Extension of UNRES force field to treat polypeptide chains with D-amino-acid residues

Adam K. Sieradzan[1,2,*], **Ulrich H.E. Hansmann**[2], **Harold A. Scheraga**[3], and **Adam Liwo**[1]

[1]Faculty of Chemistry, University of Gdańsk, Sobieskiego 18, 80-952 Gdańsk, Poland
[2]Department of Chemistry and Biochemistry, Oklahoma University, Norman, OK, 73019, U.S.A
[3]Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca, N.Y., 14853-1301, U.S.A

Abstract

Coarse-grained force fields for protein simulations are usually designed and parameterized 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 $C^a \cdots C^a$ virtual-bond axis and two consecutive $C^a \cdots C^a$ 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 $C_{i-1}^{\alpha} \cdots C_i^{\alpha} \left(\lambda^{(1)} \right)$ and $C_i^{\alpha} \cdots C_{i+1}^{\alpha} \left(\lambda^{(2)} \right)$ 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 non-equivalent 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

[*]Corresponding author; phone: +48585235430; fax: +48585235472; adasko@sun1.chem.univ.gda.pl.

residues. For all-atom force fields, except those which contain sine torsional terms, there is no problem to handle the D-configuration; however, coarse-grained force fields require additional parameterization 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 last 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 parameterized[13–15,17,19–22,24] as a potential of mean force of polypeptide chains immersed in water. 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.[25–29] In CASP4, we predicted, for the first time, the complete structure of a 70-residue protein, bacteriocin S (target T0102); with the introduction of correlation terms, we also were able to treat proteins with β-structure.[27] During the CASP5 exercise in 2004, we predicted nearly the complete structure (203 out of 235-residues) of a six-a-helix bundle protein (target T0198), complete or nearly complete structures of three other smaller a- and a + b-proteins, and significant portions of structures of other proteins.[28] It has also been extended to run coarse-grained molecular dynamics[8,9,30,31] and to study oligomeric proteins.[32,33] Recently, trans-cis isomerization of peptide groups has been introduced into UNRES.[22] UNRES has also been applied to study biological problems such as amyloid formation,[34,35] signaling[36] and chaperone dynamics.[37]

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.

$$U = w_{SC} \sum_{i<j} U_{SC_i SC_j} + w_{SCp} \sum_{i \neq j} U_{SC_i p_j} + w_{pp}^{VDW} \sum_{i<j-1} U_{p_i p_j}^{VDW} + w_{pp}^{el} f_2(T) \sum_{i<j-1} U_{p_i p_j}^{el}$$
$$+ w_{tor} f_2(T) \sum_i U_{tor}(\gamma_i) + w_{tord} f_3(T) \sum_i U_{tord}(\gamma_i, \gamma_{i+1})$$
$$+ w_b \sum_i U_b(\theta_i) + w_{rot} \sum_i U_{rot}(\alpha_{SC_i}, \beta_{SC_i}) + w_{bond} \sum_i U_{bond}(d_i) \qquad (1)$$
$$+ w_{corr}^{(3)} f_3(T) U_{corr}^{(3)} + w_{corr}^{(4)} f_4(T) U_{corr}^{(4)} + w_{corr}^{(5)} f_5(T) U_{corr}^{(5)} + w_{corr}^{(6)} f_6(T) U_{corr}^{(6)}$$
$$+ w_{turn}^{(3)} f_3(T) U_{turn}^{(3)} + w_{turn}^{(4)} f_4(T) U_{turn}^{(4)} + w_{turn}^{(6)} f_6(T) U_{turn}^{(6)}$$

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 ith virtual bond, which is either a $C^a \cdots C^a$ virtual bond or $C^a \cdots$ 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 work[16] and which reflect the dependence of the first generalized-cumulant term in those factors on temperature, as discussed in refs 16 and 38. The factors f_n are defined by eq 2.

$$f_n(T) = \frac{\ln\left[\exp(1) + \exp(-1)\right]}{\ln\left\{\exp\left[(T/T_\circ)^{n-1}\right] + \exp\left[-(T/T_\circ)^{n-1}\right]\right\}} \qquad (2)$$

where $T_\circ = 300$ K.

The term $U_{SC_i SC_j}$ represents the mean free energy of the hydrophobic (hydrophilic) interactions between the side chains, which implicitly contains the contributions from the interactions of the side chain with the solvent. The term $U_{SC_i p_j}$ 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 ($U_{p_i p_j}^{VDW}$) and the average electrostatic energy between peptide-group dipoles ($U_{p_i p_j}^{el}$); 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_{tord}, 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, 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.[13,23,39]

The energy-term weights are determined by force-field calibration to reproduce the structure and folding thermodynamics of selected training proteins.[16,40] As a training proteins LysM domain from E. coli (and R + protein; PDB code: 1E0G),[41] the Fbp28Ww domain from Mus musculus (a protein; PDB code: 1E0L),[42] the albumin-binding GA module (an R protein; PDB code: 1GAB),[43] and the IgG-bindingdomain from streptococcal protein G (an R + protein; PDB code: 1IGD)[44] and namely the tryptophan cage (PDB code: 1L2Y)[45] and the tryptophan zipper 2 (PDB code: 1LE1)[46] were used.

The present UNRES is parameterized 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_{tord}), 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 ith 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^a 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 potentials[12] 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, $U_{corr}^{(m)}$. 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 non-glycine and non-proline residues) in terms of the angles $\lambda^{(1)}$ and $\lambda^{(2)}$ for rotation of the peptide groups about the $C^a \cdots C^a$ virtual bonds of terminally-blocked amino-acid residues,[13] as given by eq 3.

$$e_X (\lambda^{(1)}, \lambda^{(2)}) = \mathbf{b_{1X}^T u} + \mathbf{b_{2X}^T v} + \mathbf{u^T C_X u} + \mathbf{v^T D_X v} + \mathbf{u^T E_X v} + \ldots \quad (3)$$

with

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

where $\mathbf{b_{1X}}$ and $\mathbf{b_{2X}}$ are the 2-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 = 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.

$$e_x\left(\lambda^{(1)},\lambda^{(2)}\right)=e_X\left(-\lambda^{(1)},-\lambda^{(2)}\right)=\mathbf{b_{1x}^T u}+\mathbf{b_{2x}^T v}+\mathbf{u^T C_x u}+\mathbf{v^T D_x v}+\mathbf{u^T E_x v}+\dots \quad (5)$$

with

$$\mathbf{b_{1x}=Ub_{1x}}=\begin{pmatrix} b_{11X} \\ -b_{21X} \end{pmatrix}, \quad \mathbf{b_{2x}=Ub_{2x}}=\begin{pmatrix} b_{12X} \\ -b_{22X} \end{pmatrix}$$

$$\mathbf{C_x=UC_xU^T}=\begin{pmatrix} c_{X11} & -c_{X12} \\ -c_{X12} & c_{X22} \end{pmatrix},$$

$$\mathbf{D_x=UD_xU^T}=\begin{pmatrix} d_{X11} & -d_{X12} \\ -d_{X12} & d_{X22} \end{pmatrix}, \quad (6)$$

$$\mathbf{E_x=UE_xU^T}=\begin{pmatrix} e_{X11} & -e_{X12} \\ -e_{X21} & e_{X22} \end{pmatrix}$$

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 work[14] 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.

