Determination of the Potentials of Mean Force for Rotation about $C^{\alpha}-C^{\alpha}$ Virtual Bonds in Polypeptides from the ab Initio Energy Surfaces of Terminally Blocked Glycine, Alanine, and Proline†

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

Faculty of Chemistry, University of Gdańsk, Sobieskiego 18, 80-952 Gdańsk, Poland, Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca, N.Y., 14853-1301, U.S.A., and Academic Computer Center in Gdańsk TASK, Narutowicza 11/12, 80-952 Gdańsk, Poland

Received: October 30, 2002; In Final Form: January 13, 2003

The diabatic energy surfaces of terminally-blocked amino-acid residues (modeling the bulk of backbone-local interaction patterns in proteins: *trans*-N-acetyl-glycyl-*trans*-N′-methylamide, *trans*-N-acetyl-glycyl-N,N′-dimethylamide, *trans*-N-acetyl-L-alanyl-*trans*-N′-methylamide, *trans*-N-acetyl-L-alanyl-N,N′-dimethylamide, *trans*-N-acetyl-L-prolyl-*trans*-N′-methylamide, and *trans*-N-acetyl-L-prolyl-N,N′-dimethylamide) were calculated at the Møller−Plesset (MP2) ab initio level of theory with the 6-31G(d,p) basis set. The dihedral angles for rotation of the peptide groups about the $C^{\alpha}-C^{\alpha}$ virtual-bond axes ($\lambda^{(1)}$ and $\lambda^{(2)}$) were used as variables; for the proline-derived peptide, only $\lambda^{(2)}$ was variable, because of the presence of the pyrrolidine-ring constraint. The resulting energy maps were compared with those obtained with the ECEPP/3 force field. On the basis of the MP2/6-31G(d,p) energy surfaces of terminally blocked single residues, the torsional potentials of mean force for rotation about the $C^{\alpha}-C^{\alpha}$ virtual-bond axes and the double-torsional potentials of mean force for rotations about two consecutive virtual bond axes of all pairs and triplets of the prototypes of L-*trans*-amino acid residues were determined by numerical integration, fitted to one- and two-dimensional Fourier series in the virtual-bond-dihedral angles γ of the C^{α} trace of a polypeptide chain, and implemented in the united-residue force field.

1. Introduction

Local interactions within the polypeptide backbone are one of the determinants of protein architecture,[1−3] because they are essential for determining the formation of protein secondary structure, which in turn leads to well-defined regular architectures consisting of interacting segments of α-helical and β-sheet structure. A good theoretical description of these interactions is, therefore, necessary for theoretical simulation of protein folding and energy-based prediction of protein structure. Basic backbone-local interactions are encoded in the energy surfaces of terminally-blocked amino-acid residues sometimes referred to as Ramachandran maps,[4] although the original Ramachandran surfaces were constructed based only on steric considerations. This nomenclature will also be used for brevity in the current article.

The Ramachandran surfaces are usually expressed in terms of backbone dihedral angles ϕ and ψ, because the other backbone degrees of freedom, namely the dihedral angles ω as well as bond lengths and bond angles, are subject to only small variations from the equilibrium values. They have been characterized in detail at the empirical force field level.[2,5−8] However, studies at higher level of theory, viz., quantum-mechanical ab initio and density functional theory (DFT), were usually limited to optimizing the geometry and computing the energies of selected conformations;[9−23] detailed maps at the Hartree−Fock and post-Hartree−Fock level have been constructed only for the *formyl* peptidyl amides.[24−26] Only recently was a DFT study reported[27] for the Ramachandran surface of Ac−Ala−NHMe and, in an earlier paper,[20] the energy profiles were calculated for this molecule with the semiempirical AM1 and PM3[28] methods.

