UNRES – a united-residue force field for energy-based prediction of protein structure – origin and significance of multibody terms

Adam Liwo,[1,2] Jarosław Pillardy,[2] Cezary Czaplewski,[1,2] Jooyoung Lee,[2]
Daniel R. Ripoll,[3] Małgorzata Groth,[1] Sylwia Rodziewicz-Motowidło,[1]
Rajmund Kaźmierkiewicz,[1] Ryszard J. Wawak,[2] Stanisław Ołdziej,[1]
and Harold A. Scheraga[2]

[1]Faculty of Chemistry, University of Gdańsk, ul. Sobieskiego 18, 80-952 Gdańsk, Poland
[2]Baker Laboratory of Chemistry and Chemical Biology, Cornell University,
Ithaca, N.Y. 14853-1301, U.S.A.
[3]Cornell Theory Center, Ithaca, NY 14853-3801, U.S.A.

United-residue models of polypeptide chains [3,5,19–22,24,31,33] have long been of interest, because they enable one to carry out global conformational searches of proteins in real time, which in turn can facilitate *ab initio* protein structure predictions based solely on Anfinsen's thermodynamic hypothesis [1], according to which the native structure of a protein is a global minimum of its potential energy surface [32]. In the last few years, we developed a united-residue force field [20–22,24], hereafter referred to as UNRES, in which a polypeptide chain is represented by a sequence of α-carbon (C^{α}) atoms linked by virtual bonds with attached united side chains (SC) and united peptide groups (p) located in the middle between the consecutive α-carbons (Figure 1). Only the united peptide groups and united side chains serve as interaction sites, the α-carbons serving to define the geometry. All the virtual bond lengths (i.e. C^{α}–C^{α} and C^{α}–SC) are fixed; the C^{α}–C^{α} distance is taken as 3.8 Å which corresponds to *trans* peptide groups, while the side-chain angles (α_{SC} and β_{SC}), as well the virtual-bond angles (θ and γ) can vary. The energy of the virtual-bond chain is expressed by eq. (1).

$$
\begin{aligned}
U = & \sum_{i<j} U_{SC_i SC_j} + \sum_{i \neq j} U_{SC_i p_j} \\
& + w_{el} \sum_{i<j-1} U_{p_i p_j} + w_{tor} \sum_{i} U_{tor}(\gamma_i) \\
& + w_{loc} \sum_{i} [U_b(\theta_i) + U_{rot}(\alpha_{SC_i}, \beta_{SC_i})] \\
& + w_{corr} U_{corr}
\end{aligned}
\tag{1}
$$

The term $U_{SC_i SC_j}$ corresponds to the mean free energy of the hydrophobic (hydrophilic) interactions between the side chains. It therefore implicitly contains the contributions from the interactions with the solvent. The terms $U_{SC_i p_j}$ denote the excluded-volume potential of the side-chain – peptide-group interactions. The peptide-group interaction potential ($U_{p_i p_j}$) accounts mainly for the electrostatic interactions between them or, in other words, for their tendency to form backbone hydrogen bonds. U_{tor}, U_b, and U_{rot} denote the energies of virtual-dihedral angle torsions, virtual-angle bending, and side-chain rotamers; these terms reflect the local propensities of the polypeptide chain. The term U_{corr} is the *correlation* or *multibody* contribution and the w's are the weights of the energy terms.

In contrast to all-atom force fields, the multibody terms are not just a small addition; the multibody terms are an essential ingredient of coarse-grain united-residue force fields. This is because coarse-grain potentials are mean-field potentials, corresponding to the restricted free energy, $F(\mathbf{X})$,
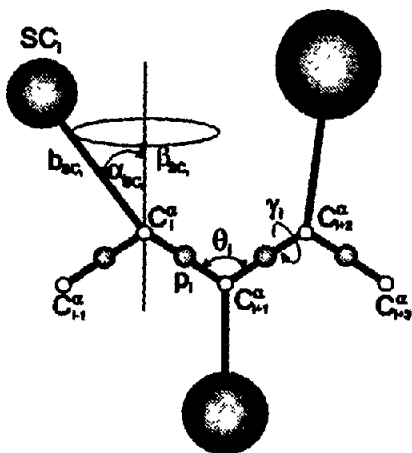
Figure 1: The UNRES model of polypeptide chains.