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

where e_X and e_Y are the energy surfaces for the terminally-blocked residues of type X and Y respectively, 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), β = $1/RT$ where R is the universal gas constant and T is the absolute temperature. For Y = Pro, the C-terminal blocking group of the X residue is pyrrolidine; otherwise it is the methyl group.

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

(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 $C^a \cdots C^a$ 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 $C^a \cdots C^a$ bonds, defined by Nishikawa et al.[47] and also shown in Figure 3. The local angles of consecutive residues are, in turn, related to the chain-wide angles $\lambda_1, \lambda_2, \ldots, \lambda_n$ shown in Figure 2 by eq 10.

$$\lambda_i^{(1)} = \lambda_i, \quad i = 1, 2, \ldots, n$$
$$\lambda_i^{(2)} = \lambda_{i+1}, \quad i = 1, 2, \ldots, n-1 \quad (10)$$
$$\lambda_n^{(2)} = \lambda_n$$

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 $\lambda^{(2).48}$ 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 non-proline 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 eq 8 and eq 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)

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

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

(12)

where a_i, b_i, …, 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).

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

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 (eqs 11 or 12), $\Delta U(\gamma)$ is the difference between the U expressed by eqs 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 good fit to numerically calculated PMF values. The target function of eq 13 was minimized with respect to the coefficients a_0, a_1 …, b_6 (for single-torsional potentials; eq 11) or c_0, c_1, …, $h_{8,8}$ (for double-torsional potentials; eq 12) by using the Secant Unconstrained Minimization SoLver (SUMSL) procedure.[49] 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.

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

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

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-acetylL-alanyl-N′-pyrrolidylamide, trans-N-acetyl-glycyl-N′-methyl amide, trans-N-acetyl-glycyl-N′-pyrrolidylamide, trans-N-acetyl-L-prolyl-N′-methyl amide, trans-N-acetyl-L-prolyl-N′-pyrrolidylamide are shown in the Fig. 4 and 5, respectively. The potential-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 4C) resemble those obtained in our earlier study (Figure 4a and 4c 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-NMe$_2$ and Ac-Ala$_L$-NMe$_2$ potential-energy surfaces used in our previous study[14] 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 4D). The most significant difference is that the high-energy regions $\lambda^{(1)} = 0°$ and $\lambda^{(2)} = 0°$, and $\lambda^{(1)} = 0°$ and $\lambda^{(2)} = 180°$, are noticeably smaller for both Ac-Gly-Pyr and Ac-Ala$_L$-Pyr compared to those for Ac-Gly-NMe$_2$ and Ac-Ala$_L$-NMe$_2$, 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 more complex shape than those of Ac-Gly-NMe$_2$ and Ac-Ala$_L$-NMe$_2$, which is understandable in view of the more complex shape of the Pyr blocking group compared to that of the NMe$_2$ group.

Another difference between the potential-energy maps of NMe$_2$- and Pyr-blocked residues is the shape and topology of the energy minima. In Ac-Ala$_L$-NMe$_2$, two clearly-separated minima at ($\lambda^{(1)} = -60°$, $\lambda^{(2)} = 90°$) and and ($\lambda^{(1)} = -150°$, $\lambda^{(2)} = 150°$), 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 ($\lambda^{(1)} = -75°$, $\lambda^{(2)} = 90°$ and $\lambda^{(1)} = -135°$, $\lambda^{(2)} = 135°$) and

nearly merged to create a broad low-energy region comprising the C^7_{eq} 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 use of a different C-terminal blocking group and minimization of a different energy function, can be observed between Ac-Pro-NMe$_2$ (Figure 6 in ref 14) and Ac-Pro$_L$-Pyr (Figure 5). The Ac-Pro$_L$-Pyr potential-energy plot exhibits only one minimum, while that of Ac-Pro$_L$-NMe$_2$ has two maxima. The maximum that appeared at $\lambda^{(2)} = 30°$ for Ac-Pro$_L$-NMe$_2$ has coalesced with the minimum at $\lambda^{(2)} \approx 60°$ into an inflexion point at $\lambda^{(2)} \approx 20°$. 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 groups (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 Supplementary Material.

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 $\gamma = 180°$ and $\gamma = \pm60°$, respectively. These minima correspond to extended, type I′ or type II′ (for $\gamma = -60°$) and type I or type II (for $\gamma = 60°$) 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^7_{eq} conformation with a C=O$\cdots$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 4C) and, consequently, has high entropy. No minimum corresponding to an α-helical structure is present ($\gamma \approx 60°$), 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-aminoacid-residue sequences about $\gamma = 0°$ (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^7_{eq} conformation (for the L-configuration) or C^7_{ax} conformation (for the D-configuration) or in the portion of the extended-structure region adjacent to the

C^7_{eq} (C^7_{ax}) 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^7_{eq} and

C_{ax}^{7} conformations corresponds to low energy of intra-residue interaction. It should be noted that four minima are also present for the Ala_L-Ala_D-NHMe systems; the global minimum occurs at $\gamma \approx -60°$ and the second one occurs at $\gamma \approx 30°$. Of course, for the Ac-Ala_D-Ala_L-NHMe sequence, these minima occur at $\gamma \approx 60°$ and $\gamma \approx -30°$, 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 C^7 (axial or equatorial) but also extended conformations, which are capable of forming inter-strand 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 8), which is a heterodimer of two chains, each with alternating D- and L-residue sequences, confirms this observation. Our observations of torsional poteitials behaviour were confirmed by conducting MREMD (multiple replica exchange molecular dynamics) of gramicin D (Fig 8C). The structure with the lowest potential energy obtained in 250K is in left-handed double-helix. The RMSD is 3.70Å and is a result of shift by one residue of turn and flexibile N-terminal fragments.

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 $\gamma \approx \pm 60°$ coalesce into a broad minimum region centered at $\gamma = 0°$ and the extended-chain minimum at $\gamma = 180°$ becomes very shallow. Similar differences occur for all glycine-containing systems. For the other systems, the torsional potentials for Pyr-blocked dipeptides have 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 $\gamma = 0°$ to $\gamma = 90°$, 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.[50,51] Thus, longer-range than intraresidue 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 4 $U_{Ala_L-Ala(D,L)-Ala(D,L)-NHMe}(\gamma_1, \gamma_2)$ and the 4 $U_{Ala_L-Ala(D,L)-Ala(D,L)-Pyr}(\gamma_1, \gamma_2)$ 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 Supplementary Material. Two-dimensional Fourier series (eq 12) were subsequently fitted to the PMFs; the coefficients are collected in Table S2 of Supplementary Material.