Protein conformations can be modeled both at the atomic-detailed (all-atom) or coarse-grained (united-residue) level. In both cases, empirical force fields, rather than first-principle energy calculations, must be implemented for tractability of the systems under consideration. All-atom force fields are parametrized to reproduce the energy relations between characteristic conformations (minima on the Ramachandran maps) of model peptides (usually terminally blocked glycine and terminally blocked alanine residues) obtained from state-of-the-art ab initio calculations.[6−30] While this treatment ignores the detailed structure of the whole energy surfaces, it seems to be sufficient for practical purposes; the Ramachandran maps calculated with commonly used empirical force fields qualitatively reproduce those obtained with ab initio calculations. A single artifact that should raise some concern is the presence of the A (right-handed-α-helical;α_r) conformation[2] as a stable minimum in the Ramachandran maps calculated with the ECEPP,[31] AMBER,[6,29,30] or CHARMM[5] empirical force fields, as opposed to the results of ab initio calculations of formyl peptidyl amides.[24,25] However, the terminal formyl blocking group used in these calculations[24,25] does not model the environment of a residue within a polypeptide chain exactly. A weak minimum close to the α_r conformation was obtained in the restricted Hartree−

* Corresponding author. Phone: (607) 255-4034; fax: (607) 254-4700; e-mail: has5@cornell.edu.

† Part of the special issue "A. C. Albrecht Memorial Issue".

‡ University of Gdańsk

§ Cornell University

|| Academic Computer Center in Gdańsk

10.1021/jp0223410 CCC: $25.00
Published on Web 03/13/2003

Fock (RHF) ab initio study of Iwaoka et al.[26] that included the polarizable-continuum model (PCM and I-PCM)[32,33] of solvation on formyl-blocked glycine and alanine residues.

The case of united-residue (UNRES) force fields is more complex. Most of them do not specify an explicit polypeptide backbone, but only the C^{α} trace geometry.[34−46] Therefore, the Ramachandran maps are hidden in such force fields in the "torsional" and "multi-torsional" potentials for rotational variation of the virtual-bond dihedral angles.[34,38,45,46] These potentials can be determined as knowledge-based potentials from the distribution functions of virtual-bond dihedral angles,[37,38,45] calculated from the Protein Data Bank (PDB),[47] or by averaging an appropriate all-atom energy function over the degrees of freedom that vanish when passing from the all-atom to a united-residue representation.[34,40,46]

In the last 10 years, we have started to develop the UNRES force field[40,45,46,48−51] for the physics-based prediction of protein structure *based solely on the amino-acid sequence* without ancillary information from structural databases. This force field performed reasonably well in two consecutive Community Wide Experiments on the Critical Assessment of Techniques for Protein Structure Prediction, CASP3[52,53] and CASP4.[54] The UNRES force field is a restricted free energy (RFE) function for a polypeptide chain corresponding to averaging of the atomic detailed energy function over the secondary degrees of freedom. Recently,[46] we presented a general theory for the construction of the components of coarse-grained energy functions by factorization of the total RFE of the system under consideration, with the factors corresponding to the smaller parts of the system. Within the framework of this formalism, the virtual-torsional potentials arise naturally by integrating over the Ramachandran maps of consecutive amino-acid residues, given a UNRES configuration; they correspond to second-order correlation terms between the local interactions within consecutive amino acid residues.[46] In that work, we calculated the virtual-bond torsional potentials preliminarily by using the Ramachandran maps calculated with the ECEPP/3 force field.[31] The aim of the present work was to use ab initio quantum mechanical methods to calculate high-quality Ramachandran maps of model peptides in order to revise the torsional potentials in the UNRES force field. Moreover, we also determined double-torsional potentials that, so far, have not been used in UNRES. For this purpose, energy maps were calculated for terminally blocked glycine, L-alanine, and L-proline residues, with L-alanine serving as a prototype of all amino acids except glycine and proline, as in our earlier work.[40,45,46] We also compared the computed Ramachandran maps with those obtained with the empirical force fields and with earlier quantum mechanical calculations.[24,25]

2. Methods

2.1. The UNRES Force Field. In the UNRES model,[48,50,51,55] a polypeptide chain is represented as a sequence of α-carbon (C^{α}) atoms linked by virtual bonds with attached united side chains (SC) and united peptide groups (p). Each united peptide group is located in the middle of two consecutive α-carbons. Only these united peptide groups and the united side chains serve as interaction sites; the α-carbons serve only to define the chain geometry (Figure 1). All virtual bond lengths (i.e. $C^{\alpha}-C^{\alpha}$ and $C^{\alpha}-SC$) are fixed; the distance between neighboring C^{α}'s is 3.8 Å, corresponding to *trans* peptide groups, while the

