

Extension of UNRES Force Field to Treat Polypeptide Chains with D-Amino Acid Residues

Adam K. Sieradzan,^{*,†,‡} Ulrich H. E. Hansmann,[‡] Harold A. Scheraga,[§] and Adam Liwo[†]

[†]Faculty of Chemistry, University of Gdańsk, Sobieskiego 18, 80-952 Gdańsk, Poland

[‡]Department of Chemistry and Biochemistry, Oklahoma University, Norman, Oklahoma 73019, United States

[§]Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853-1301, United States

S Supporting Information

ABSTRACT: Coarse-grained force fields for protein simulations are usually designed and parametrized to treat proteins composed of natural L-amino acid residues. However, D-amino acid residues occur in bacterial, fungal (e.g., gramicidins), as well as human-designed proteins. For this reason, we have extended the UNRES coarse-grained force field developed in our laboratory to treat systems with D-amino acid residues. We developed the respective virtual-bond-torsional and double-torsional potentials for rotation about the $C^{\alpha}\cdots C^{\alpha}$ virtual-bond axis and two consecutive $C^{\alpha}\cdots C^{\alpha}$ virtual-bond axes, respectively, as functions of virtual-bond-dihedral angles γ . In turn, these were calculated as potentials of mean force (PMFs) from the diabatic energy surfaces of terminally blocked model compounds for glycine, alanine, and proline. The potential-energy surfaces were calculated by using the *ab initio* method of molecular quantum mechanics at the Møller–Plesset (MP2) level of theory and the 6-31G(d,p) basis set, with the rotation angles of the peptide groups about $C_{i-1}^{\alpha}\cdots C_i^{\alpha}$ ($\lambda^{(1)}$) and $C_i^{\alpha}\cdots C_{i+1}^{\alpha}$ ($\lambda^{(2)}$) used as variables, and the energy was minimized with respect to the remaining degrees of freedom. The PMFs were calculated by numerical integration for all pairs and triplets with all possible combinations of types (glycine, alanine, and proline) and chirality (D or L); however, symmetry relations reduce the number of nonequivalent torsional potentials to 13 and the number of double-torsional potentials to 63 for a given C-terminal blocking group. Subsequently, one- (for torsional) and two-dimensional (for double-torsional potentials) Fourier series were fitted to the PMFs to obtain analytical expressions. It was found that the torsional potentials of the x – Y and X – y types, where X and Y are Ala or Pro, respectively, and a lowercase letter denotes D-chirality, have global minima for small absolute values of γ , accounting for the double-helical structure of gramicidin A, which is a dimer of two chains, each possessing an alternating D-Tyr–L-Tyr sequence, and similar peptides. The side-chain and correlation potentials for D-amino acid residues were obtained by applying the reflection about the $C_{i-1}^{\alpha}\cdots C_i^{\alpha}\cdots C_{i+1}^{\alpha}$ plane to the respective potentials for the L-amino acid residues.

1. INTRODUCTION

Less than 1% of all known protein and peptide structures contain D-amino acids.¹ However, D-amino acids are abundant in natural proteins and peptides,² such as cone snail venoms,³ the murein of bacteria,⁴ fungus-produced antibiotics,^{5,6} etc., and it is estimated that the D-amino acid market will reach \$ 3.7 billion by the year 2017.⁷ Therefore, molecular simulations must be able to treat polypeptides and proteins containing D-amino acid residues. For all-atom force fields, except those which contain sine torsional terms, there is no problem handling the D configuration; however, coarse-grained force fields require additional parametrization for this purpose. Because the coarse-grained force fields offer an enormous advantage for reducing the cost and broadening of the time step of simulations,^{8,9} it is worth investing this effort in extending the range of application of such approaches. However, most of the known coarse-grained force fields developed for proteins can treat only the natural L-amino acids.

In the past several years, we have been developing the coarse-grained UNited RESidue (UNRES) force field for simulation of protein structure and dynamics.^{10–22} Unlike most of the other force fields, UNRES has been derived^{13,23} and parametrized^{13–15,17,19–22,24} as a potential of mean force of polypeptide chains immersed in water. UNRES was initially used in the prediction of single-chain protein structure as a global minimum in the effective potential-energy surface;^{25–28} then it was

extended to run coarse-grained molecular dynamics^{8,9,29,30} and to study oligomeric proteins.^{31,32} Recently, trans–cis isomerization of peptide groups was introduced into UNRES.²² UNRES has also been applied to study biological problems such as amyloid formation,^{33,34} signaling,³⁵ and chaperone dynamics.³⁶

UNRES proved to be a powerful tool in predicting protein structure, as assessed in the Critical Assessment of Techniques for Protein Structure Prediction (CASP) experiments.^{25,27,28,37,38} In CASP3, our prediction of the 89-residue α -helical protein HDEA from *E. coli* was assessed as the best of all predictions of all groups. For this protein, we predicted the 60-residue C-terminal fragment within 4.2 Å C^{α} RMSD.²⁶ The version of the UNRES force field used at that time treated α -helical proteins only because it did not contain sufficient correlation terms. The correlation terms were introduced prior to the CASP4 exercise, and we were able to predict fragments of proteins with β -structure. We also predicted the complete structure of a 70-residue protein, bacteriocin S (target T0102).²⁷ During the CASP5 exercise in 2004, we predicted nearly the complete structure (203 out of 235-residues) of a six- α -helix bundle protein (target T0198), complete or nearly complete structures of three other smaller α and α + β proteins, and significant portions of structures of other proteins.²⁸ In CASP9, we

Received: July 2, 2012

Published: September 5, 2012

obtained³⁸ the longest of all groups matching segments for the following three targets: T0534-D2 (156 residues; α -helical structure), T0537-D2 (31 residues; $\alpha + \beta$ structure), and T0578-D1 (208 residues; $\alpha + \beta$ structure), where D1 and D2 denote domains 1 and 2, respectively.

In this work, we extend the UNRES force field to treat systems with D-amino acid residues. For this purpose, we revise the local potentials that depend on amino acid chirality, namely the virtual-bond torsional and double torsional potentials, the side-chain local potentials, and the correlation potentials. Of those, introduction of the torsional, double-torsional, and virtual-bond-angle potentials (which also depend on the adjacent virtual-bond-dihedral angles) requires reparameterization of the force field, while the other potentials for D-amino acid residues are obtained in a trivial manner by applying symmetry operations to the respective potentials for the natural L-amino acid residues, as shown in section 2.2. In this work, we have determined the torsional and double-torsional potentials for D-amino acid residues. The virtual-bond-angle potentials for D-amino acid residues will be introduced later.

2. METHODS

2.1. UNRES Representation of Polypeptide Chain. In the UNRES model,^{10–13,15–23} a polypeptide chain is represented as a sequence of α -carbon (C^α) atoms with attached united side chains (SC's) and united peptide groups (p's) positioned halfway between two consecutive C^α 's. Only the united side chains and united peptide groups act as interaction sites, while the C^α atoms assist only in the definition of geometry (Figure 1). The effective energy function is defined as the restricted free energy (RFE) or the potential of mean force (PMF) of the chain constrained to a given coarse-grained conformation along with the surrounding solvent.^{13,18,23} This effective energy function is expressed by eq 1.

$$\begin{aligned}
 U = & w_{SC} \sum_{i < j} U_{SC,SC_i} + w_{SCp} \sum_{i \neq j} U_{SC,p_j} + w_{pp}^{VDW} \sum_{i < j-1} U_{pp_j}^{VDW} \\
 & + w_{pp2}^{el} f_2(T) \sum_{i < j-1} U_{pp_j}^{el} + w_{torf_2}(T) \sum_i U_{tor}(\gamma_i) \\
 & + w_{torf_3}(T) \sum_i U_{tor}(\gamma_i, \gamma_{i+1}) + w_b \sum_i U_b(\theta_i) \\
 & + w_{rot} \sum_i U_{rot}(\alpha_{SC_i}, \beta_{SC_i}) + w_{bond} \sum_i U_{bond}(d_i) \\
 & + w_{corr3}^{(3)} f_3(T) U_{corr}^{(3)} + w_{corr4}^{(4)} f_4(T) U_{corr}^{(4)} + w_{corr5}^{(5)} f_5(T) U_{corr}^{(5)} \\
 & + w_{turn6}^{(6)} f_6(T) U_{turn}^{(6)} + w_{turn3}^{(3)} f_3(T) U_{turn}^{(3)} + w_{turn4}^{(4)} f_4(T) U_{turn}^{(4)} \\
 & + w_{turn6}^{(6)} f_6(T) U_{turn}^{(6)}
 \end{aligned} \quad (1)$$

where the U 's are energy terms, θ_i is the backbone virtual-bond angle, γ_i is the backbone virtual-bond-dihedral angle, α_i and β_i are the angles defining the location of the center of the united side chain of residue i (Figure 1), and d_i is the length of the i th virtual bond, which is either a $C^\alpha \cdots C^\alpha$ virtual bond or $C^\alpha \cdots SC$ virtual bond. Each energy term is multiplied by an appropriate weight, w_x , and the terms corresponding to factors of order higher than 1 are additionally multiplied by the respective temperature factors which were introduced in our recent work¹⁶ and which reflect the dependence of the first generalized-cumulant term in those factors on temperature, as discussed in refs 16 and 39. The factors f_n are defined by eq 2.

$$f_n(T) = \frac{\ln[\exp(1) + \exp(-1)]}{\ln\{\exp[(T/T_0)^{n-1}] + \exp[-(T/T_0)^{n-1}]\}} \quad (2)$$

where $T_0 = 300$ K.