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 ($\lambda^{(1)} = 15°$, $\lambda^{(2)} = -15°$) and the minimum region extends along the $\gamma_2 = \gamma_1$ axis towards positive values of γ_1. Thus, the double-torsional potentials for all-L-residue polypeptide chains diminish the propensity towards 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 towards $\gamma_1 = \gamma_2 = 180°$ and also a shallower minimum appears at $\gamma_1 = 120°$ and $\gamma_2 = 120°$. Thus, for sequences with alternating chirality, the double-torsional potentials have the opposite effect as that of the 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 $\gamma_1 = 15°$, $\gamma_2 = -165°$ and for the Ac-Ala$_L$-Ala$_D$-Ala$_D$-NHMe system (Figure 9G) the global minimum occurs at $\gamma_1 = -15°$, $\gamma_2 = 180°$, as in β-bulge structures.[52] 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 contribute to final structure of protein and cannot be separated in easy manner. Structure of gramicidin D obtained from simulation (Fig. 8) cannot be solely contributed to torsional interactions but is a sum of all interactions.

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. Based on 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 earlier[14] 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-Ala 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 $C^a \cdots C^a \cdots C^a$ 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 in currently underway in our laboratory. We also plan to run series of simulation of proteins eg. gramicidin D containing D-aminoacid residues to check their folding pathways and properties.

The UNRES software is an open source software and can be obtained from webpage: www.unres.pl

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. Nucl Acid Res. 2000; 28:235–242.

2. Kreil G. Annu Rev Biochem. 1997; 66:337–345. [PubMed: 9242910]

3. Craig AG, Bandyopadhayay P, Olivera BM. Eur J Biochem. 1999; 264:271–275. [PubMed: 10491070]

4. Caparros M, Pisabarro AG, de Pedro MA. J Bacteriol. 1992; 174:5549–5559. [PubMed: 1512190]

5. Wenger RM. Helv Chim Act. 1984; 67:502–525.

6. Wade D, Boman A, Wahlin B, Drain CM, Andreu D, Boman HG, Merrifield RB. Proc Natl Acad Sci USA. 1990; 87:4761–4765. [PubMed: 1693777]

7. Global Industry Analysts, I. *PRWEB*

8. Liwo A, Khalili M, Scheraga HA. Proc Natl Acad Sci USA. 2005; 102:2362–2367. [PubMed: 15677316]

9. Khalili M, Liwo A, Jagielska A, Scheraga H. J Phys Chem B. 2005; 109:13798–13810. [PubMed: 16852728]

10. Liwo A, Pincus MR, Wawak RJ, Rackovsky S, Scheraga HA. Protein Sci. 1993; 2:1715–1731. [PubMed: 8251944]

11. Liwo A, Ołdziej S, Pincus MR, Wawak RJ, Rackovsky S, Scheraga HA. J Comput Chem. 1997; 18:849–873.

12. Liwo A, Pincus MR, Wawak RJ, Rackovsky S, Ołdziej S, Scheraga HA. J Comput Chem. 1997; 18:874–887.

13. Liwo A, Czaplewski C, Pillardy J, Scheraga HA. J Chem Phys. 2001; 115:2323–2347.

14. Ołdziej S, Kozłowska U, Liwo A, Scheraga HA. J Phys Chem A. 2003; 107:8035–8046.

15. Liwo A, Ołdziej S, Czaplewski C, Kozłowska U, Scheraga HA. J Phys Chem B. 2004; 108:9421–9438.

16. Liwo A, Khalili M, Czaplewski C, Kalinowski S, Ołdziej S, Wachucik K, Scheraga H. J Phys Chem B. 2007; 111:260–285. [PubMed: 17201450]

17. Kozłowska U, Liwo A, Scheraga HA. J Phys: Cond Matter. 2007; 19:285203.

18. Liwo, A.; Czaplewski, C.; Ołdziej, S.; Rojas, AV.; KaŸmierkiewicz, R.; Makowski, M.; Murarka, RK.; Scheraga, HA. Coarse-Graining of Condensed Phase and Biomolecular Systems. Voth, G., editor. Vol. Chapter 8. CRC Press; 2008. p. 107-122.

19. Kozłowska U, Maisuradze GG, Liwo A, Scheraga HA. J Comput Chem. 2010; 31:1154–1167. [PubMed: 20017135]

20. Makowski M, Liwo A, Sobolewski E, Scheraga HA. J Phys Chem B. 2011; 115:6119–6129. [PubMed: 21500792]

21. Makowski M, Liwo A, Scheraga HA. J Phys Chem B. 2011; 115:6130–6137. [PubMed: 21500791]

22. Sieradzan AK, Scheraga HA, Liwo A. J Chem Theor Comput. 2012; 8:1334–1343.

23. Liwo A, KaŸmierkiewicz R, Czaplewski C, Groth M, Ołdziej S, Wawak RJ, Rackovsky S, Pincus MR, Scheraga HA. J Comput Chem. 1998; 19:259–276.

24. Kozłowska U, Liwo A, Scheraga HA. J Comput Chem. 2010; 31:1143–1153. [PubMed: 20073062]

25. Liwo A, Lee J, Ripoll DR, Pillardy J, Scheraga HA. Proc Natl Acad Sci, U S A. 1999; 96:5482–5485. [PubMed: 10318909]

26. Lee J, Liwo A, Ripoll DR, Pillardy J, Saunders JA, Gibson KD, Scheraga HA. Int J Quant Chem. 2000; 77:90–117.

27. Pillardy J, Czaplewski C, Liwo A, Lee J, Ripoll DR, KaŸmierkiewicz R, Oldziej S, Wedemeyer WJ, Gibson KD, Arnautova YA, Saunders J, Ye YJ, Scheraga HA. Proc Natl Acad Sci USA. 2001; 98:2329–2333. [PubMed: 11226239]

28. Ołdziej S, et al. Proc Natl Acad Sci USA. 2005; 102:7547–7552. [PubMed: 15894609]

29. Liwo A, He Y, Scheraga HA. Phys Chem Chem Phys. 2011; 13:16890–16901. [PubMed: 21643583]

30. Khalili M, Liwo A, Rakowski F, Grochowski P, Scheraga H. J Phys Chem B. 2005; 109:13785–13797. [PubMed: 16852727]

31. Rakowski F, Grochowski P, Lesyng B, Liwo A, Scheraga HA. J Chem Phys. 2006; 125:204107. [PubMed: 17144690]

32. Saunders JA, Scheraga HA. Biopolymers. 2003; 68:300–317. [PubMed: 12601791]

33. Rojas AV, Liwo A, Scheraga HA. J Phys Chem B. 2007; 111:293–309. [PubMed: 17201452]

34. Rojas A, Liwo A, Browne D, Scheraga HA. J Mol Biol. 2010; 404:537–552. [PubMed: 20888834]

35. Rojas A, Liwo A, Scheraga H. J Phys Chem B. 2011; 115:12978–12983. [PubMed: 21939202]

36. He Y, Weinstein AL, Scheraga HH. J Mol Biol. 2011; 405:298–314. [PubMed: 21050858]

37. Golas EI, Maisuradze GG, Senet P, Ołdziej S, Czaplewski C, Scheraga HA, Liwo A. J Chem Theory Comput. 2012; 8:1334–1343. [PubMed: 23087598]