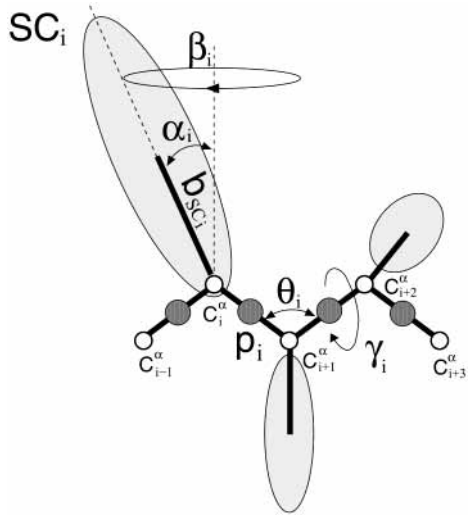


Figure 1. The UNRES model of polypeptide chains. The interaction sites are side-chain centroids of different sizes (SC) and the peptide-bond centers (p) are indicated by shaded circles, whereas the N-carbon atoms (small empty circles) are introduced only to assist in defining the geometry. The virtual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

side-chain angles (α_i and β_i), virtual-bond (θ), and dihedral (γ) angles can vary. The energy of the virtual-bond chain is expressed by eq 1.

$$U = \sum_{i<j} U_{SC_iSC_j} + w_{SC_p} \sum_{i\neq j} U_{SC_ip_j} + w_{el} \sum_{i<j-1} U_{p_ip_j} + w_{tor} \sum_i U_{tor}(\gamma_i) + w_b \sum_i U_b(\theta_i) + w_{rot} \sum_i U_{rot}(\alpha_i, \beta_i) + \sum_{m=1}^{N_{corr}} w^m_{corr} U^m_{corr} \quad (1)$$

The term $U_{SC_iSC_j}$ represents the mean free energy of the hydrophobic (hydrophilic) interactions between the side chains, which implicitly contains the contributions from the interactions of the side chain with the solvent. The term $U_{SC_ip_j}$ denotes the excluded-volume potential of the side-chain−peptide-group interactions. The peptide-group interaction potential ($U_{p_ip_j}$) accounts mainly for the electrostatic interactions (i.e., the tendency to form backbone hydrogen bonds) between peptide groups p_i and p_j. U_{tor}, U_b, and U_{rot} represent the energies of virtual-dihedral angle torsions, virtual-bond angle bending, and side-chain rotamers; these terms account for the local propensities of the polypeptide chain. Details of the parametrization of all of these terms are provided in earlier publications.[45,48] Finally, the terms U^m_{corr}, $m = 1,2,...$ N_{corr} are the *correlation* or *multibody* contributions from a cumulant expansion[46] of the RFE, and the w's are the weights of the energy terms. The multibody terms are indispensable for reproduction of regular α-helical and β-sheet structures. The UNRES force field has been derived as an RFE function of an all-atom polypeptide chain plus the surrounding solvent, where the all-atom energy function is averaged over the degrees of freedom that are lost when passing from the all-atom to the simplified system [i.e., the degrees of freedom of the solvent, the dihedral angles (χ) for rotation about the bonds in the side chains, and the torsional

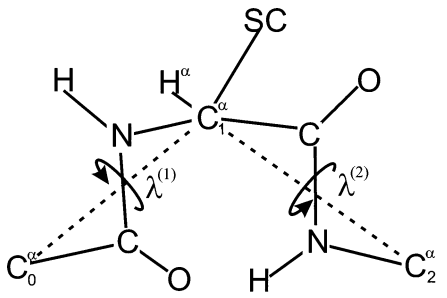


Figure 2. Definition of the dihedral angles $\lambda^{(1)}$ and $\lambda^{(2)}$) for rotation of the peptide groups about the $C^{\alpha}-C^{\alpha}$ virtual bonds (dashed) of a peptide unit.

angles (λ) for rotation of the peptide groups about the $C^{\alpha}-C^{\alpha}$ virtual bonds (the λ's are defined in the next section)]. This enabled us to derive the multibody terms U_{corr}^{m}, m = l, 2,..., N_{corr} by a generalized cumulant expansion of the RFE developed by Kubo.[56] The internal parameters of the individual U's were derived by fitting the resulting analytical expressions to the RFE surfaces of model systems[46] or by fitting the calculated distribution functions[45] to those determined from the PDB,[47] while the w's (the weights of the energy terms) were calculated by Z-score optimization of the training proteins.[45,46,57−59] The force field is now able to predict the structures of proteins containing both α-helical and β-sheet structures with reasonable accuracy, as assessed by tests on model proteins[50,52,54,55] as well as in the CASP3[52,53] and CASP4[54] blind prediction experiments.