calculated for given configurations of the "coarse-grain" interaction sites (p and SC in the case of the UNRES force field; Figure 1) and to averaging over the remaining, "less important" degrees of freedom, as expressed by eq. (2) [20].

$$F(\mathbf{X}) = -RT \ln \left\{ \frac{1}{V_Y} \right.$$

$$\left. \int_{\Omega_Y} \exp[-E(\mathbf{X};\mathbf{Y})/k_B T] dV_Y \right\} \quad (2)$$

with

$$V_Y = \int_{\Omega_Y} dV_Y.$$

where $E(\mathbf{X};\mathbf{Y})$ is the original energy function, $\mathbf{X}$ denotes the vector of the degrees of freedom of the "coarse-grain" system (the virtual bond angles θ, the virtual-bond dihedral angles γ, and the polar angles β_{SC} and γ_{SC} in UNRES), $\mathbf{Y}$ denotes the vector of the degrees of freedom over which the average is computed (e.g., the positions and orientations of solvent molecules, the side-chain dihedral angles χ, etc.), R is the gas constant, T is the absolute temperature, Ω_Y is the region of the $\mathbf{Y}$ subspace of variables over which the integration is carried out and V_Y is the volume of this region. Expanding $F(\mathbf{X})$ of eq. (2) in the cumulant series in $\beta = 1/RT$, we obtain [20]:

$$F(\mathbf{X}) \equiv F(\beta, \mathbf{X}) = U_1 - \frac{1}{2}(U_2 - U_1^2)\beta$$

$$+ \frac{1}{6}(U_3 - 3U_1 U_2 + 2U_1^3)\beta^2$$

$$- \frac{1}{24}(U_4 - 3U_2^2 - 4U_1 U_3 + 12U_1^2 U_2$$

$$- 6U_1^4)\beta^3 + \dots$$

$$= \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k!} C_k(\mathbf{X})\beta^{k-1} \quad (3)$$

where C_k is the k-th order cumulant and U_1, U_2, $\dots U_n$ are consecutive energy moments:

$$U_k = \frac{1}{V_Y} \int_{\Omega_Y} E(\mathbf{X};\mathbf{Y})^k dV_Y \quad (4)$$

Even if the original energy function $E(\mathbf{X};\mathbf{Y})$ contains at most pairwise terms, the restricted free energy $F(\mathbf{X})$ will in general contain higher-order terms that arise from the presence of higher energy moments in the cumulant expansion [eq. (3)]. The early version of UNRES [22,24] did not contain multibody terms and was therefore good only for inverse folding (i.e., it could recognize a native fold corresponding to a given amino-acid sequence in the data base of decoys taken from the PDB), but was not capable of de novo folding of a protein [22,24]. The capability of de novo folding was achieved only after introducing multibody or correlation terms in the backbone electrostatic interactions [20,23]. Similar conclusions about the role of backbone hydrogen bonding and other multibody terms have also been drawn by other workers [5,10,11].

The side-chain (U_{SCSC}) and the components of the local-interaction potential (U_{tor}, U_b, and U_{rot}) were parameterized based on distribution and correlation functions determined [22,24] from a set of 195 high-resolution non-homologous structures from the Protein Data Bank (PDB) [2]. The peptide-group interaction potential $U_{p_i p_j}$ and the correlation terms pertaining to backbone hydrogen bonding (U_{corr}) were parameterized by averaging the all-atom ECEPP/2 [27,28] potential. Finally, the relative weights of the energy terms were determined so as to maximize the energy gap between
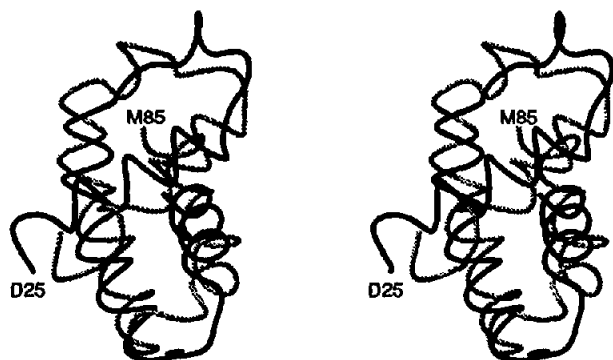
194