38. Shen H, Liwo A, Scheraga HA. J Phys Chem B. 2009; 113:8738–8744. [PubMed: 19480420]

39. Kolinski A, Skolnick J. J Chem Phys. 1992; 97:9412–9426.

40. He Y, Xiao Y, Liwo A, Scheraga HA. J Comput Chem. 2009; 30:2127–2135. [PubMed: 19242966]

41. Bateman A, Bycroft M. J Mol Biol. 2000; 299:1113–1119. [PubMed: 10843862]

42. Macias MJ, Gervais V, Civera C, Oschkinat H. Nat Struct Biol. 2000; 7:375–379. [PubMed: 10802733]

43. Johansson MU, de Chateau M, Wikstrom M, Forsen S, Drakenberg T, Bjorck L. J Mol Biol. 1997; 266:859–865. [PubMed: 9086265]

44. Derrick JP, Wigley DB. J Mol Biol. 1994; 243:906–918. [PubMed: 7966308]

45. Neidigh JW, Fesinmeyer RM, Andersen NH. Nat Struct Biol. 2002; 9:425–430. [PubMed: 11979279]

46. Cochran AG, Skelton NJ, Starovasnik MA. Proc Natl Acad Sci USA. 2001; 98:5578–5583. [PubMed: 11331745]

47. Nishikawa K, Momany FA, Scheraga HA. Macromolecules. 1974; 7:797–806. [PubMed: 4437206]

48. Liwo A, Pincus MR, Wawak RJ, Rackovsky S, Scheraga HA. Protein Sci. 1993; 2:1697–1714. [PubMed: 7504550]