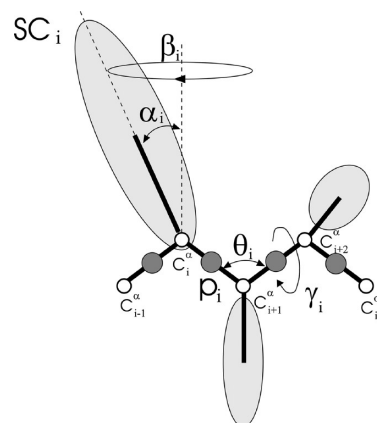


Figure 1. The UNRES model of polypeptide chains. The interaction sites are peptide-group centers (p), and side-chain centers (SC) attached to the corresponding α -carbons with different $C^\alpha \cdots SC$ bond lengths, d_{SC} . The peptide groups are represented as dark gray circles, and the side chains are represented as light gray ellipsoids of different size. The α -carbon atoms are represented by small open circles. The geometry of the chain can be described either by the virtual-bond vectors dC_i (from C_i^α to C_{i+1}^α), $i = 1, 2, \dots, n-1$, and dX_i (from C_i^α to SC_i), $i = 2, \dots, n-1$, represented by thick lines, where n is the number of residues, or in terms of virtual-bond lengths, backbone virtual-bond angles θ_i , $i = 1, 2, \dots, n-2$; backbone virtual-bond-dihedral angles γ_i , $i = 1, 2, \dots, n-3$; and the angles α_i and β_i , $i = 2, 3, \dots, n-1$ that describe the location of a side chain with respect to the coordinate frame defined by C_{i-1}^α , C_i^α , and C_{i+1}^α .

The term U_{SC,SC_i} represents the mean free energy of the hydrophobic (hydrophilic) interactions between the side chains, which implicitly contains the contributions from the interactions of the side chain with the solvent. The term U_{SC,p_i} denotes the excluded-volume potential of the side-chain–peptide-group interactions. The peptide-group interaction potential is split into two parts: the Lennard-Jones interaction energy between peptide-group centers ($U_{pp_j}^{VDW}$) and the average electrostatic energy between peptide-group dipoles ($U_{pp_j}^{el}$); the second of these terms accounts for the tendency to form backbone hydrogen bonds between peptide groups p_i and p_j . The terms U_{tor} , U_{tor_d} , U_b , U_{rot} , and U_{bond} are the virtual-bond-dihedral angle torsional terms, virtual-bond dihedral angle double-torsional terms, virtual-bond angle bending terms, side-chain rotamer, and virtual-bond-deformation terms; these terms account for the local properties of the polypeptide chain. The terms $U_{corr}^{(m)}$ represent correlation or multibody contributions from the coupling between backbone-local and backbone-electrostatic interactions (represented by the interactions between peptide-group dipoles^{13,15}), and the terms $U_{turn}^{(m)}$ are correlation contributions involving m consecutive peptide groups. They are, therefore, termed turn contributions. The multibody terms are indispensable for reproduction of regular α -helical and β -sheet structures.^{13,23,40}

The energy-term weights are determined by force-field calibration to reproduce the structure and folding thermodynamics of selected training proteins.^{16,41} The force field calibrated with the 1GAB protein¹⁶ was used in the present study.

The present UNRES is parametrized for L-amino acid residues. To extend it to D-amino acid residues, the terms that depend on chirality must be changed. These are the torsional (U_{tor}), double-torsional (U_{tor_d}), side-chain-rotamer (U_{rot}), correlation (U_{corr}), and virtual-bond-angle (U_b) potentials.

In the subsequent sections, we will describe the modification of these potentials for D-amino acid residues. Because the torsional and double-torsional potentials require the greatest effort, we will devote most of the description to these potentials.

2.2. Effect of Amino Acid Chirality on the Side-Chain-Rotamer and Correlation Potentials. As mentioned in section 2.1, amino acid chirality affects the torsional, double-torsional, side-chain rotamer, virtual-bond-bending, and correlation potentials. We begin the discussion with the side-chain rotamer potentials, U_{rot} because their dependence on chirality is straightforward. The latest potentials derived in ref 19 have the form of polynomials in the local coordinates of a side chain, in which, for the i th residue, the x axis is the bisector of the $C_{i-1}^\alpha \cdots C_i^\alpha \cdots C_{i+1}^\alpha$, virtual-bond angle; the y axis is perpendicular to the x axis, lies in the plane of the three C^α atoms, and runs in the direction from C_{i-1}^α to C_{i+1}^α ; and the z axis forms a right-handed orthogonal coordinate system with the x and y axes (see Figure 1 in ref 19 for the illustration of the definition of the local-coordinate system). The change of chirality is achieved just by reflection of a side-chain center in the xy plane; consequently, the sign at all coefficients that pertain to the local z coordinate in eq 3 of ref 19 must be changed. Similarly, the D chirality can be introduced in the older (statistical) side-chain-rotamer potentials¹² by changing the sign of the coefficients in eq 8 of ref 12 corresponding to odd powers of the polar angle β , which changes sign when the chirality is changed.

Changing chirality is also straightforward for the correlation potential, $U_{\text{corr}}^{(m)}$. These potentials contain the coefficients of the second-order Fourier expansion of the local-interaction-energy surfaces, e_X , where X denotes a basic type of amino acid residue (glycine, alanine, and proline, where alanine represents all nonglycine and nonproline residues) in terms of the angles $\lambda^{(1)}$ and $\lambda^{(2)}$ for rotation of the peptide groups about the $C^\alpha \cdots C^\alpha$ virtual bonds of terminally blocked amino acid residues,¹³ as given by eq 3.

$$e_X(\lambda^{(1)}, \lambda^{(2)}) = \mathbf{b}_{1X}^T \mathbf{u} + \mathbf{b}_{2X}^T \mathbf{v} + \mathbf{u}^T \mathbf{C}_X \mathbf{u} + \mathbf{v}^T \mathbf{D}_X \mathbf{v} + \mathbf{u}^T \mathbf{E}_X \mathbf{v} + \dots \quad (3)$$

with

$$\mathbf{u} \equiv \mathbf{u}(\lambda^{(1)}) = \begin{pmatrix} \cos \lambda^{(1)} \\ \sin \lambda^{(1)} \end{pmatrix}, \mathbf{v} \equiv \mathbf{v}(\lambda^{(2)}) = \begin{pmatrix} \cos \lambda^{(2)} \\ \sin \lambda^{(2)} \end{pmatrix} \quad (4)$$

where \mathbf{b}_{1X} and \mathbf{b}_{2X} are the two-dimensional vectors of first-order Fourier coefficients and the matrices \mathbf{C}_X , \mathbf{D}_X , and \mathbf{E}_X are the 2×2 matrices of second-order Fourier coefficients of the second-order expansion of the energy surface in the angles $\lambda^{(1)}$ and $\lambda^{(2)}$ for residue X . Additionally, \mathbf{C}_X represents symmetric traceless matrices; therefore, $c_{X,22} = -c_{X,11}$ and $d_{X,22} = -d_{X,11}$.

The energy surface of a D-amino acid residue, e_X , where the lowercase letter denotes the D chirality, $x = \text{Ala}_D$ or Pro_D (because glycine does not have a side chain and, consequently, is not chiral), is obtained from that of the corresponding L-amino acid residue by applying the inversion operation to the angular variables, as given by eq 5.

$$e_x(\lambda^{(1)}, \lambda^{(2)}) = e_X(-\lambda^{(1)}, -\lambda^{(2)}) \\ = \mathbf{b}_{1X}^T \mathbf{u} + \mathbf{b}_{2X}^T \mathbf{v} + \mathbf{u}^T \mathbf{C}_X \mathbf{u} + \mathbf{v}^T \mathbf{D}_X \mathbf{v} + \mathbf{u}^T \mathbf{E}_X \mathbf{v} + \dots \quad (5)$$

with

$$\mathbf{b}_{1x} = \mathbf{U} \mathbf{b}_{1X} = \begin{pmatrix} b_{11X} \\ -b_{21X} \end{pmatrix}, \mathbf{b}_{2x} = \mathbf{U} \mathbf{b}_{2X} = \begin{pmatrix} b_{12X} \\ -b_{22X} \end{pmatrix} \\ \mathbf{C}_x = \mathbf{U} \mathbf{C}_X \mathbf{U}^T = \begin{pmatrix} c_{X11} & -c_{X12} \\ -c_{X12} & c_{X22} \end{pmatrix}, \\ \mathbf{D}_x = \mathbf{U} \mathbf{D}_X \mathbf{U}^T = \begin{pmatrix} d_{X11} & -d_{X12} \\ -d_{X12} & d_{X22} \end{pmatrix}, \\ \mathbf{E}_x = \mathbf{U} \mathbf{E}_X \mathbf{U}^T = \begin{pmatrix} e_{X11} & -e_{X12} \\ -e_{X21} & e_{X22} \end{pmatrix} \quad (6)$$

where \mathbf{U} is the matrix for reflection in the $C_{i-1}^\alpha \cdots C_i^\alpha \cdots C_{i+1}^\alpha$ plane.

$$\mathbf{U} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad (7)$$

Thus, introducing the D chirality in the correlation terms is achieved by changing the signs of the coefficients of the second-order expansion of e_X at the terms with odd sine powers.