Figure 2: The predicted structure of the 25-85 fragment of HDEA (grey ribbon) superposed on the experimental structure (PDB code 1BG8) [34] (dark ribbon); the C^α RMS deviation is 4.2 Å.

the native structure and the average energy of the non-native structures [24]; this was accomplished by the minimization of the so-called Z-score function [6–9].

The version of the UNRES force field described above with the Conformational Space Annealing (CSA) global-optimization procedure [16–18] was first tested on two simple helical proteins: the 10-58 fragment of the B-domain of staphylococcal protein A (a three-helix-bundle topology) and apo-calbindin D9K (a 75-residue protein with the topology of a four-helix bundle with an EF-hand motif) [15]. In both cases, the native structure and its mirror image were located; the native structure was lower in energy for protein A, and higher for apo-calbindin. A full-blown blind-prediction test was performed within the CASP3 experiment. We submitted predictions for seven targets, one of which, for the periplasmic protein HDEA from E. coli, turned out to be the most accurate one among all the models submitted [30], including homology modeling and threading (Figure 2). We also achieved very good results for DNA b helicase, a 116-residue protein. These two proteins were assessed as particularly difficult targets [30], because of their rare folds. Our predictions of most of the other targets were also fairly accurate [13,14,21,30]. However, the version of the UNRES force field that includes only the multibody terms pertaining to hydrogen bonding pro-

duces too-distorted β-structure.

In our present work, we introduced two new types of multibody terms: U_{tor}^{mult} and $U_{el,loc}$ that arise from the coupling between local interactions involving more than two consecutive peptide units and local and backbone electrostatic interactions, respectively. These terms comprise the third – sixth order terms in the cumulant expansion for the restricted free energy [eq. (3)]. These new terms should extend the performance of our procedure to proteins that contain β-structure. By local-interaction energy (E_{loc}), we denote the conformational energy of an isolated peptide unit. E_{loc} can be modeled by the energy surface of an N-acetyl-N'-methylamide derivative of the amino acid residue under study [35]. It is usually expressed as a function of the dihedral angles ϕ and ψ; however, for the purpose of implementing it in eqs. (2) and (3), we express it as a function of the dihedral angles of rotation λ_1 and λ_2 about the C^α–C^α bonds forming the peptide unit [29] (Figure 3), which provides a clear separation of the degrees of freedom into the "coarse" or "important" ones [X of eq. (2)] and "fine" or "less important" ones [Y of eq. (2)], the "less important" ones being the dihedral angles λ [25,26]. The dihedral angles ϕ and ψ describe both the "coarse" and "fine" shape of the polypeptide backbone and cannot therefore be implemented directly in the calculation of the restricted free energy.

The lowest-energy conformations (i.e., region C) [35] of terminally-blocked L-amino acid residues (the smallest peptide units) lie exactly in the region of the (ϕ, ψ)-dihedral angle space characteristic of
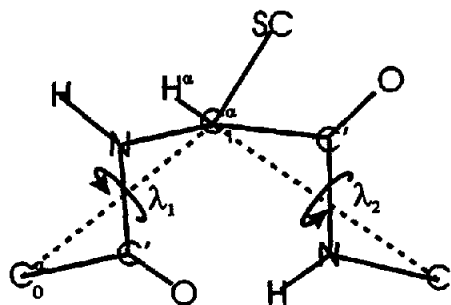

Figure 3: The definition of the dihedral angles λ_1 and λ_2 of rotation about the C^α–C^α virtual bonds of a peptide unit [29].

β-structure. Therefore, the mixed local and electrostatic energy moments in the cumulant expansion of the restricted free energy of the polypeptide chain [eq. (3)] contribute to the stabilization of β-structures.