49. Gay DM. ACM Trans Math Software. 1983; 9:503–524.

50. Krause E, Bienert M, Schmieder P, Wenschuh H. J Am Chem Soc. 2000; 122:4865–4870.

51. Mitchell JB, Smith J. Proteins: Struct Func Bionif. 2003; 50:563–571.

52. Shepherd NE, Hoang HN, Abbenante G, Fairlie DP. J Am Chem Soc. 2009; 131:15877–15886. [PubMed: 19807085]
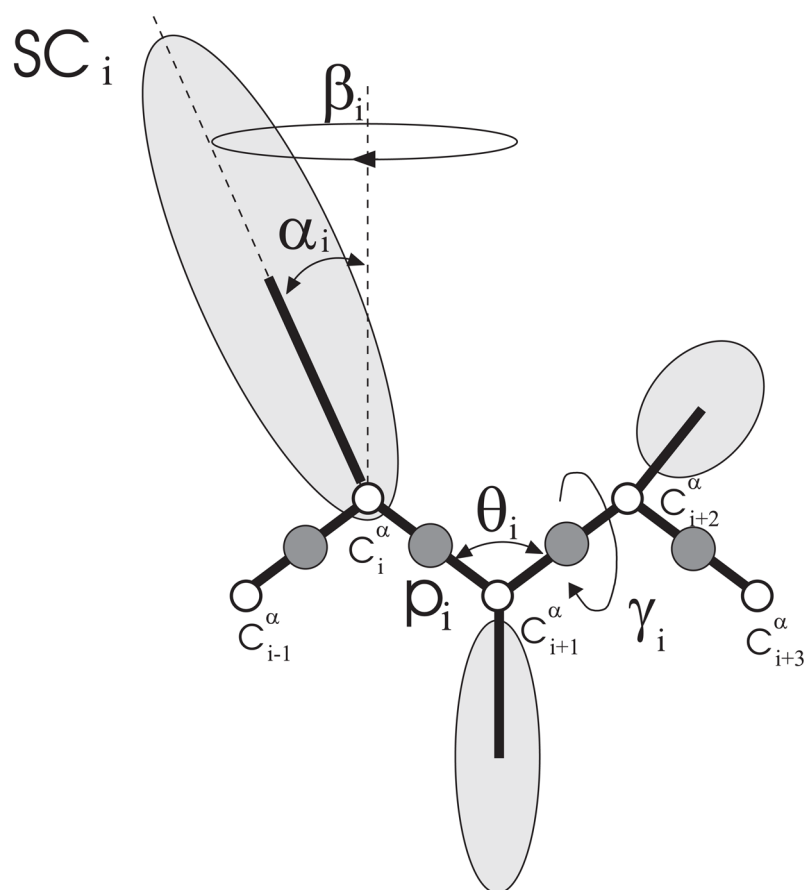
Fig. 1.
The UNRES model of polypeptide chains. The interaction sites are peptide-group centers
(p), and side-chain centers (SC) attached to the corresponding a-carbons with different $C^a \cdots$
SC bond lengths, d_{SC}. The peptide groups are represented as dark gray circles and the side
chains are represented as light gray ellipsoids of different size. The a-carbon atoms are
represented by small open circles. The geometry of the chain can be described either by the
virtual-bond vectors $\mathbf{dC}_i$ (from C_i^α to C_{i+1}^α), $i = 1, 2, \ldots, n - 1$, and $\mathbf{dX}_i$ (from C_i^α to SC_i), $i =$