2.3. Determination of Torsional and Double-Torsional Potentials. As opposed to U_{rot} the virtual-bond-torsional, U_{tor} , and double-torsional potentials, U_{tord} , that contain residues with mixed chirality must be redetermined. The procedure developed in our earlier work¹⁴ was used, in which the respective potentials of mean force are calculated first from the energy maps of terminally blocked amino acid residues, and then a one- (for U_{tor}) or two-dimensional (for U_{tord}) Fourier series is fitted to them. The harmonic-entropy contribution is also considered. The potentials of mean force corresponding to U_{tor} and U_{tord} are thus defined by eqs 8 and 9, respectively. The terminally blocked di- and tripeptide systems used for the derivation of the torsional and double-torsional potentials are illustrated in Figure 2A and B, respectively.

$$U_{XY}(\gamma) = -\beta^{-1} \ln \left\{ \frac{1}{(2\pi)^3} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \right. \\ \times [\det \mathbf{H}^*(\lambda_1, \gamma - \pi - \lambda_2) \times \det \mathbf{H}^*(\lambda_2, \lambda_3)]^{-1/2} \\ \times \exp[-\beta[e_X(\lambda_1, \gamma - \pi - \lambda_2) + e_Y(\lambda_2, \lambda_3)]] d\lambda_1 d\lambda_2 d\lambda_3 \} \\ + \beta^{-1} \ln \left\{ \frac{1}{(2\pi)^3} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} [\det \mathbf{H}^*(\lambda_1, \lambda_2)]^{-1/2} \right. \\ \times \exp[-\beta e_X(\lambda_1, \lambda_2)] d\lambda_1 d\lambda_2 \} \\ + \beta^{-1} \ln \left\{ \frac{1}{(2\pi)^3} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} [\det \mathbf{H}^*(\lambda_2, \lambda_3)]^{-1/2} \right. \\ \times \exp[-\beta e_Y(\lambda_2, \lambda_3)] d\lambda_1 d\lambda_2 \} \quad (8)$$

where e_X and e_Y are the energy surfaces for the terminally blocked residues of types X and Y , respectively, \mathbf{H}^* is the energy Hessian computed over all variables except for the angles λ_i and λ_j (the terms with $\det \mathbf{H}^*$ account for the harmonic-entropy contribution to the PMF), and $\beta = 1/RT$, where R is the universal gas constant and T is the absolute temperature. For $Y = \text{Pro}$, the C-terminal blocking group of the X residue is pyrrolidine; otherwise it is the methyl group.

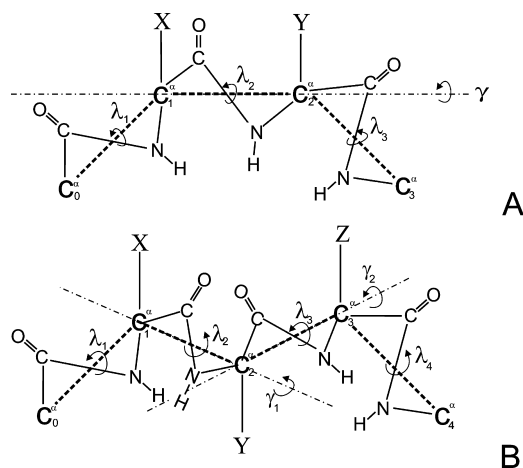


Figure 2. Illustration of the model terminally blocked dipeptides (a) and tripeptides (b) constructed to compute the integrals of eqs 8 and 9. Each of the symbols X, Y, and Z denotes the side chains of Ala_L, Ala_D, Gly, Pro_L, or Pro_D.

$$\begin{aligned}
 U_{XYZ}(\gamma_1, \gamma_2) = & -\beta^{-1} \ln \left\{ \frac{1}{(2\pi)^4} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \right. \\
 & [\det \mathbf{H}^*(\lambda_1, \gamma_1 - \pi - \lambda_2) \times \det \mathbf{H}^*(\lambda_2, \gamma_2 \\
 & \quad - \pi - \lambda_3) \times \det \mathbf{H}^*(\lambda_3, \lambda_4)]^{-1/2} \\
 & \times \exp[-\beta[e_X(\lambda_1, \gamma_1 - \pi - \lambda_2) \\
 & \quad + e_Y(\lambda_2, \gamma_2 - \pi - \lambda_3) \\
 & \quad + e_Z(\lambda_3, \lambda_4)]] d\lambda_1 d\lambda_2 d\lambda_3 d\lambda_4 \\
 & \quad \left. - [U_{XY}(\gamma_1) + U_{YZ}(\gamma_2)] \right\} \\
 & + \beta^{-1} \ln \left\{ \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} [\det \mathbf{H}^*(\lambda_1, \lambda_2)]^{-1/2} \right. \\
 & \quad \times \exp[-\beta e_X(\lambda_1, \lambda_2)] d\lambda_1 d\lambda_2 \left. \right\} \\
 & + \beta^{-1} \ln \left\{ \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} [\det \mathbf{H}^*(\lambda_2, \lambda_3)]^{-1/2} \right. \\
 & \quad \times \exp[-\beta e_X(\lambda_2, \lambda_3)] d\lambda_2 d\lambda_3 \left. \right\} \\
 & + \beta^{-1} \ln \left\{ \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} [\det \mathbf{H}^*(\lambda_3, \lambda_4)]^{-1/2} \right. \\
 & \quad \times \exp[-\beta e_X(\lambda_3, \lambda_4)] d\lambda_3 d\lambda_4 \left. \right\}
 \end{aligned} \quad (9)$$

It should be noted that, because the single-torsional potentials U_{XY} and U_{YZ} are subtracted from the RFE of the terminally blocked tripeptide, U_{XYZ} contains only that part of the free energy of the rotation about the $C^\alpha \cdots C^\alpha$ virtual-bond angles that cannot be accounted for by the single-torsional contributions.

The presence of $\gamma_1 - \pi - \lambda_2$ in eqs 8 and 9 and $\gamma_2 - \pi - \lambda_3$ in eq 9 arises from the fact that λ_2 is shared between residues X and Y in the dipeptide and in the tripeptide systems, and λ_3 is shared between residues Y and Z in the tripeptide system. The diabatic energy surface, e_X , of a terminally blocked amino acid residue X can be expressed as a function of two *local* angles $\lambda^{(1)}$ and $\lambda^{(2)}$ for rotation about the consecutive virtual $C^\alpha \cdots C^\alpha$

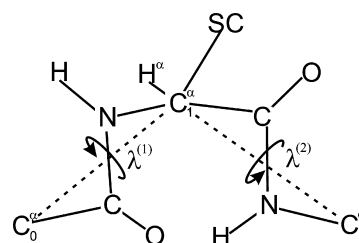


Figure 3. Definition of the dihedral angles $\lambda^{(1)}$ and $\lambda^{(2)}$ for rotation of a peptide unit about the $C^\alpha \cdots C^\alpha$ virtual bonds of a peptide unit.⁴²

bonds, defined by Nishikawa et al.⁴² and also shown in Figure 3. The local angles of consecutive residues are, in turn, related to the chain-wide angles $\lambda_1, \lambda_2, \dots, \lambda_n$ shown in Figure 2 by eq 10.

$$\begin{aligned}
 \lambda_i^{(1)} &= \lambda_i, \quad i = 1, 2, \dots, n \\
 \lambda_i^{(2)} &= \lambda_{i+1}, \quad i = 1, 2, \dots, n-1 \\
 \lambda_n^{(2)} &= \lambda_n
 \end{aligned} \quad (10)$$

The integrals in eqs 8 and 9 were calculated by numerical quadrature by summing the values of the integrand over the nodes of a multidimensional grid in the angles λ . The grid size was the same as that with which the corresponding potential-energy surfaces were obtained. The presence of a proline residue reduced the dimensions of a grid by one, because the $\lambda^{(1)}$ angle of the proline residue is determined by the requirement to obey the constraints accruing from the presence of the pyrrolidine ring. The presence of proline residues also made it necessary to interpolate linearly between the closest points of the original grid to estimate the energy and the logarithm of the determinant of Hessian values for residues preceding proline; otherwise, all energies were taken directly from the *ab initio* values computed at grid points. This was caused by the fact that $\lambda^{(1)}$ of a proline residue is a function of $\lambda^{(2)}$.⁴³ Thus, even if $\lambda^{(2)}$ is on the grid, $\lambda^{(1)}$ need not be, if the residue is proline. The angle $\lambda^{(1)}$ is, in turn, related by eq 10 to $\lambda^{(2)}$ of the preceding residue; therefore, the last angle also, generally, takes values outside the grid points. For nonproline residues, $\lambda^{(1)}$ is unrestricted. Consequently, the virtual-bond dihedral angles γ are from a grid of the same spacing as the angles λ , and, by eq 10, all points taken to evaluate the integrals that occur in eqs 8 and 9 are on the grid.

In the UNRES force field, the torsional potentials are expressed as a one- and two-dimensional Fourier series, respectively (eqs 11 and 12)

$$U_{XY}(\gamma) = a_0 + \sum_{i=1}^{n1} a_i \cos(i\gamma) + b_i \sin(i\gamma) \quad (11)$$

$$\begin{aligned}
 U_{XYZ}(\gamma_1, \gamma_2) = & c_0 + \sum_{i=1}^{n2} c_i^{(1)} \cos(i\gamma_1) + d_i^{(1)} \sin(i\gamma_1) \\
 & + \sum_{i=1}^{n2} c_i^{(2)} \cos(i\gamma_2) + d_i^{(2)} \sin(i\gamma_2) \\
 & + \sum_{i=1}^{n2} \sum_{j=1}^{i-1} e_{ij} \cos[j\gamma_1 + (i-j)\gamma_2] + f_{ij} \cos[j\gamma_1 - (i-j)\gamma_2] \\
 & \times g_{ij} \sin[j\gamma_1 + (i-j)\gamma_2] + h_{ij} \sin[j\gamma_1 - (i-j)\gamma_2]
 \end{aligned} \quad (12)$$

where a_i, b_i, \dots, h_{ij} are coefficients and $n1$ and $n2$ are the orders of the expression; we used $n1 = 6$ and $n2 = 8$. The coefficients in eqs 11 and 12 were calculated by fitting the U_{XY} and U_{XYZ}

surfaces (obtained by numerical integration) to eqs 11 and 12, respectively. As in previous work,¹⁴ the target function F consisted of the sum of the squares of the errors over the grid points in γ_1 and γ_2 and an entropy-like or smoothing term preventing too large a variation of the fitted functional expression between the points of the grid (eq 13).

$$F = \sum_{i=1}^N [U_i^{\text{num}} - U_i(X)]^2 + \alpha \int_{\gamma} \Delta U(\gamma) \exp[\Delta U(\gamma)] d\gamma \quad (13)$$

where U_i^{num} denotes the value of a torsional or a double-torsional potential at grid point i , obtained by numerical integration of the energy surfaces of terminally blocked amino acid residues (eq 8 or 9), $U_i(X)$ denotes the value of U calculated by using the Fourier series (eq 11 or 12), $\Delta U(\gamma)$ is the difference between the U expressed by eq 11 or 12 and the value at that point obtained by linear interpolation between the grid points, and α is the weight of the Shannon-entropy term; we used $\alpha = 0.2$. The value of α was set after several trial calculations to provide both smooth potentials and a good fit to numerically calculated PMF values. The target function of eq 13 was minimized with respect to the coefficients $a_0, a_{1,\dots}, b_6$ (for single-torsional potentials; eq 11) or $c_0, c_{1,\dots}, h_{8,8}$ (for double-torsional potentials; eq 12) by using the Secant Unconstrained Minimization Solver (SUMSL) procedure.⁴⁴ The Shannon-entropy term (the last term in eq 13) prevents the fitted functions from oscillating between grid points and, therefore, enables one to obtain quite smooth functions despite using a large number of Fourier terms. It takes a zero value when the function is linear between the grid points and increases considerably if the function oscillates between the grid points.