The torsional terms, U_{tor} are very important, because they define the local conformational preferences of the polypeptide chain (e.g., the energetic preference for right-handed helices over left-handed helices). These potentials originate from the coupling between the local interactions within the neighboring peptide units.[46,55] Until now, we determined these potentials by averaging the all-atom energy calculated with the use of the ECEPP/3 force field.[31] However, as mentioned in the Introduction, better quality of the all-atom energy surfaces is necessary to reproduce the energetics of local interactions correctly.

2.2. Definition and Calculation of the Ramachandran Surfaces. Because our goal is to obtain energy terms in the UNRES force field by averaging the all-atom energy surface over the secondary degrees of freedom with conservation of the UNRES geometry, the dihedral angles ϕ and ψ commonly used as variables in the description of the Ramachandran surfaces are not appropriate; their variation changes the shape of the polypeptide chain. Therefore, the angles $\lambda_i^{(1)}$ and $\lambda_i^{(2)}$ for rotation of the peptide groups about the $C^{\alpha}_{i-1}-C^{\alpha}_{i}$ and $C^{\alpha}_{i}-C^{\alpha}_{i+1}$ virtual-bond axes of a peptide unit centered at C^{α}_{i}, introduced by Nishikawa et al.[60] (Figure 2), are more appropriate. These angles are close to ϕ and ψ in values, except in regions near ($\lambda^{(1)} = -90°$, $\lambda^{(2)} = 90°$) and ($\lambda^{(1)} = 90°$, $\lambda^{(2)} = -90°$);[60] they would be exactly equal to ϕ and ψ for a "linear" peptide group with the C′ and N atoms lying on the $C^{\alpha}-C^{\alpha}$ lines. Variation of $\lambda^{(1)}$ and $\lambda^{(2)}$ by rotation about the virtual bonds leaves the virtual-bond dihedral angles (γ) and, thereby, the shape of the virtual-bond chain, unchanged. It should be noted that, in the case of proline, $\lambda^{(1)}$ is determined by $\lambda^{(2)}$, because of the constraint imposed by the presence of a rigid pyrrolidine ring, and the Ramachandran map is therefore a function of $\lambda^{(2)}$ alone.

The energy maps in $\lambda^{(1)}$ and $\lambda^{(2)}$ were calculated by using an ab initio method with the 6-31G(d,p) basis set for the following model systems: *trans*-N-acetyl-glycyl-*trans*-N′-methylamide (Ac−Gly−NHMe), *trans*-N-acetyl-glycyl-N,N′-dimethylamide (Ac−Gly−NMe_2), *trans*-N-acetyl-L-alanyl-*trans*-N′-methylamide (Ac−Ala−NHMe), *trans*-N-acetyl-L-alanyl-N,N′-dimethylamide (Ac−Ala−NMe_2), *trans*-N-acetyl-L-prolyl-*trans*-N′-methylamide (Ac−Pro−NHMe), and *trans*-N-acetyl-L-prolyl-N,N′-dimethylamide (Ac−Pro−NMe_2). The NHMe-blocked systems represent cases in which the next peptide residue in the chain is not a proline, while the NMe_2-blocked systems represent cases in which the next residue is a proline. First, initial geometries of all systems were prepared by using the standard ECEPP/3 amino acid-residue geometry from the ECEPP/3 database.[31] The grid in the ($\lambda^{(1)}$, $\lambda^{(2)}$) space was 15° (a total of 576 points, which were then used to derive the torsional potentials). As mentioned above, for proline, $\lambda^{(2)}$ was the only variable, while $\lambda^{(1)}$ was calculated for each $\lambda^{(2)}$ to satisfy the constraints arising from the presence of the pyrrolidine ring as in our earlier work.[61] For all structures, constrained energy optimization was carried out at the RHF level, with the angles $\lambda^{(1)}$ and $\lambda^{(2)}$ constrained to those from the respective grid point. For optimized geometries, single-point second-order Møller−Plesset perturbation (MP2) calculations were carried out to account for electron correlation. For comparison, the energy surfaces of the Ac−Pro−NHMe system were also calculated by using the above optimization procedure with the semiempirical AMl and PM3 methods included in the MOPAC package,[28] without, however, the MP2 procedure, which is not applicable to semiempirical methods. All ab initio calculations were carried out with the program GAMESS.[62]