$2, \ldots, n - 1$, represented by thick lines, where n is the number of residues, or in terms of
virtual-bond lengths, backbone virtual-bond angles θ_i, $i = 1, 2, \ldots, n - 2$, backbone virtual-
bond-dihedral angles γ_i, $i = 1, 2, \ldots, n - 3$, and the angles a_i and β_i, $i = 2, 3, \ldots, n - 1$ that
describe the location of a side chain with respect to the coordinate frame defined by
$C_{i-1}^\alpha, C_i^\alpha$, and $C_i^\alpha C_{i+1}^\alpha$.
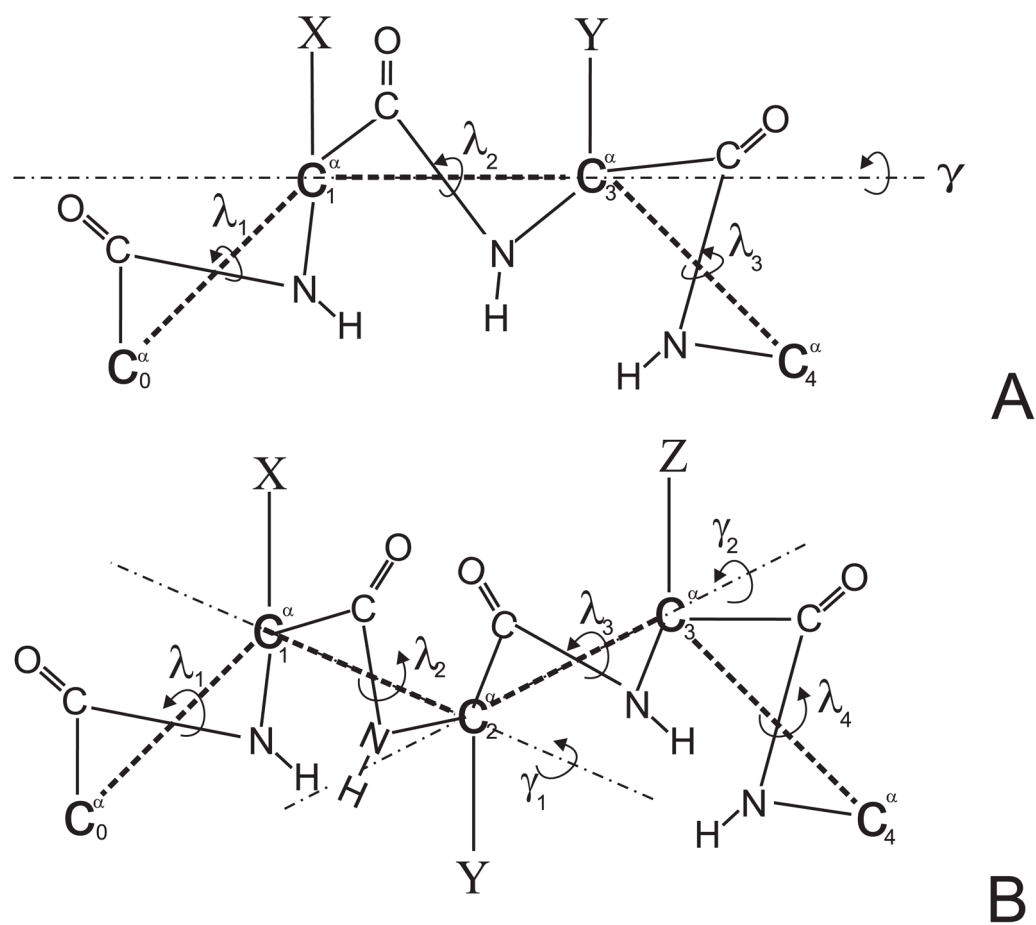
Fig. 2.
Illustration of the model terminally blocked dipeptides (a) and tripeptides (b) constructed to compute the integrals of eqs 8 and 9. Each X,Y and Z denotes sidechains of Ala_L, Ala_D, Gly, Pro_L or Pro_D.
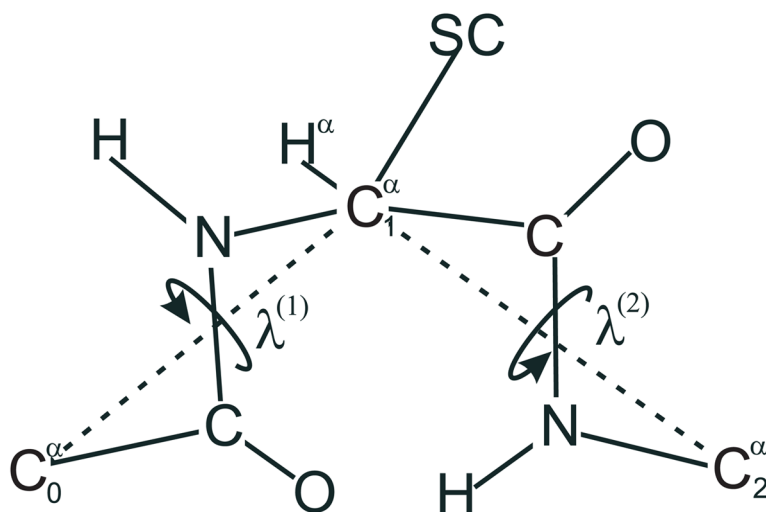
Fig. 3.
Definition of the dihedral angles $\lambda^{(1)}$ and $\lambda^{(2)}$ for rotation of a peptide unit about the $C^a \cdots C^a$ virtual bonds of a peptide unit.[47]
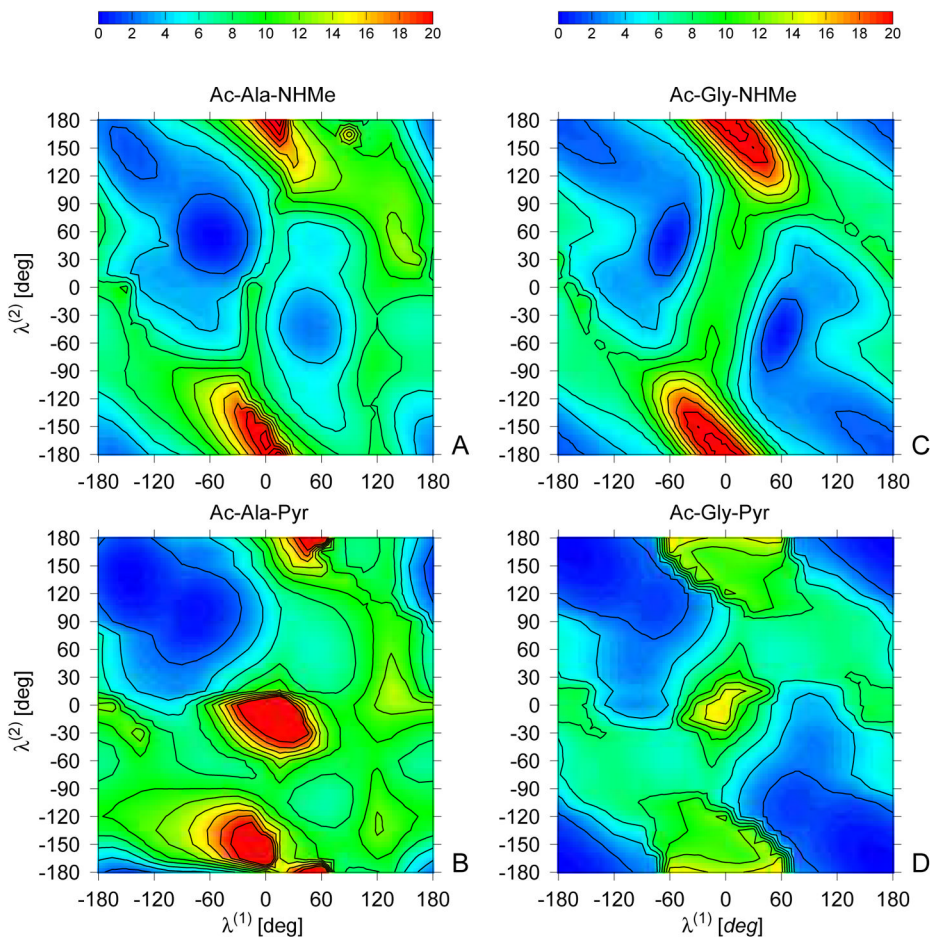
Fig. 4.