It can easily be demonstrated that the symmetry relations expressed by eqs 14 and 15 hold; there are only 2×13 independent torsional potentials and 2×63 independent double-torsional potentials, respectively. The two sets of torsional and double-torsional potentials appear because a regular or proline residue, respectively, can follow a di- or tripeptide fragment in a polypeptide chain; if the following residue is proline, the model system implemented to derive the respective potential is blocked by the N-pyrrolidine group.

$$U_{\text{tor};\overline{XY}}(\gamma) = U_{\text{tor};XY}(-\gamma) \quad (14)$$

$$U_{\text{tor};\overline{XYZ}}(\gamma_1, \gamma_2) = U_{\text{tor};XYZ}(-\gamma_1, -\gamma_2) \quad (15)$$

where X, Y, and Z denote amino acid residue types of a given chirality (L or D) and a bar over a symbol denotes the change of enantiomer chirality; this operation does not affect glycine.

2.4. Calculations of the Energy Surfaces for Terminally Blocked Amino Acid Residues. In our previous work,¹⁴ we computed the energy maps of terminally blocked amino acid residues by using *ab initio* molecular quantum mechanics with the 6-31G(d,p) basis set. Because less powerful resources were available at that time, the energy was minimized in the Restricted Hartree–Fock (RHF) scheme, and subsequently, the single-point MP2 energy correction was computed for the energy-minimized conformation. Moreover, interfacing a residue to the following proline residue was modeled by blocking it with an N',N'-dimethyl group. In this work, all energy maps were recalculated by minimizing the energy in the MP2/6-31G(d,p) scheme; additionally, as mentioned in section 2.3, we used the C-terminal pyrrolidine group and not the N,N-dimethylamine blocking group for

residues preceding proline. Consequently, all diabatic energy maps of *trans*-N-acetyl-alanyl-N'-methyl amide, *trans*-N-acetyl-alanyl-N'-pyrrolidylamide, *trans*-N-acetyl-glycyl-N'-methyl amide, *trans*-N-acetyl-glycyl-N'-pyrrolidylamide, *trans*-N-acetyl-prolyl-N'-methyl amide, and *trans*-N-acetyl-prolyl-N'-pyrrolidylamide were computed. The angles $\lambda^{(1)}$ and $\lambda^{(2)}$ were used as grid variables; both varied from -180° to 180° with a 15° step. The energy was minimized with respect to all other degrees of freedom. Energy Hessians (second-derivative matrices or force-constant matrices) were calculated at every grid point to determine the harmonic-entropy contribution to the PMF (eqs 8 and 9).

3. RESULTS AND DISCUSSION

3.1. Energy Maps of Terminally Blocked Amino Acid Residues. The potential-energy surfaces of *trans*-N-acetyl-L-alanyl-N'-methyl amide, *trans*-N-acetyl-L-alanyl-N'-pyrrolidylamide, *trans*-N-acetyl-glycyl-N'-methyl amide, *trans*-N-acetyl-glycyl-N'-pyrrolidylamide, *trans*-N-acetyl-L-prolyl-N'-methyl amide, and *trans*-N-acetyl-L-prolyl-N'-pyrrolidylamide are shown in the Figures 4 and 5, respectively. The potential-

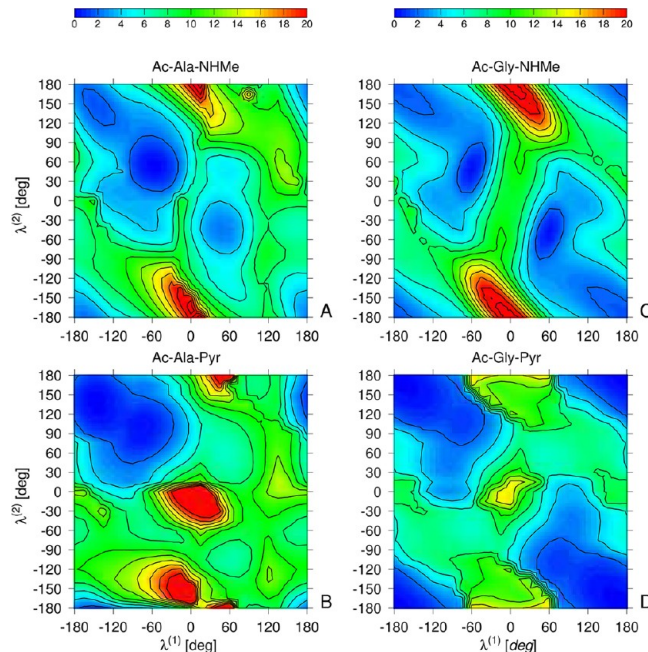


Figure 4. Contour plots of the diabatic potential-energy surfaces of Ac-Ala_L-NHMe (A), Ac-Ala_L-Pyr (B), Ac-Gly-NHMe (C), and Ac-Gly-Pyr (D) obtained with the MP2/6-31G(d,p) *ab initio* method. Energies (kcal/mol) are expressed as relative values with respect to the global-minimum value for a given system, which occurs at ($\lambda^{(1)} = -60^\circ$, $\lambda^{(2)} = 60^\circ$) for Ac-Ala_L-NHMe, ($\lambda^{(1)} = -60^\circ$, $\lambda^{(2)} = 60^\circ$) or ($\lambda^{(1)} = 60^\circ$, $\lambda^{(2)} = -60^\circ$) for Ac-Gly-NHMe, ($\lambda^{(1)} = -135^\circ$, $\lambda^{(2)} = 135^\circ$) for Ac-Ala_L-Pyr, and ($\lambda^{(1)} = -180^\circ$, $\lambda^{(2)} = 180^\circ$) for Ac-Gly-Pyr (d). The energy color scale (kcal/mol) is on the top of the figure.

energy surfaces of terminally blocked D-amino acid residues can be obtained from those shown in Figures 4 and 5 by applying the inversion-symmetry operation (eq 5). The potential-energy maps for the Ac-Ala_L-NHMe and Ac-Gly-NHMe systems (Figure 4A and C) resemble those obtained in our earlier study (Figure 4a and c in ref 14). It should be noted that, in that study,¹⁴ we included the second-order electron-correlation contribution to the energy only as a correction calculated for RHF-energy-minimized conformations. However, the Ac-Gly-NMe₂ and Ac-Ala_L-NMe₂ potential-energy surfaces used in our

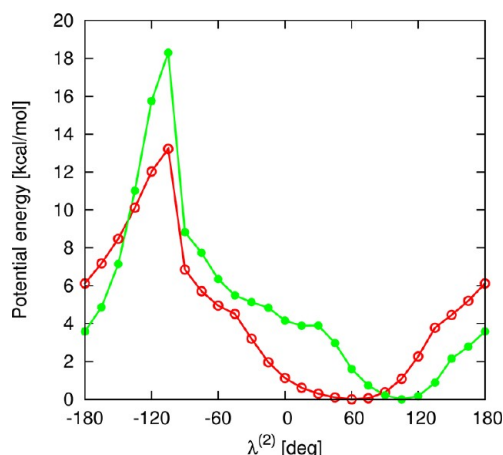


Figure 5. Relative energy profiles (kcal/mol) of Ac-Pro_L-NHMe (red line) and Ac-Pro_L-Pyr (green line) in the $\lambda^{(2)}$ angle calculated with the MP2/6-31G(d,p) ab initio method.

previous study¹⁴ to represent the residues preceding proline (Figure 5 in ref 14) exhibit some differences with respect to the Ac-Gly-Pyr and Ac-Ala_L-Pyr potential-energy surfaces (Figure 4B and D). The most significant difference is that the high-energy regions $\lambda^{(1)} = 0^\circ$ and $\lambda^{(2)} = 0^\circ$ and $\lambda^{(1)} = 0^\circ$ and $\lambda^{(2)} = 180^\circ$ are noticeably smaller for both Ac-Gly-Pyr and Ac-Ala_L-Pyr compared to those for Ac-Gly-NMe₂ and Ac-Ala_L-NMe₂, respectively. This feature results from the lack of specific hydrogen-bonding interactions, especially those in the C_{eq}⁷ conformation, which create narrow and deep energy minima. With the bulky pyrrolidine group in place of a C-terminal amide-hydrogen atom, a broader but shallower minimum is present. Moreover, the high-energy regions of Ac-Gly-Pyr and Ac-Ala_L-Pyr have a more complex shape than those of Ac-Gly-NMe₂ and Ac-Ala_L-NMe₂, which is understandable in view of the more complex shape of the Pyr blocking group compared to that of the NMe₂ group.

Another difference between the potential-energy maps of NMe₂- and Pyr-blocked residues is the shape and topology of the energy minima. In Ac-Ala_L-NMe₂, two clearly separated minima at $(\lambda^{(1)} = -60^\circ, \lambda^{(2)} = 90^\circ)$ and $(\lambda^{(1)} = -150^\circ, \lambda^{(2)} = 150^\circ)$, respectively, can be observed, which correspond to the C_{eq}⁷ and extended conformation, respectively (it should be noted, though, that the hydrogen bond characteristic of the C_{eq}⁷ conformation is absent because of replacement of the amide-hydrogen atom with a methyl group). For Ac-Ala_L-Pyr, these two minima are shifted to $(\lambda^{(1)} = -75^\circ, \lambda^{(2)} = 90^\circ)$ and $(\lambda^{(1)} = -135^\circ, \lambda^{(2)} = 135^\circ)$ and nearly merged to create a broad low-energy region comprising the C_{eq}⁷ and extended conformations. This feature is also likely to result from use of the full MP2/6-31G(d,p) energy function in the present study.

