in and out from any vertex must be equal. Based on this observation, all directed graphs corresponding to nonvanishing integrals Υ in eq. (21) can easily be constructed, as shown in Figure 9.

Once a directed graph is constructed, each of its edges has two arrows, which show the relation between the signs of the λ values corresponding to the vertices linked by this edge. For vertices j and k, for example, the arrows pointing in the same direction show that the λ_j and λ_k enter with opposite signs; this implies that $t_m = -1$. Therefore, the contribution of this edge to the expression for the integral Υ in eq. (27) equals $-\tilde{\zeta}_{jk}$. If the two λ values have the same sign, $t_m = +1$ and the contribution is equal to ζ_{jk}. The arguments and the phase angles of each edge are shown for the two directed graphs shown at the top of Figure 9; for the six remaining graphs, only the contributions to the products constituting the integrals Υ are shown.

As mentioned in the remark under eq. (26), there are more than one nonvanishing integrals $\Upsilon(i_1, j_1, \ldots, i_k, j_k; s_1, t_1, \ldots, s_k, t_k)$, with value $\tilde{\Upsilon}(i_1, j_1, \ldots, i_k, j_k; t_1, \ldots, t_k)$; they correspond to the same set of the numbers, $t_1, \ldots, t_k$, and different sets of the numbers, $s_1, \ldots, s_k$, such that eqs. (25) are satisfied. By counting the number of the directed graphs that correspond to a given value of $\tilde{\Upsilon}(i_1, j_1, \ldots, i_k, j_k; t_1, \ldots, t_k)$, we obtain the weight, $\gamma(i_1, j_1, \ldots, i_k, j_k, t_1, \ldots, t_k)$, of the respective term in eq. (28) and, further, all the integrals, $\mathcal{I}$, of eq. (17).

References

1. M. Levitt, *J. Mol. Biol.*, **104,** 59 (1976).
2. M. R. Pincus and H. A. Scheraga, *J. Phys. Chem.*, **81,** 1579 (1977).
3. A. Godzik, A. Koliński, and J. Skolnick, *J. Comput.-Aided Mol. Design*, **7,** 397 (1993).
4. M. J. Sippl, *J. Comput.-Aided Mol. Design*, **7,** 473 (1993).
5. A. Liwo, M. R. Pincus, R. J. Wawak, S. Rackovsky, and H. A. Scheraga, *Prot. Sci.*, **2,** 1697 (1993).
6. A. Liwo, M. R. Pincus, R. J. Wawak, S. Rackovsky, and H. A. Scheraga, *Prot. Sci.*, **2,** 1715 (1993).
7. V. N. Maiorov and G. M. Crippen, *Proteins*, **20,** 167 (1994).
8. M.-H. Hao and H. A. Scheraga, *J. Phys. Chem.*, **100,** 14540 (1996).
9. A. Liwo, S. Ołdziej, M. R. Pincus, R. J. Wawak, S. Rackovsky, and H. A. Scheraga, *J. Comput. Chem.*, **18,** 849 (1997).
10. A. Liwo, M. R. Pincus, R. J. Wawak, S. Rackovsky, S. Ołdziej, and H. A. Scheraga, *J. Comput. Chem.*, **18,** 874 (1997).
11. F. C. Bernstein, T. F. Koetzle, G. J. B. Williams, E. F. Meyer, Jr., M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi, *J. Mol. Biol.*, **112,** 535 (1977).
12. S. Miyazawa and R. L. Jernigan, *Macromolecules*, **18,** 534 (1985).
13. K. Nishikawa, F. A. Momany, and H. A. Scheraga, *Macromolecules*, **7,** 797 (1974).
14. F. A. Momany, R. F. McGuire, A. W. Burgess, and H. A. Scheraga, *J. Phys. Chem.*, **79,** 2361 (1975).
15. G. Némethy, M. S. Pottle, and H. A. Scheraga, *J. Phys. Chem.*, **87,** 1883 (1983).
16. A. Liwo, S. Ołdziej, J. Ciarkowski, G. Kupryszewski, M. R. Pincus, R J. Wawak, S. Rackovsky, and H. A. Scheraga, *J. Prot. Chem.*, **13,** 375 (1994).
17. L. E. Reichl, *A Modern Course in Statistical Physics*, University of Texas Press, Austin, TX, 1980, pp. 142–144.
18. M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics*, Griffin, London, 1968, pp. 67–71.
19. Mal'cev, A. I. (1970) *Algorithms and Recursive Functions*, Wolters-Noordhoff, Groningen, p. 67.
20. D. A. McQuarrie, *Statistical Mechanics*, Harper & Row, 1976, pp. 226–233.
21. H. N. V. Temperley, *Graph Theory and Applications*, Horwood, NY, 1981, pp. 82–89.
22. M. L. Abell and J. P. Braselton, *The Maple V Handbook*, Academic Press, New York, 1994.
23. G. Némethy, K. D. Gibson, K. A. Palmer, C. N. Yoon, G. Paterlini, A. Zagari, S. Rumsey, and H. A. Scheraga, *J. Phys. Chem.*, **96,** 6472 (1992).
24. L. Piela and H. A. Scheraga, *Biopolymers*, **26,** S33 (1987).
25. D. R. Ripoll and H. A. Scheraga, *Biopolymers*, **27,** 1283 (1988).
26. K. A. Olszewski, L. Piela, and H. A. Scheraga, *J. Chem. Phys.*, **96,** 4672 (1992).
27. J. Vila, R. L. Williams, M. Vásquez, and H. A. Scheraga, *Proteins*, **10,** 199 (1991).
28. D. N. Silverman and H. A. Scheraga, *Arch. Biophys. Biochem.*, **153,** 449 (1972).
29. Z. Li and H. A. Scheraga, *Proc. Natl. Acad. Sci. USA*, **84,** 6611 (1987).
30. Z. Li and H. A. Scheraga, *J. Mol. Struct.* (*Theochem*), **179,** 333 (1988).
31. D. J. Barlow and J. M. Thornton, *J. Mol. Biol.*, **201,** 601 (1988).