Contour plots of the diabatic potential-energy surfaces of Ac-Ala$_L$-NHMe (A), Ac-L-Ala$_L$-Pyr (B), Ac-Gly-NHMe (C), and Ac-Gly-Pyr (D) obtained with the MP2/6-31G(d,p) ab initio method. Energies (kcal/mol) are expressed as relative values with respect to the global-minimum value for a given system, which occurs at ($\lambda^{(1)} = -60°$, $\lambda^{(2)} = 60°$) for Ac-Ala$_L$-NHMe, ($\lambda^{(1)} = -60°$, $\lambda^{(2)} = 60°$) or ($\lambda^{(1)} = 60°$, $\lambda^{(2)} = -60°$) for Ac-Gly-NHMe, ($\lambda^{(1)} = -135°$, $\lambda^{(2)} = 135°$) for Ac-Ala$_L$-Pyr and ($\lambda^{(1)} = 180°$, $\lambda^{(2)} = 180°$) for Ac-Gly-Pyr (d). The energy color scale (kcal/mol) is on the top of the figure.
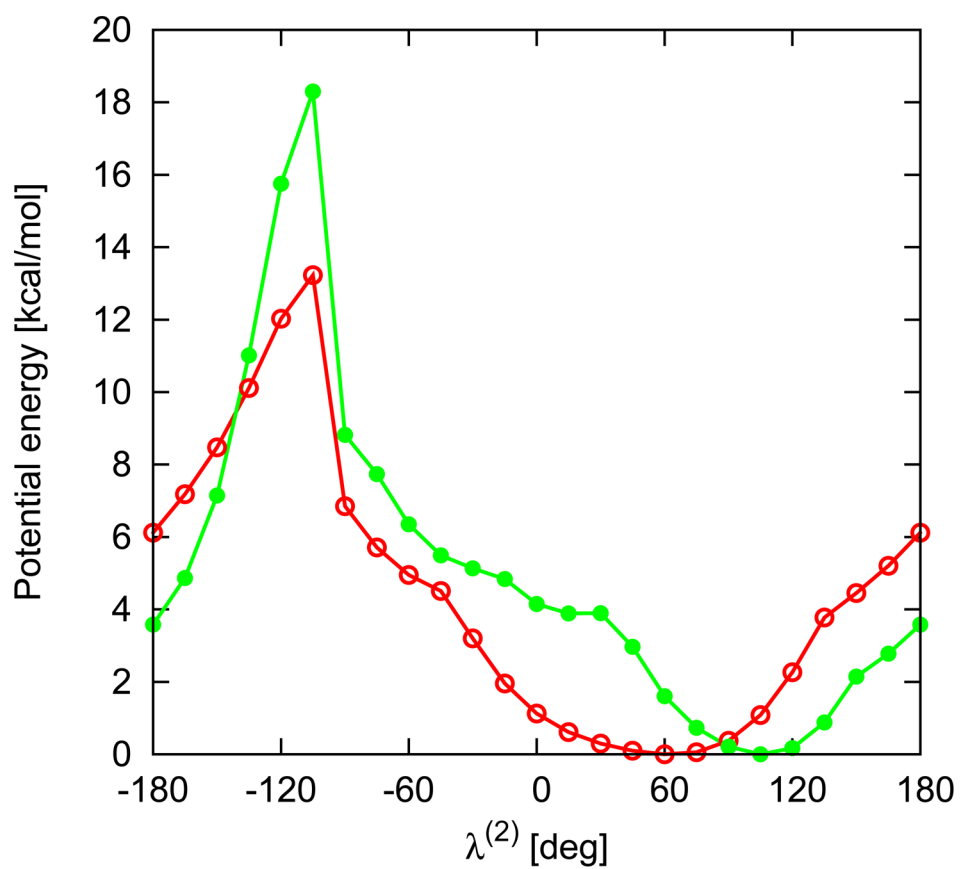
Fig. 5.
Relative energy profiles (kcal/mol) of Ac-Pro$_L$-NHMe (red line) and Ac-Pro$_L$-Pyr (green line) in $\lambda^{(2)}$ angle calculated with MP2/6-31G(d,p) ab initio method.
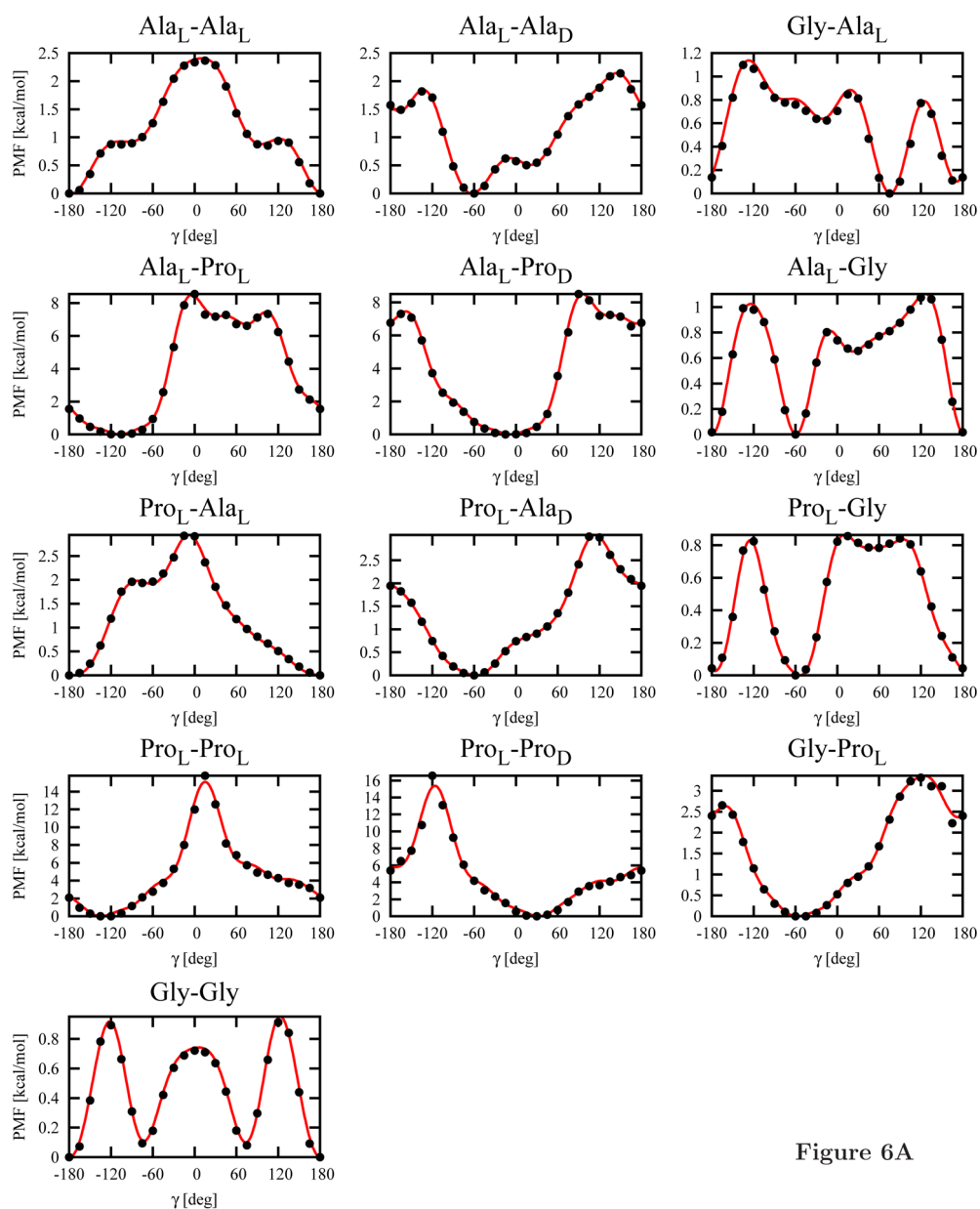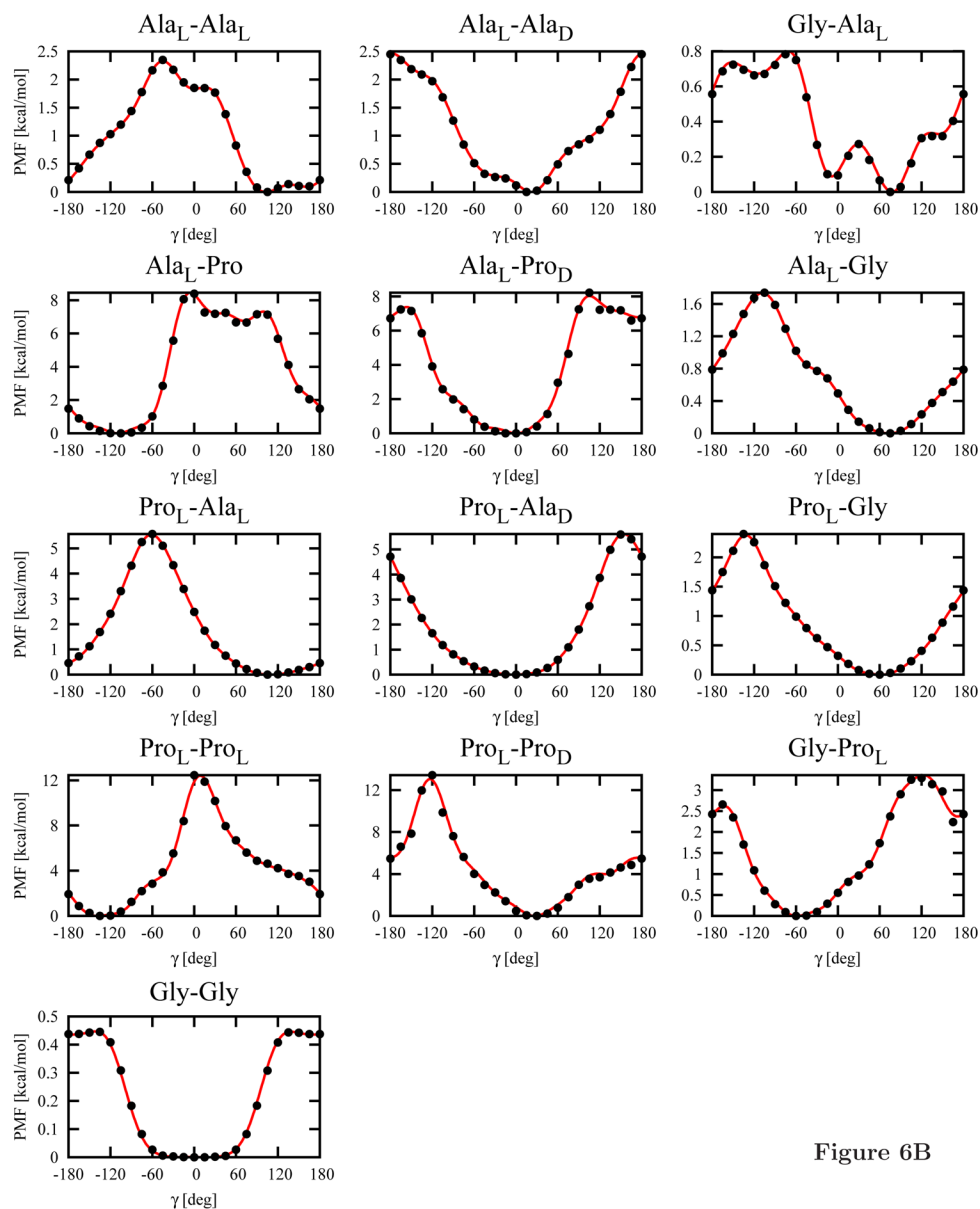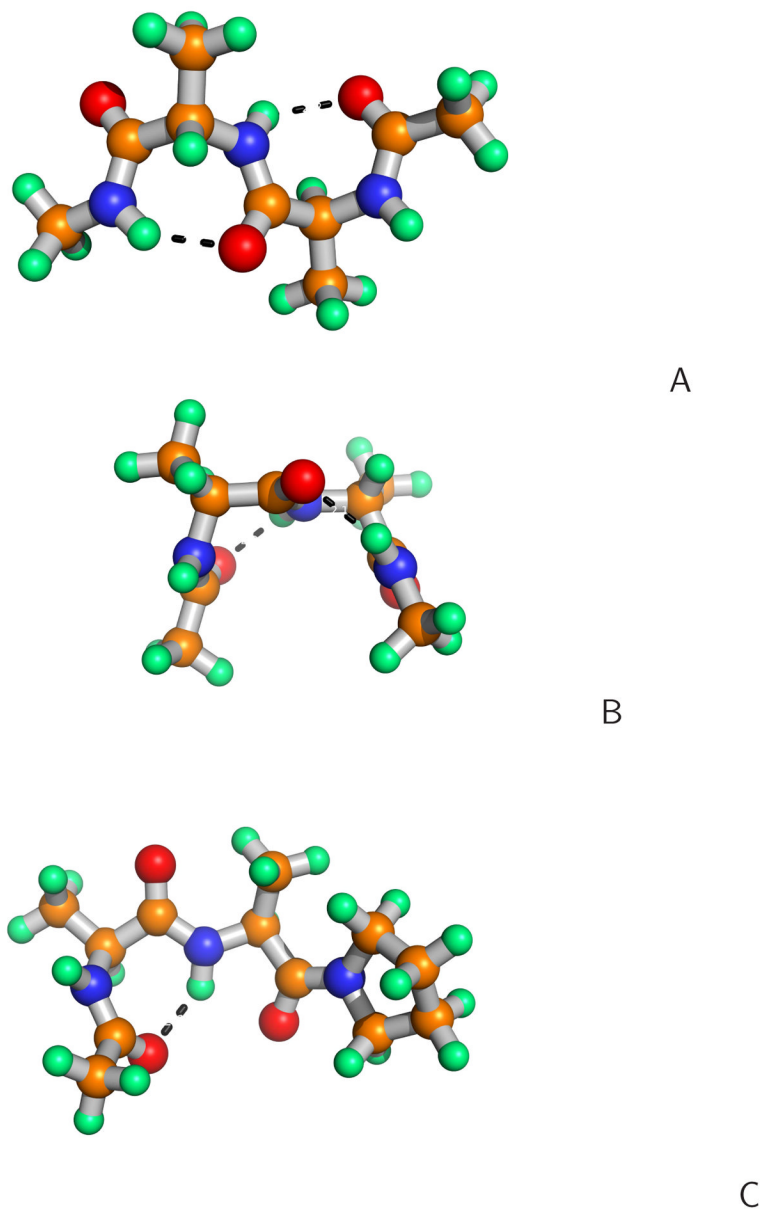
Figure 6A