Yet bigger differences, resulting from the use of a different C-terminal blocking group and minimization of a different energy function, can be observed between Ac-Pro_L-NMe₂ (Figure 6 in ref 14) and Ac-Pro_L-Pyr (Figure 5). The Ac-Pro_L-Pyr potential-energy plot exhibits only one maximum, while that of Ac-Pro_L-NMe₂ has two maxima. The maximum that appeared at $\lambda^{(2)} = 30^\circ$ for Ac-Pro_L-NMe₂ has coalesced with the minimum at $\lambda^{(2)} \approx 60^\circ$ into an inflection point at $\lambda^{(2)} \approx 20^\circ$. These differences and inclusion of the harmonic-entropy contribution result in different torsional and double torsional potentials involving a proline residue.

3.2. Torsional Potentials. The 13 independent torsional potentials (see section 2.3) for each of the C-terminal blocking group (NHMe or Pyr; a total of 26 potentials) calculated from the MP2/6-31G(d,p) potential-energy maps by using eq 8 are shown in Figure 6. Subsequently, Fourier series (eq 11) were fitted to them. The Fourier coefficients are collected in Table S1 of the Supporting Information.

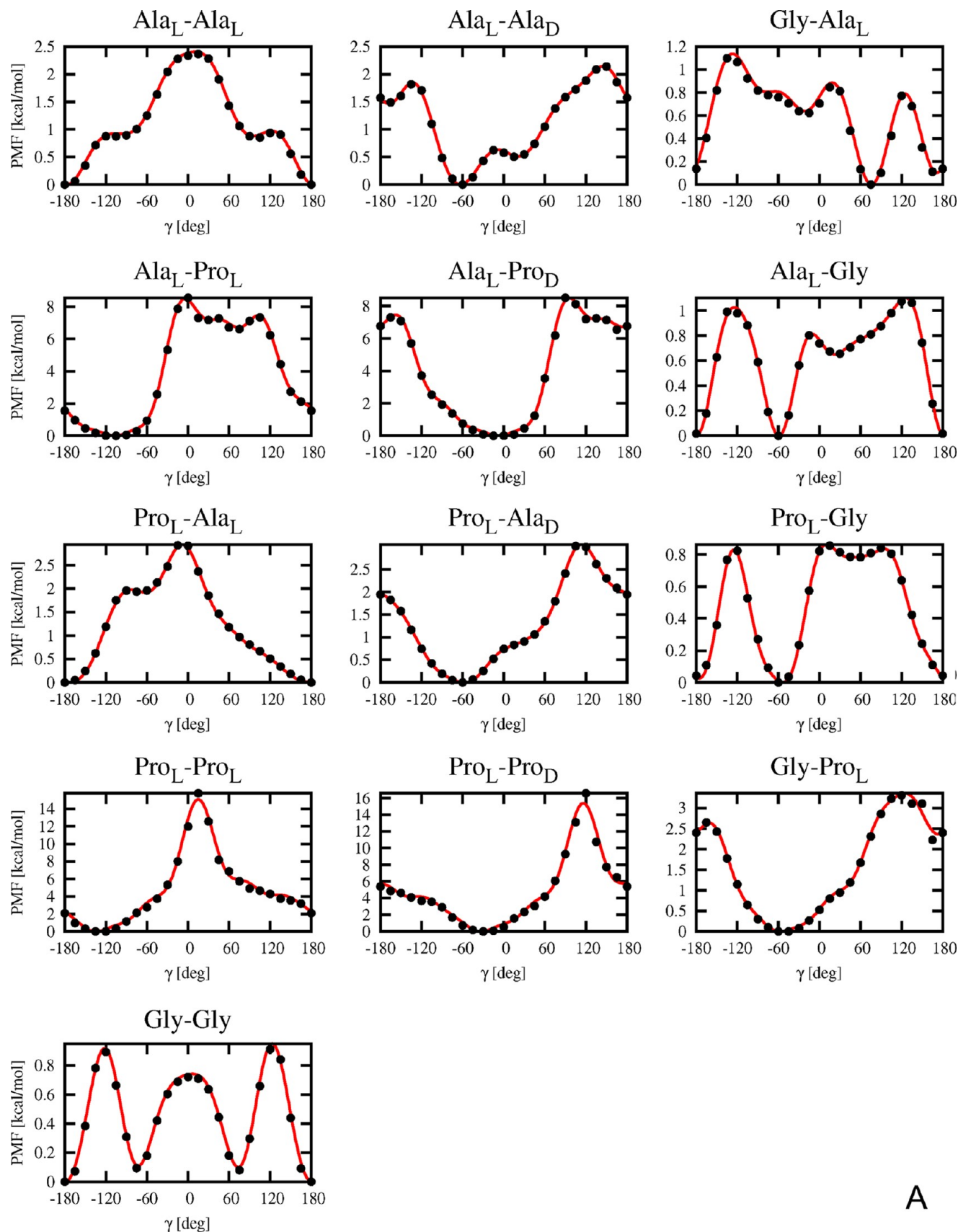
Despite the more elaborate methodology for calculating the energy surfaces of terminally blocked amino acid residues, the torsional potentials of the L-amino acid residue sequences blocked with the NHMe group at the C-terminus (Figure 6A) resemble the previously determined potentials (Figure 8a in ref 14). As previously observed,¹⁴ three minima are present in the Gly-Gly potentials at $\gamma = 180^\circ$ and $\gamma = \pm 60^\circ$, respectively. These minima correspond to extended, type I' or type II' (for $\gamma = -60^\circ$) and type I or type II (for $\gamma = 60^\circ$) β turns, respectively. As in our previous work,¹⁴ the only minimum of the Ala_L-Ala_L torsional potential corresponds to the extended structure (Figure 7A). This feature results from the fact that, for the extended structure, both residues can form the energetically favorable C_{eq}⁷ conformation with a C=O...H-N hydrogen bond, as illustrated in Figure 7A. On the other hand, both residues can also be in the extended conformation, which forms a broad low-energy region (Figure 4A and C) and, consequently, has high entropy. No minimum corresponding to an α -helical structure is present ($\gamma \approx 60^\circ$), which is understandable because that structure is stabilized only by interactions between the second- and third-neighbor residues, which are not included in the torsional potentials.

As mentioned in section 2.3, the torsional potentials for the system in which both amino acid residues are in the D configuration can be obtained by inverting the potentials for L-amino acid residue sequences about $\gamma = 0^\circ$ (eq 14); the same holds for the torsional potentials of the Gly-X and X-Gly sequences, where X denotes alanine or proline. Conversely, the potentials for the sequences of mixed-chirality dipeptides exhibit a different pattern, as shown in Figure 6. The energy minima occur for small absolute values of γ regardless of whether Ala or Pro is involved and regardless of the type of the C-terminal blocking group. This feature arises from the fact that, for small values of γ , both residues can occur in the energetically favorable C_{eq}⁷ conformation (for the L configuration) or C_{ax}⁷ conformation (for the D configuration) or in the portion of the extended-structure region adjacent to the C_{eq}⁷ (C_{ax}⁷) region. This situation is illustrated in Figure 7B for the Ala_L-Ala_D dipeptide. Moreover, the presence of three minima for the Ac-Gly-Gly-NHMe system (Figure 6A) can also now be explained because glycine is an achiral amino acid and any sequence of C_{eq}⁷ and C_{ax}⁷ conformations corresponds to a low energy of intraresidue interaction. It should be noted that four minima are also present for the Ala_L-Ala_D-NHMe systems; the global minimum occurs at $\gamma \approx -60^\circ$, and the second one occurs at $\gamma \approx 30^\circ$. Of course, for the Ac-Ala_D-Ala_L-NHMe sequence, these minima occur at $\gamma \approx 60^\circ$ and $\gamma \approx -30^\circ$, respectively, by eq 14.

The location of the minima for the alternating sequences of L- and D-amino acid residues has very interesting structural implications, namely twisting of a pair of a two-strand β sheet into a double helix. This feature arises because not only C⁷ (axial or equatorial) but also extended conformations, which are capable of forming interstrand hydrogen bonds, contribute to the partition function at the minima of the torsional potentials for the DL- or LD-chirality sequences. The doubly helical structure of gramicidin D (PDB code: 1AL4; Figure 8A