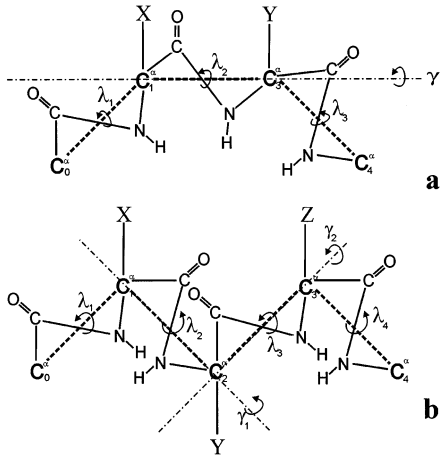


Figure 3. Illustration of the model terminally blocked dipeptides (a) and tripeptides (b) constructed to compute the integrals of eqs 7 and 8. *X*, *Y*, and *Z* denote Ala, Gly, or Pro.

2.3. Calculation of the Torsional and Double-Torsional Potentials. The torsional and double-torsional potentials can be calculated from the Ramachandran maps of the terminally-blocked single residues as the second- and the third-order RFE factors, respectively.[46] Let us consider model terminally-blocked di- and tripeptides sketched in Figure 3, parts a and b, respectively. We compute the contributions of backbone local interactions to their RFE surfaces by averaging the Ramachandran surfaces over the angles λ for rotation of the peptide groups about the $C^{\alpha}-C^{\alpha}$ virtual-bond axes subject to fixed virtual-bond dihedral angles (γ). Let us define the angles λ_i, i = 1,2,.. .n, where n is the number of residues in a chain by eqs 2 and 3.[46]

$$\lambda_i = \lambda_{i+1}^{(1)} \quad i = 1, 2, \cdots, n-2 \tag{2}$$

$$\lambda_{n-1} = \lambda_{n-1}^{(2)} \tag{3}$$

Making use of eq 10 of ref 60, and of our definition of the virtual-bond dihedral angles *(γ)* (Figure 1), we obtain

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

$$\lambda_i^{(2)} = \gamma_{i+2} - \pi - \lambda_i \qquad i = 1, 2, \cdots, n-2 \tag{5}$$

$$\lambda_{n-1}{}^{(2)} = \lambda_{n-1} \tag{6}$$

Following ref 46, we define the energy, $E_{\mathrm{loc};i}$, of backbone local interactions of the amino acid residue centered at C_i^α as the energy of all interactions at the all-atom level, between all backbone atoms starting at C_{i-1}^α and ending at C_{i+1}^α. For an alanine-type residue, we additionally include the C_i^β and the H_i^α atoms in the interaction list; for a glycine-type residue, we include $H_i^{\alpha 1}$ and $H_i^{\alpha 2}$; for a proline-type residue, we include all atoms of the pyrrolidine ring. Thus, $E_{\mathrm{loc};i}$ (hereafter referred to as e_X, X being the type of residue i (Gly, Ala, or Pro)) can be calculated effectively as the Ramachandran surface of residue X. Following the general theory presented in ref 46, the torsional potentials for rotation about the $C^\alpha-C^\alpha$ virtual bonds between residue types X and Y (Figure 3a), where X, Y = Gly, Ala, or Pro (with Ala representing all residue types except glycine and proline) are defined by eq 7.