Fig. 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.

Fig. 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°, $\lambda^{(2)}$ = 60°, $\lambda^{(3)}$ = −60°, $\lambda^{(4)}$ = 60°), (B) Ac-Ala$_L$-Ala$_D$-NHMe ($\lambda^{(1)}$ = −60°, $\lambda^{(2)}$ = 60°, $\lambda^{(3)}$ = 60°, $\lambda^{(4)}$ = −60°), (C) Ac-Ala$_L$-Ala$_L$-Pyr ($\lambda^{(1)}$ = −60°, $\lambda^{(2)}$ = 60°, $\lambda^{(3)}$ = −135°, $\lambda^{(4)}$ = 135°) systems.
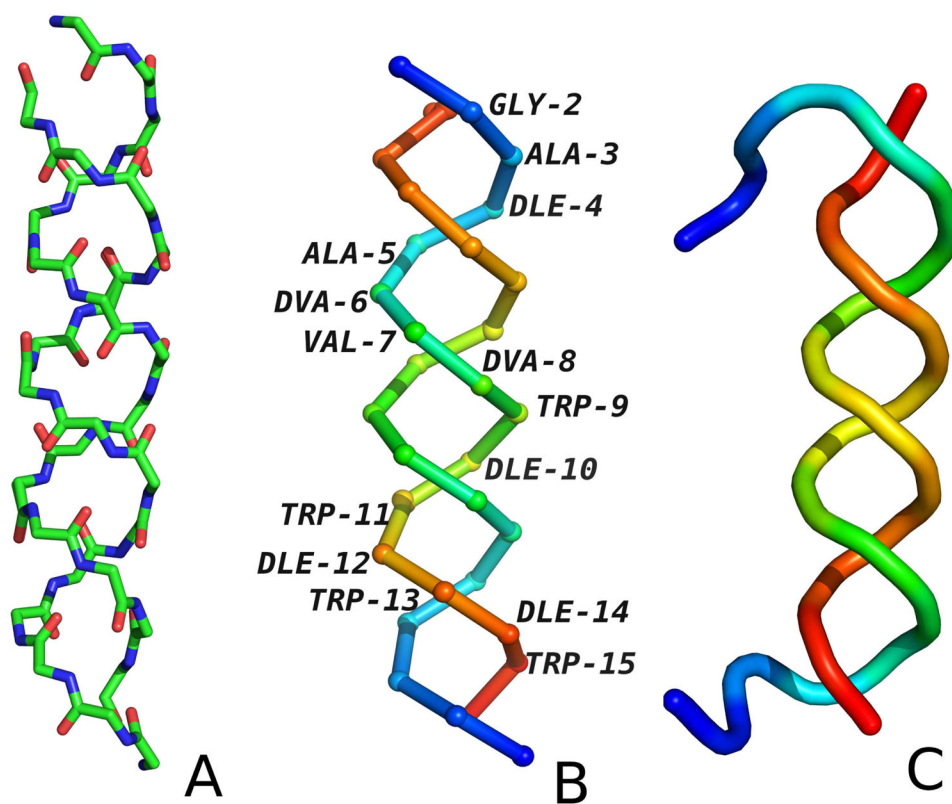
Fig. 8.
The experimental structure of gramicidin D (PDB code: 1AL4). (A) The backbone with all atoms shown to indicate the hydrogen-bonding contacts, (B) The C^a trace with residues colored from blue to red from the N-terminus to the C-terminus for each chain. (C) The cartoon representation of structure obtained from MREMD simulation at 250K colored from blue to red from the N-terminus to the C-terminus for each chain.
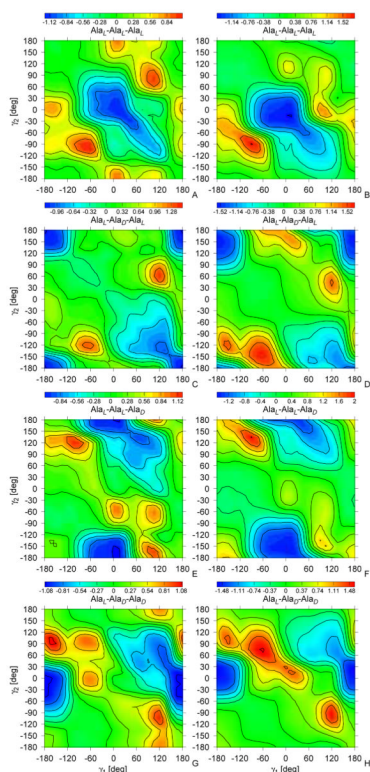
Fig. 9.
Sample contour plots of the double-torsional potentials for alanine-type tripeptides,
$U_{\text{L–Ala–(D,L)–Ala–(D,L)–Ala–NHMe}}(\gamma^{(1)}, \gamma^{(2)})$ (A, C, E, G) and
$U_{\text{L–Ala–(D,L)–Ala–(D,L)–Ala–Pyr}}(\gamma^{(1)}, \gamma^{(2)})$ (B, D, F, H). The energy scale (kcal/mol) is on the top of the figure.