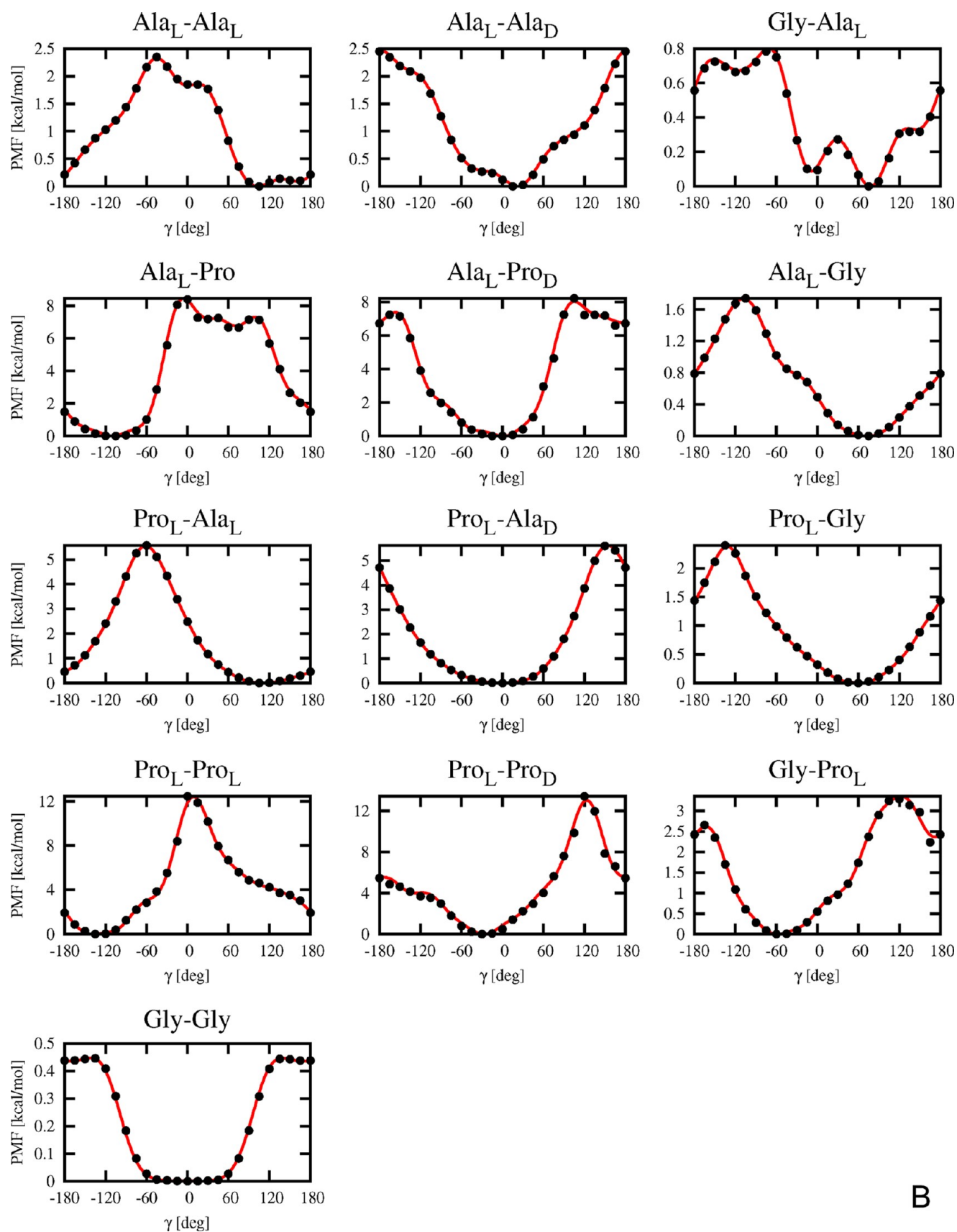
and B), which is a heterodimer of two chains, each with alternating D- and L-residue sequences, confirms this observation. The native-like structure (with 3.7 Å C $^{\alpha}$ RMSD) remained

the lowest-energy structure after 40 000 000 steps of multiplexed replica-exchange molecular dynamics simulations run at 16 temperatures from 250 to 400 K, four replicas per



A

Figure 6. continued



B

Figure 6. Plots of the torsional potentials $U_{XY}(\gamma)$, where X and Y denote Gly, Ala_L, Pro_L, Ala_D, or Pro_D, calculated from the MP2/6-31G(d,p) potential-energy surfaces of terminally blocked amino acid residues by using eq 8 (circles) and Fourier fit to these points (eq 11; lines). Panel A: Ac-X-Y-NHMe systems. Panel B: Ac-X-Y-Pyr systems.

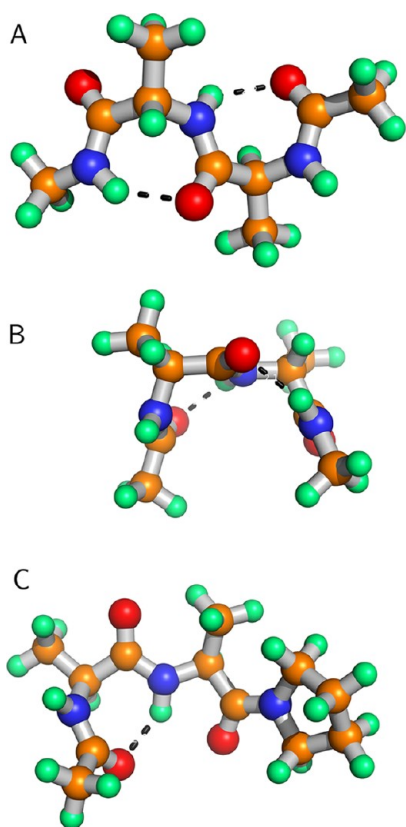


Figure 7. The structures of the dipeptides with the most significant statistical weight in the partition function at the global minimum of the torsional PMF for the (A) Ac-Ala_L-Ala_L-NHMe ($\lambda^{(1)} = -60^\circ$, $\lambda^{(2)} = 60^\circ$, $\lambda^{(3)} = -60^\circ$, $\lambda^{(4)} = 60^\circ$), (B) Ac-Ala_L-Ala_D-NHMe ($\lambda^{(1)} = -60^\circ$, $\lambda^{(2)} = 60^\circ$, $\lambda^{(3)} = 60^\circ$, $\lambda^{(4)} = -60^\circ$), (C) Ac-Ala_L-Ala_L-Pyr ($\lambda^{(1)} = -60^\circ$, $\lambda^{(2)} = 60^\circ$, $\lambda^{(3)} = -135^\circ$, $\lambda^{(4)} = 135^\circ$) systems.

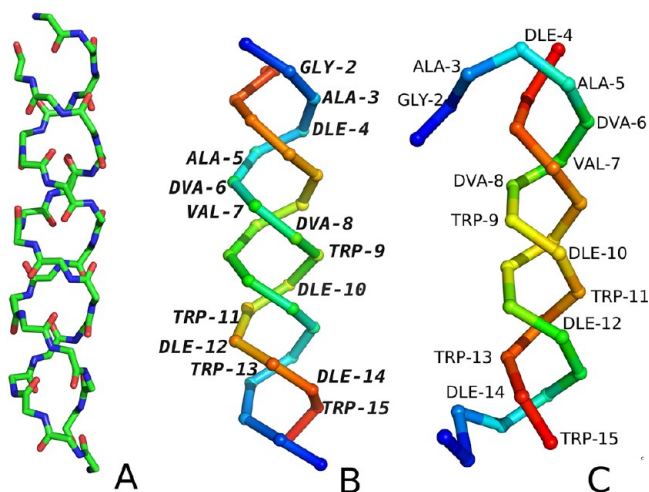


Figure 8. The experimental and computed structures of gramicidin D. (A) The backbone of the experimental 1AL4 structure with all atoms shown to indicate the hydrogen-bonding contacts. (B) The C α trace of the experimental 1AL4 structure. (C) The lowest-energy C α trace of the structures obtained from using UNRES/MREMD simulations starting from the experimental 1AL4 structure. In panels B and C, residues are colored from blue to red from N- to C-terminus.

temperature. The 3.7 Å RMSD results from a shift of the chains in the calculated structure by one turn with respect to each

other, compared to the experimental structure. The calculated structure is shown in Figure 8C.

As can be seen from Figure 6B, the torsional potentials for the Pyr-blocked dipeptides are different from those for NHMe-blocked ones. The most significant difference occurs for the Gly-Gly dipeptide, for which the two minima at $\gamma \approx \pm 60^\circ$ coalesce into a broad minimum region centered at $\gamma = 0^\circ$, and the extended-chain minimum at $\gamma = 180^\circ$ becomes very shallow. Similar differences occur for all glycine-containing systems. For the other systems, the torsional potentials for Pyr-blocked dipeptides have a less pronounced fine structure compared to those for the NHMe-blocked systems. This is understandable because the large blocking group, which does not have a proton donor, prevents the formation of local hydrogen bonds, and thus, their formation cannot be reflected in the minima of the torsional potentials.

An analysis of the plots (Figure 6) also reveals that L-proline is an efficient helix breaker. For the Ala_L-Pro_L-NHMe dipeptide, the torsional-potential values are about 9 kcal/mol in the entire region from $\gamma = 0^\circ$ to $\gamma = 90^\circ$, i.e., for all possible values of γ corresponding to α -helical conformations (Figure 6A). If D-proline is inserted in the sequence, such high values of the torsional potential do not occur. This result contradicts the experimental observation that D-proline is a strong α -helix breaker.^{45,46} Thus, longer-range than intrasidue interactions must determine helix destabilization by D-proline.

3.3. Double-Torsional Potentials. All 126 independent potentials (63 for each type of the C-terminal blocking group) were calculated from the MP2/6-31G(d,p) potential-energy surfaces of terminally blocked glycine, alanine, and proline, by using eq 9. To provide an example, the four $U_{\text{Ala}_L-\text{Ala}_{(D,L)}-\text{Ala}_{(D,L)}-\text{NHMe}(\gamma_1, \gamma_2)$ and the four $U_{\text{Ala}_L-\text{Ala}_{(D,L)}-\text{Ala}_{(D,L)}-\text{Pyr}(\gamma_1, \gamma_2)$ PMF surfaces are shown in Figure 9; the surfaces for the system in which the first residue is Ala_D are related to those by eq 15. The other PMF surfaces are shown in Figure S1 of the Supporting Information. Two-dimensional Fourier series (eq 12) were subsequently fitted to the PMFs; the coefficients are collected in Table S2 of the Supporting Information.

As observed for the torsional potentials (section 3.2), the double-torsional potentials are similar to those derived in our previous work,¹⁴ some differences arising for proline-containing tripeptides and the tripeptides blocked with the pyrrolidine C-terminal group. As for the torsional potentials discussed in the previous section, the difference between PMF minima and maxima is usually larger when the tripeptide is blocked with the pyrrolidine group. Even though the double-torsional potentials account for the third-order correlations, the range of their variation is comparable to that of the torsional potential and they, therefore, cannot be ignored in the UNRES effective energy function.