The contributions to consecutive energy moments [needed to compute the cumulant expansion for $F(X)$ in eq. (3)] that include the products of the local and electrostatic energies have the following general form:

$$U_{k;el,loc} = \frac{1}{(2\pi)^{N_k}} \int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} E_{loc}^{j} E_{el}^{k-j} \mathrm{d}\lambda_{i_1} \mathrm{d}\lambda_{i_2}$$
$$\ldots \lambda_{i_{N_k}}, \quad j = 1, 2, \ldots, k-1; \quad k = 2, 3, \ldots \tag{5}$$

where $\lambda_{i_1}, \lambda_{i_2} \ldots \lambda_{i_{N_k}}$ indicate the dihedral angles λ involved in integration (their number, N_k, will vary depending on the contribution to the energy moment). Likewise, the contributions needed to determine multiple-torsional terms (U_{tor}^{mult}) are expressed by eq. (6).

$$U_{k;tor}^{mult} = \frac{1}{(2\pi)^{N_k}} \int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} E_{loc}^{k} \mathrm{d}\lambda_{i_1} \mathrm{d}\lambda_{i_2} \ldots \lambda_{i_{N_k}},$$
$$k = 2, 3, \ldots \tag{6}$$

To a very good approximation, $E_{loc}(\lambda_1, \lambda_2)$ can be expressed as a second-order Fourier series in λ_1 and λ_2. In our dipole model of the peptide group [25], the energy of electrostatic interaction between peptide groups is a second-order Fourier series in the angles λ (this follows directly from the energetics of the dipole-dipole interactions [25]). Therefore, all energy moments involving the local and/or the electrostatic energy can be calculated analytically. In our earlier work [20], we developed an algorithm for calculating the moments of the electrostatic-interaction energy; this algorithm was used to derive the hydrogen-bonding multibody terms in UNRES. We have now generalized this algorithm to compute the energy moments involving both electrostatic and local interactions, and derived the terms in the cumulant expansion for $F(X)$ [eq. (3)]
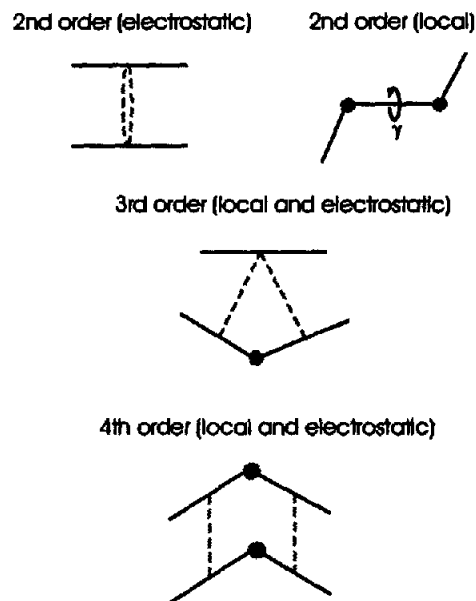

Figure 4: Representation of some cumulant contributions to the restricted free energy in the original version of UNRES. A filled circle indicates local interaction energy in the respective peptide unit and a dashed line indicates the energy of electrostatic interaction between the two peptide groups that it connects.

up to the sixth order. Instead of writing the complicated formulas for the component terms here, we present them as graphs in Figures 4, 5a and 5b. Figure 4 includes the terms already present in the original version of UNRES, while Figures 5a and 5b show the new terms.