The global minimum for the Ac-Ala_L-Ala_L-Ala_L-NHMe double-torsional potential (Figure 9A) occurs at ($\gamma_1^{(1)} = 15^\circ$, $\gamma_2^{(2)} = -15^\circ$) and the minimum region extends along the $\gamma_2 = \gamma_1$ axis toward positive values of γ_1 . Thus, the double-torsional potentials for all-L-residue polypeptide chains diminish the propensity toward extended chain conformations and drive the chain at right-handed α -helical conformations. When the L-alanine residue in the middle of the sequence is replaced by D-alanine (Figure 9C), the global minimum is shifted toward $\gamma_1 = \gamma_2 = 180^\circ$, and also a shallower minimum appears at $\gamma_1 = 120^\circ$ and $\gamma_2 = 120^\circ$. Thus, for sequences with alternating chirality, the double-torsional potentials have the opposite effect as that of the

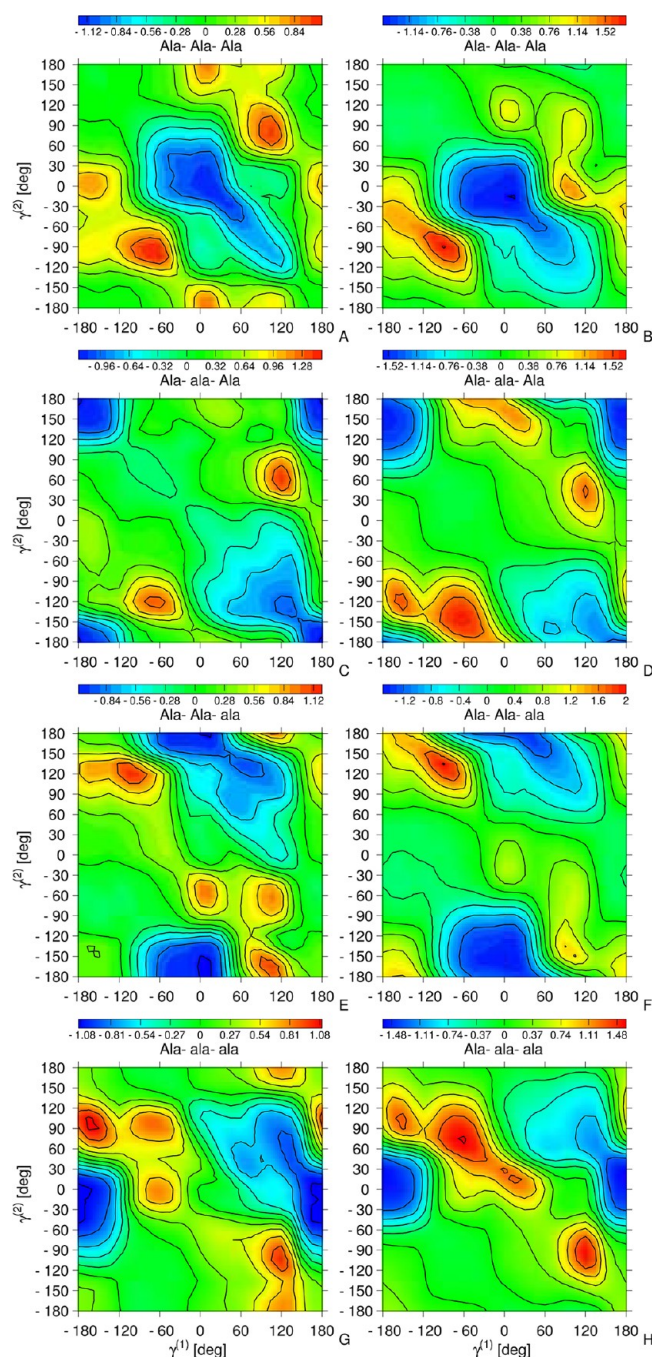


Figure 9. Sample contour plots of the double-torsional potentials for alanine-type tripeptides, $U_{L-Ala-(D,L)-Ala-(D,L)-Ala-NHMe}(\gamma^1, \gamma^2)$ (A, C, E, G) and $U_{L-Ala-(D,L)-Ala-(D,L)-Ala-Pyr}(\gamma^1, \gamma^2)$ (B, D, F, H). The energy scale (kcal/mol) is on the top of the figure.

torsional potentials which favor small values of γ (Figure 6). For the $Ac-Ala_L-Ala_L-Ala_D-NHMe$ system sequence (Figure 9E), the global minimum is located at $\gamma_1 = 15^\circ$, $\gamma_2 = -165^\circ$ and for the $Ac-Ala_L-Ala_D-Ala_D-NHMe$ system (Figure 9G), the global minimum occurs at $\gamma_1 = -15^\circ$, $\gamma_2 = 180^\circ$, as in β -bulge structures.⁴⁷

The double-torsional potentials for the Pyr-blocked systems (Figure 9B, D, and F) are close to those of the NHMe-blocked system except that the minimum regions are less structured.

It should be stressed that both torsional and double torsional potentials are not solely responsible for structure formation,

even at the local level; in particular, the turn-correlation interactions ($U_{turn}^{(m)}$) play a significant role here. The discussion of this and preceding section should, therefore, be considered as the description of the features of the torsional and double-torsional energy surfaces rather than as the description of the structural features of UNRES-generated conformations. For the torsional potentials, it was possible to relate the features of the energy surface to the structure of gramicidin D (Figure 8). However, because the double-torsional potentials make a smaller contribution to the UNRES energy and depend on the geometry of a larger chain fragment, thus making it more probable to ignore other important energy contributions, we refrain from discussing the correspondence between the features of double-torsional energy surfaces and experimental structures.

4. SUMMARY

In this work, the potential-energy surfaces of terminally blocked glycine, L-alanine, and L-proline were calculated at the MP2/6-31G(d,p) *ab initio* level. The C-terminal blocking group was either the N-methylamino group to account for the situation in which the next residue in the chain is not a proline residue or pyrrolidine to account for the connection of the residue to a proline residue next in the amino acid sequence. For Gly and Ala_L , the energy maps were constructed on the grid of the angles $\lambda^{(1)}$ and $\lambda^{(2)}$ for rotation about the $C^\alpha \cdots C^\alpha$ virtual bonds (Figure 3); for L-proline, one-dimensional potential-energy surfaces in $\lambda^{(2)}$ were constructed because, for L-proline, $\lambda^{(1)}$ is a function of $\lambda^{(2)}$ due to the constraint imposed by the presence of the rigid pyrrolidine ring. For each grid point, the energy was minimized with respect to all other degrees of freedom. Energy Hessians were calculated at each grid point to account for harmonic-entropy contributions. On the basis of the calculated potential-energy surfaces, the energy surfaces of D-alanine and D-proline were constructed by applying the inversion operation.

On the basis of the calculated energy surfaces, the torsional and double torsional potentials for rotation about one or two consecutive $C^\alpha \cdots C^\alpha$ virtual-bond axes (defined by the virtual-bond-dihedral angle γ or two consecutive virtual-bond-dihedral angles γ_1 and γ_2), respectively, were calculated, and one- or two-dimensional Fourier series, respectively, were fitted to them to obtain the potentials for use in the UNRES force field. The potentials were calculated for all possible combinations of D- and L-amino acid residues and for both the N-methylamide and pyrrolidine blocking groups. Because the change of the chirality of all residues is equivalent to inversion of the sign of the angle γ or of the angles γ_1 and γ_2 , respectively, there are only 13 and 63 independent torsional and double-torsional potentials for each type of C-terminal blocking group, respectively. For L-amino acid residues and the NHMe C-terminal blocking group, the potentials are very similar to those determined in our previous work;¹⁴ they are, however, more accurate than the earlier¹⁴ ones because of full minimization of the MP2/6-31G(d,p) energy function instead of energy minimization in the RHF/6-31G(d,p) approximation and adding the second-order correction to energy and because of use of the pyrrolidine blocking group to account for a succeeding proline residue instead of using the N',N'-dimethylamine group.

For sequences with alternating chirality, the torsional and double-torsional potentials exhibit shifted low-energy regions with respect to those for all-L- or all-D-amino acid residue sequences. The torsional potentials for the Ala_L-Ala_L sequence exhibit the global energy minimum at the extended conformation,

which corresponds to the sequence of two C_{eq}^7 or extended conformations of consecutive residues. When the chirality of one of the residues is changed, the preferred absolute value of the γ angle is small; this corresponds to a sequence of C_{eq}^7 and C_{ax}^7 conformations or vice versa, depending on the chirality of the first residue or of two extended conformations adjacent to the respective C^7 region. This feature is reflected in the experimental conformation of gramicidin D, which forms a dimer which has a conformation of a left-handed double helix; this double helix can be considered as a distorted two-stranded intermolecular β sheet and has the hydrogen-bond pattern characteristic of a β sheet (Figure 8).

We have also demonstrated that the correlation and the side-chain-rotamer potentials of D-amino acid residues can be obtained from those of L-amino acid residues by reflection of the respective coefficients in the xy plane of a local coordinate system associated with the $C^\alpha \cdots C^\alpha \cdots C^\alpha$ unit of a given residue. To complete the extension of UNRES to systems containing D-amino acid residues, the virtual-bond-angle potentials must be determined; these potentials also depend on the virtual-bond-angles γ_1 and γ_2 adjacent to a given virtual-bond-angle θ . This work is currently underway in our laboratory. The force field will need to be calibrated with the structures and thermodynamic data of peptides and proteins containing D-amino acid residue. Gramicidin D, which already gives reasonable results with the state-of-the-art UNRES force field, appears to be a good training protein because its structure depends on the change of the local features of the force field following the introduction of D-amino acid residues, while appropriate thermodynamic data appear to be those of helix destabilization upon introduction of a D-amino acid residue into a sequence.⁴⁵

The modified force field has already been used to study the folding and association of the synthetic homotetrameric $\beta\beta\alpha$ (BBAT1) protein, which contains a D-proline residue.⁴⁸ The treatment of D-amino acid residues along with the respective torsional potentials has been incorporated into the UNRES open-source package available at www.unres.pl, and the upgraded version of UNRES with the new feature will be available soon at the above Web page.