In Figure 4, the upper left graph represents the averaging of the square of the energy of the electrostatic interactions between the peptide groups; it corresponds to the contribution to U_{pp} of eq. (1) coming from a pair of interacting peptide groups. The upper right graph corresponds to the averaging of the products of local-interaction energy of two consecutive peptide units. As a result of this, the average becomes dependent on the virtual dihedral angle γ centered at the C^{α}-C^{α} virtual bond connecting the two units. This term formally corresponds to U_{tor} of eq. (1). Both of the graphs described above come from second moments of the energy of the all-atom chain. The middle and the bottom graphs represent the dominant three- and four-body contributions to the restricted free en-
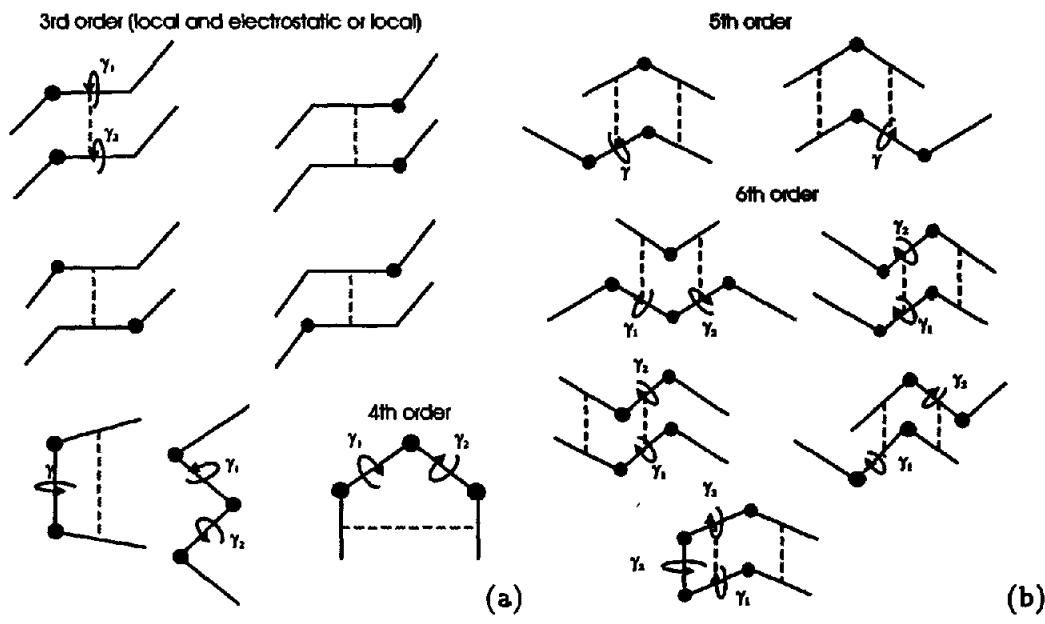
Figure 5: Graphical representation of the new multibody terms in the cumulant expansion for the restricted free energy of the polypeptide chain, corresponding to U_{tor}^{mult} (bottom middle graph in part a) and the third- and fourth-order terms in $U_{el,loc}$ (a) and the fifth- and sixth-order terms in $U_{el,loc}$.

ergy. They were derived in our earlier work [20] as the components of the third and the fourth-order term in the cumulant expansion of the backbone electrostatic energy. However, they also include the averaging of the electrostatic energy between neighboring peptide groups, which is a part of the local interaction energy of the peptide unit to which these peptide group belong. Therefore, these terms should be considered as parts of $U_{el,loc}$.

The third-order components of $U_{el,loc}$ presented in Figure 5a describe the correlation between the energy of the electrostatic interaction between two non-contiguous peptide groups and the energy of local interaction of the neighboring peptide units. As shown, these terms are dependent not only on the relative orientation of the two peptide groups, but also on the virtual-bond dihedral angles γ_1 and γ_2 centered on the corresponding C^α–C^α virtual-bond axes. In other words, a given orientation of two peptide groups invokes a certain local fold of the respective portion of the polypeptide chain, or, long-range interactions have the capacity of propagating an ordered local structure. Figure 6 shows that a parallel or antiparallel orientation of two interacting peptide groups, as in β-sheets, favors extended virtual-bond dihedral angles and, thus

propagates extended configurations of the polypeptide chain as in β-strands (note the minimum at $\gamma_1 = \gamma_2 = \pm 180°$). The fifth- and sixth-order components displayed in Figure 5b should also be important, because they propagate the local fold of the chain, if two pairs of neighboring peptide groups are in contact.

The double torsional terms (U_{tor}^{mult}; represented as the middle graph in the bottom of Figure 5a) should also be important with regard to the formation of ordered structures. Their importance has already been pointed out by other workers [4]. From our preliminary analysis, it appears that these double torsional terms contribute to the stabilization of left-handed extended strands [which, in turn, lead to (the observed) right-handed β-sheets]. The other two third- and fourth-order terms shown in the bottom part of Figure 5a involve local and electrostatic interaction correlations within three and four adjacent peptide units, respectively; they can be important for the correct description of the geometry of β- or γ-turns and of the geometry of α-helices.

To test the ability of UNRES augmented with the new correlation terms to reproduce the structure
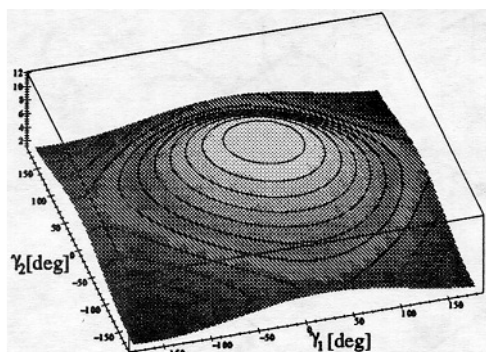
Figure 6: The dependence of the sum of the four third-order $U_{el,loc}$ terms of the upper part of Figure 5a (which is the total unit third-order contribution to $U_{el,loc}$) on the virtual-bond dihedral angles γ_1 and γ_2 shown in Figure 5a. The central peptide groups are assumed to stack on each other.

of β-sheets, we used the sequence of a 20-residue polypeptide betanova, which was recently designed as a minimum β-sheet model [12]. This peptide forms a *stable* three-stranded β-sheet in solution, as revealed by NMR spectroscopy [12]. The first calculation was carried out without including the new features of the force field (i.e., the only multibody terms included are those shown in Figure 4), while the second one was carried out with inclusion of the third-, fourth- and fifth-order correlation terms that are depicted in Figure 5; these term pertain to the coupling between local and electrostatic interactions. In both cases the CSA method [16–18] was used to find the global minimum. As shown, lack of sufficient terms responsible for the coupling between the local and backbone hydrogen-bonding interactions leads effectively to a coil structure (Figure 7a). Including the multibody terms introduced in this work leads to correct topology of betanova with correct positions of both turns and correct contacts between the side chains [12]. It should be noted that this result was obtained even without systematic calibration of the weights of the new correlation contributions to energy. At present, we are determining the weights of the new correlation terms in a systematic way by means of Z-score optimization.
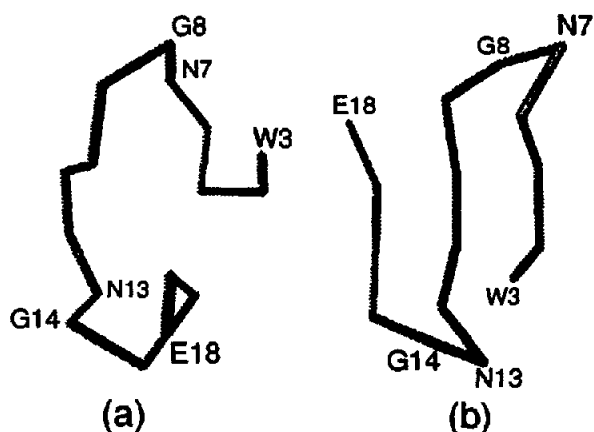

(a)   (b)

Figure 7: The lowest-energy structures of betanova calculated with the UNRES force field without inclusion of the new correlation terms (a) and with inclusion of the new correlation terms (b). The N- and C-terminal residues, as well as the turn residues in the NMR structure of betanova are marked for tracing purpose.

References

[1] ANFINSEN, C. B. Principles that govern the folding of protein chains. *Science 181* (1973), 223–230.

[2] BERNSTEIN, F. C., KOETZLE, T. F., WILLIAMS, G. J. B., JR., E. F. M., BRICE,

M. D., RODGERS, J. R., KENNARD, O., SHI-MANOUCHI, T., AND TASUMI, M. The Protein Data Bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol. 112* (1977), 535–542.

[3] CRIPPEN, G. M. Easily searched protein folding potentials. *J. Mol. Biol. 260* (1996), 467–475.

[4] DEWITTE, R. S., AND SHAKHNOVICH, E. I. Pseudodihedrals: Simplified protein backbone representation with knowledge-based energy. *Protein Sci. 3* (1994), 1570–1581.