■ ASSOCIATED CONTENT

■ Supporting Information

Table S1: Parameters of Fourier-series fit (eq 11) to the torsional potentials of mean force for all non-reducible potentials. Table S2: Coefficients of the two-dimensional Fourier expansion (eq 12) of the double-torsional potentials of mean force for all non-reducible potentials of the terminally-blocked tripeptide systems. Figure S1: Contour plots of the irreducible double-torsional potentials of mean force of the tripeptide systems considered in this work (126 systems). This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: +48585235430. Fax: +48585235472. E-mail: adasko@sun1.chem.univ.gda.pl.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by grant UMO-2011/01/N/ST4/01772 from the National Science Center of Poland, the National Institutes of Health (GM-14312 and GM-62838), and the National Science Foundation (MCB10-19767). This research was supported by an allocation of advanced computing resources provided by the National Science Foundation (<http://www.nics.tennessee.edu/>) and by the National Science Foundation through TeraGrid resources provided by the Pittsburgh Supercomputing Center. Computational resources were also provided by (a) the supercomputer resources at the Informatics Center of the Metropolitan Academic Network (IC MAN) in Gdańsk, (b) the 624-processor Beowulf cluster at the Baker Laboratory of Chemistry, Cornell University, (c) our 184-processor Beowulf cluster at the Faculty of Chemistry, University of Gdańsk, and (d) the SOONER cluster at the University of Oklahoma.

■ REFERENCES

- (1) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucl. Acid Res.* **2000**, *28*, 235–242.
- (2) Kreil, G. *Annu. Rev. Biochem.* **1997**, *66*, 337–345.
- (3) Craig, A. G.; Bandyopadhyay, P.; Olivera, B. M. *Eur. J. Biochem.* **1999**, *264*, 271–275.
- (4) Caparros, M.; Pisabarro, A. G.; de Pedro, M. A. *J. Bacteriol.* **1992**, *174*, 5549–5559.
- (5) Wenger, R. M. *Helv. Chim. Act.* **1984**, *67*, 502–525.
- (6) Wade, D.; Boman, A.; Wahlin, B.; Drain, C. M.; Andreu, D.; Boman, H. G.; Merrifield, R. B. *Proc. Natl. Acad. Sci. U. S. A.* **1990**, *87*, 4761–4765.
- (7) Global Industry Analysts, I. *PRWEB*.
- (8) Liwo, A.; Khalili, M.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 2362–2367.
- (9) Khalili, M.; Liwo, A.; Jagielska, A.; Scheraga, H. A. *J. Phys. Chem. B* **2005**, *109*, 13798–13810.
- (10) Liwo, A.; Pincus, M. R.; Wawak, R. J.; Rackovsky, S.; Scheraga, H. A. *Protein Sci.* **1993**, *2*, 1715–1731.
- (11) Liwo, A.; Oldziej, S.; Pincus, M. R.; Wawak, R. J.; Rackovsky, S.; Scheraga, H. A. *J. Comput. Chem.* **1997**, *18*, 849–873.
- (12) Liwo, A.; Pincus, M. R.; Wawak, R. J.; Rackovsky, S.; Oldziej, S.; Scheraga, H. A. *J. Comput. Chem.* **1997**, *18*, 874–887.
- (13) Liwo, A.; Czaplewski, C.; Pillardy, J.; Scheraga, H. A. *J. Chem. Phys.* **2001**, *115*, 2323–2347.
- (14) Oldziej, S.; Kozłowska, U.; Liwo, A.; Scheraga, H. A. *J. Phys. Chem. A* **2003**, *107*, 8035–8046.
- (15) Liwo, A.; Oldziej, S.; Czaplewski, C.; Kozłowska, U.; Scheraga, H. A. *J. Phys. Chem. B* **2004**, *108*, 9421–9438.
- (16) Liwo, A.; Khalili, M.; Czaplewski, C.; Kalinowski, S.; Oldziej, S.; Wachucik, K.; Scheraga, H. A. *J. Phys. Chem. B* **2007**, *111*, 260–285.
- (17) Kozłowska, U.; Liwo, A.; Scheraga, H. A. *J. Phys.: Condens. Matter* **2007**, *19*, 285203.
- (18) Liwo, A.; Czaplewski, C.; Oldziej, S.; Rojas, A. V.; Kazmierkiewicz, R.; Makowski, M.; Murarka, R. K.; Scheraga, H. A. In *Coarse-Graining of Condensed Phase and Biomolecular Systems*; Voth, G., Ed.; CRC Press: Boca Raton, FL, 2008; Chapter 8, pp 107–122.
- (19) Kozłowska, U.; Maisuradze, G. G.; Liwo, A.; Scheraga, H. A. *J. Comput. Chem.* **2010**, *31*, 1154–1167.
- (20) Makowski, M.; Liwo, A.; Sobolewski, E.; Scheraga, H. A. *J. Phys. Chem. B* **2011**, *115*, 6119–6129.
- (21) Makowski, M.; Liwo, A.; Scheraga, H. A. *J. Phys. Chem. B* **2011**, *115*, 6130–6137.
- (22) Sieradzian, A. K.; Scheraga, H. A.; Liwo, A. *J. Chem. Theory Comput.* **2012**, *8*, 1334–1343.
- (23) Liwo, A.; Kaźmierkiewicz, R.; Czaplewski, C.; Groth, M.; Oldziej, S.; Wawak, R. J.; Rackovsky, S.; Pincus, M. R.; Scheraga, H. A. *J. Comput. Chem.* **1998**, *19*, 259–276.

- (24) Kozłowska, U.; Liwo, A.; Scheraga, H. A. *J. Comput. Chem.* **2010**, *31*, 1143–1153.
- (25) Liwo, A.; Lee, J.; Ripoll, D. R.; Pillardy, J.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 5482–5485.
- (26) Lee, J.; Liwo, A.; Ripoll, D. R.; Pillardy, J.; Scheraga, H. A. *Proteins Struct. Funct. Genet.* **1999**, *Suppl. 3*, 204–208.
- (27) Pillardy, J.; Czaplewski, C.; Liwo, A.; Lee, J.; Ripoll, D. R.; Kaźmierkiewicz, R.; Oldziej, S.; Wedemeyer, W. J.; Gibson, K. D.; Arnautova, Y. A.; Saunders, J.; Ye, Y.-J.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 2329–2333.
- (28) Oldziej, S.; Czaplewski, C.; Liwo, A.; Chinchio, M.; Nancias, M.; Vila, J. A.; Khalili, M.; Arnautova, Y. A.; Jagielska, A.; Makowski, M.; Schafroth, H. D.; Kaźmierkiewicz, R.; Ripoll, D. R.; Pillardy, J.; Saunders, J. A.; Kang, Y. K.; Gibson, K. D.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 7547–7552.
- (29) Khalili, M.; Liwo, A.; Rakowski, F.; Grochowski, P.; Scheraga, H. A. *J. Phys. Chem. B* **2005**, *109*, 13785–13797.
- (30) Rakowski, F.; Grochowski, P.; Lesyng, B.; Liwo, A.; Scheraga, H. A. *J. Chem. Phys.* **2006**, *125*, 204107.
- (31) Saunders, J. A.; Scheraga, H. A. *Biopolymers* **2003**, *68*, 300–317.
- (32) Rojas, A. V.; Liwo, A.; Scheraga, H. A. *J. Phys. Chem. B* **2007**, *111*, 293–309 PMID: 17201452.
- (33) Rojas, A.; Liwo, A.; Browne, D.; Scheraga, H. A. *J. Mol. Biol.* **2010**, *404*, 537–552.
- (34) Rojas, A.; Liwo, A.; Scheraga, H. A. *J. Phys. Chem. B* **2011**, *115*, 12978–12983.
- (35) He, Y.; Liwo, A.; Weinstein, H.; Scheraga, H. J. *Mol. Biol.* **2011**, *405*, 298–314.
- (36) Golas, E. I.; Maisuradze, G. G.; Senet, P.; Oldziej, S.; Czaplewski, C.; Scheraga, H. A.; Liwo, A. *J. Chem. Theory Comput.* **2012**, *8*, 1334–1343.
- (37) Lee, J.; Liwo, A.; Ripoll, D. R.; Pillardy, J.; Saunders, J. A.; Gibson, K. D.; Scheraga, H. A. *Int. J. Quantum Chem.* **2000**, *77*, 90–117.
- (38) Liwo, A.; He, Y.; Scheraga, H. A. *Phys. Chem. Chem. Phys.* **2011**, *13*, 16890–16901.
- (39) Shen, H.; Liwo, A.; Scheraga, H. A. *J. Phys. Chem. B* **2009**, *113*, 8738–8744.
- (40) Kolinski, A.; Skolnick, J. J. *J. Chem. Phys.* **1992**, *97*, 9412–9426.
- (41) He, Y.; Xiao, Y.; Liwo, A.; Scheraga, H. A. *J. Comput. Chem.* **2009**, *30*, 2127–2135.
- (42) Nishikawa, K.; Momany, F. A.; Scheraga, H. A. *Macromolecules* **1974**, *7*, 797–806.
- (43) Liwo, A.; Pincus, M. R.; Wawak, R. J.; Rackovsky, S.; Scheraga, H. A. *Protein Sci.* **1993**, *2*, 1697–1714.
- (44) Gay, D. M. *ACM Trans. Math. Software* **1983**, *9*, 503–524.
- (45) Krause, E.; Bienert, M.; Schmieder, P.; Wenschuh, H. J. *Am. Chem. Soc.* **2000**, *122*, 4865–4870.
- (46) Mitchell, J. B.; Smith, J. *Proteins: Struct., Funct., Bioinf.* **2003**, *50*, 563–571.
- (47) Shepherd, N. E.; Hoang, H. N.; Abbenante, G.; Fairlie, D. P. *J. Am. Chem. Soc.* **2009**, *131*, 15877–15886.
- (48) Sieradzian, A. K.; Liwo, A.; Hansmann, U. H. J. *J. Chem. Theory Comput.* **2012**, *8*, 3416–3422.