[5] GODZIK, A., KOLIŃSKI, A., AND SKOLNICK, J. De novo and inverse folding predictions of protein structure and dynamics. *J. Comput. Aid. Mol. Des. 7* (1993), 397–438.

[6] GOLDSTEIN, R. A., LUTHEY-SCHULTEN, Z. A., AND WOLYNES, P. G. Protein tertiary structure recognition using optimized Hamiltonians with local interactions. *Proc. Natl. Acad. Sci. USA 89* (1992), 9029–9033.

[7] HAO, M.-H., AND SCHERAGA, H. A. Statistical thermodynamics of protein folding: Sequence dependence. *J. Phys. Chem. 98* (1994), 9882–9893.

[8] HAO, M.-H., AND SCHERAGA, H. A. How optimization of potential functions affects protein folding. *Proc. Natl. Acad. Sci. USA 93* (1996), 4984–4989.

[9] KLIMOV, D. K., AND THIRUMALAI, D. Linking rates of folding in lattice models of proteins with underlying thermodynamic characteristics. *J. Chem. Phys. 109* (1998), 4119–4125.

[10] KOLIŃSKI, A., AND SKOLNICK, J. Monte Carlo simulations of protein folding. I. Lattice model and interaction scheme. *Proteins Struct. Funct. Genet. 18* (1994), 338–352.

[11] KOLIŃSKI, A., AND SKOLNICK, J. Monte Carlo simulations of protein folding. II. Application to protein A and crambin. *Proteins Struct. Funct. Genet. 18* (1994), 353–366.

[12] KORTEMME, T., RAMIREZ-ALVARADO, M., AND SERRANO, L. Design of a 20-amino acid, three-stranded β-sheet protein. *Science 282* (1998), 253–256.

[13] LEE, J., LIWO, A., RIPOLL, D. R., PILLARDY, J., SAUNDERS, J. A., GIBSON, K. D., AND SCHERAGA, H. A. Hierarchical energy-based approach to protein-structure prediction; blind-test evaluation with CASP3 targets. *Int. J. Quant. Chem.* (1999). In press.

[14] LEE, J., LIWO, A., RIPOLL, D. R., PILLARDY, J., AND SCHERAGA, H. A. Calculation of protein conformation by global optimization of a potential energy function. *Proteins Struct. Funct. Genet. Suppl. 3* (1999), 204–408.

[15] LEE, J., LIWO, A., AND SCHERAGA, H. A. Energy-based *de novo* protein folding by conformational space annealing and an off-lattice united-residue force field: Application to the 10-55 fragment of staphylococcal protein A and to apo-calbindin D9K. *Proc. Natl. Acad. Sci. U. S. A. 96* (1999), 2025–2030.

[16] LEE, J., AND SCHERAGA, H. A. Conformational space annealing by parallel computations: extensive conformational search of Met-enkephalin and of the 20-residue membrane-bound portion of melittin. *Int. J. Quant. Chem.* (1999). In press.

[17] LEE, J., SCHERAGA, H. A., AND RACKOVSKY, S. New optimization method for conformational energy calculations on polypeptides: Conformational space annealing. *J. Comput. Chem. 18* (1997), 1222–1232.

[18] LEE, J., SCHERAGA, H. A., AND RACKOVSKY, S. Conformational analysis of the 20-residue membrane-bound portion of melittin by Conformational space annealing. *Biopolymers 46* (1998), 103–115.

[19] LEVITT, M., AND WARSHELL, A. Computer simulation of protein folding. *Nature 253* (1975), 694–698.

[20] LIWO, A., KAŹMIERKIEWICZ, R., CZAPLEWSKI, C., GROTH, M., OLDZIEJ, S., WAWAK, R. J., RACKOVSKY, S., PINCUS, M. R., AND SCHERAGA, H. A. United-residue force field for off-lattice protein-structure simulations; III. Origin of backbone hydrogen-bonding cooperativity in

united-residue potentials. *J. Comput. Chem.* *19* (1998), 259–276.

[21] LIWO, A., LEE, J., RIPOLL, D. R., PILLARDY, J., AND SCHERAGA, H. A. Protein structure prediction by global optimization of a potential energy function. *Proc. Natl. Acad. Sci., U. S. A.* *96* (1999), 5482–5485.

[22] LIWO, A., OŁDZIEJ, S., PINCUS, M. R., WAWAK, R. J., RACKOVSKY, S., AND SCHERAGA, H. A. A united-residue force field for off-lattice protein-structure simulations. I. Functional forms and parameters of long-range side-chain interaction potentials from protein crystal data. *J. Comput. Chem.* *18* (1997), 849–873.

[23] LIWO, A., PILLARDY, J., KAŹMIERKIEWICZ, R., WAWAK, R. J., GROTH, M., CZAPLEWSKI, C., OŁDZIEJ, S., AND SCHERAGA, H. A. Prediction of protein structure using a knowledge-based off-lattice united-residue force field and global optimization methods. *Theoretical Chemistry Accounts* *101* (1999), 16–20.

[24] LIWO, A., PINCUS, M. R., WAWAK, R. J., RACKOVSKY, S., OŁDZIEJ, S., AND SCHERAGA, H. A. A united-residue force field for off-lattice protein-structure simulations. II: Parameterization of local interactions and determination of the weights of energy terms by Z-score optimization. *J. Comput. Chem.* *18* (1997), 874–887.

[25] LIWO, A., PINCUS, M. R., WAWAK, R. J., RACKOVSKY, S., AND SCHERAGA, H. A. Calculation of protein backbone geometry from α-carbon coordinates based on peptide-group dipole alignment. *Protein Sci.* *2* (1993), 1697–1714.

[26] LIWO, A., PINCUS, M. R., WAWAK, R. J., RACKOVSKY, S., AND SCHERAGA, H. A. Prediction of protein conformation on the basis of a search for compact structures; test on avian pancreatic polypeptide. *Protein Sci.* *2* (1993), 1715–1731.

[27] MOMANY, F. A., MCGUIRE, R. F., BURGESS, A. W., , AND SCHERAGA, H. A. Energy parameters in polypeptides. VII. Geometric parameters, partial atomic charges, non-bonded interactions, hydrogen bond interactions and intrinsic torsional potential for the naturally occurring amino-acids. *J. Phys. Chem.* *79* (1975), 2361–2381.

[28] NÉMETHY, G., POTTLE, M. S., AND SCHERAGA, H. A. Energy parameters in polypeptides. 9. Updating of geometrical parameters, nonbonded interactions, and hydrogen bond interactions for the naturally occurring amino acids. *J. Phys. Chem.* *87* (1983), 1883–1887.

[29] NISHIKAWA, K., MOMANY, F. A., AND SCHERAGA, H. A. Low-energy structures of two dipeptides and their relationship to bend conformations. *Macromolecules* *7* (1974), 797–806.

[30] ORENGO, C. A., BRAY, J. E., HUBBARD, T., CONTE, L. L., AND SILLITOE, I. Analysis and assessment of ab initio three-dimensional prediction, secondary structure, and contacts prediction. *Proteins Struct. Funct. Genet. Suppl.* *3* (1999), 149–170.

[31] PINCUS, M. R., AND SCHERAGA, H. A. An approximate treatment of long-range interactions in proteins. *J. Phys. Chem.* *81* (1977), 1579–1583.

[32] SCHERAGA, H. A. Recent developments in the theory of protein folding: searching for the global energy minimum. *Biophys. Chem.* *59* (1996), 329–339.

[33] SIPPL, M. J. Boltzmann's principle, knowledge-based mean fields and protein folding. An approach to the computational determination of protein structures. *J. Comput.-Aid. Mol. Des.* *7* (1993), 473–501.

[34] YANG, F., GUSTAFSON, K. R., BOYD, M. R., AND WLODAWER, A. Crystal structure of *Escherichia coli* HDEA. *Nature Struct. Biol.* *5* (1998), 763–764.

[35] ZIMMERMAN, S. S., M. S. POTTLE, G. N., AND SCHERAGA, H. A. Conformational analysis of the twenty naturally occurring amino acid residues using ECEPP. *Macromolecules* *10* (1977), 1–9.