$$\begin{aligned} U_{\mathrm{XY}}(\gamma) = -RT \ln\Bigg\{ & \frac{1}{(2\pi)^3} \int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \\ & \exp\Big[-\frac{1}{RT}[e_{\mathrm{X}}(\lambda_1, \gamma - \pi - \lambda_2) + e_{\mathrm{Y}}(\lambda_2, \lambda_3)]\Big] d\lambda_1 d\lambda_2 d\lambda_3 \Bigg\} + \\ & RT \ln\Bigg\{\frac{1}{(2\pi)^3}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \exp\Big[-\frac{1}{RT} e_{\mathrm{X}}(\lambda_1, \lambda_2)\Big] d\lambda_1 d\lambda_2\Bigg\} + \\ & RT \ln\Bigg\{\frac{1}{(2\pi)^3}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \exp\Big[-\frac{1}{RT} e_{\mathrm{X}}(\lambda_2, \lambda_3)\Big] d\lambda_2 d\lambda_3\Bigg\} \end{aligned} \tag{7}$$

where e_X and e_Y are the Ramachandran surfaces for the terminally-blocked residues of type X and Y, respectively, R is the gas constant, T is the absolute temperature, and use was made of eqs 2−6. When Y = Pro, the Ramachandran map of Ac−X−NMe_2 is used; otherwise, that of Ac−X−NHMe is entered into the expression. It should be noted that the last two double integrals over the Ramachandran surfaces of residues X and Y are constants independent of γ; therefore, they do not actually contribute to the expression for UNRES energy (eq 1), which has the sense of a relative energy only. The temperature was set at T = 298 K in all calculations.

The definition of torsional potentials given by eq 7 is different from that of the pioneering work on UNRES force fields by Levitt[34] and from our earlier UNRES force field.[40] In our earlier paper,[40] the energy surfaces of the *whole* terminally-blocked dipeptides were calculated and averaged over the λ angles explicitly; Levitt[34] did the averaging by varying the dihedral angles ϕ and ψ with constraints imposed on the angle γ. Then, all nonlocal contributions arising from the interactions between peptide groups 0 and 2, as well as between the side chains X and Y, were subtracted to obtain the torsional potentials. This results in torsional potentials that contain part of the correlation contributions arising from the coupling between the Ramachandran surfaces and the electrostatic interactions between peptide groups p_0 and p_2, which are now accounted for by explicit correlation contributions in UNRES. A consequence of this fact is that those earlier torsionals contain a minimum corresponding to a right-handed α-helical conformation, while the torsionals defined by eq 7 do not; this minimum appears only when the third-order correlation contribution, describing the coupling between backbone local and backbone-electrostatic interactions of a terminally-blocked Ala−Ala dipeptide, is added to the torsional potentials defined by eq 7. Thus, eq 7 provides the "pure" contributions arising from the coupling between the Ramachandran surfaces of consecutive residues alone. It should also be noted that this newer method for computing the torsional potentials is much less expensive than calculating the energy surfaces of *whole* terminally-blocked dipeptides; all that is needed are the Ramachandran surfaces of single terminally-blocked amino acid residues.

Similarly, in this work, we introduce the double-torsional potentials for rotation about two consecutive $C^\alpha-C^\alpha$ axes, between residue types X and Y, and Y and Z (Figure 3b), expressed by eq 8.

$$\begin{aligned} U_{\mathrm{XYZ}} = -RT \ln\Bigg\{ & \frac{1}{(2\pi)^4}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \\ & \exp\Big[-\frac{1}{RT}[e_{\mathrm{X}}(\lambda_1, \gamma_1 - \pi - \lambda_2) + e_{\mathrm{Y}}(\lambda_2, \gamma_2 - \pi - \lambda_3) + \\ & e_{\mathrm{Z}}(\lambda_3, \lambda_4)]\Big] d\lambda_1 d\lambda_2 d\lambda_3 d\lambda_4\Bigg\} - [U_{\mathrm{XY}}(\gamma_1) + U_{\mathrm{YZ}}(\gamma_2)] + \\ & \mathrm{RT} \ln\Bigg\{\frac{1}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \exp\Big[-\frac{1}{RT} e_{\mathrm{X}}(\lambda_1, \lambda_2)\Big] d\lambda_1 d\lambda_2\Bigg\} + \\ & RT \ln\Bigg\{\frac{1}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \exp\Big[-\frac{1}{RT} e_{\mathrm{Y}}(\lambda_2, \lambda_3)\Big] d\lambda_2 d\lambda_3\Bigg\} + \\ & RT \ln\Bigg\{\frac{1}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \exp\Big[-\frac{1}{RT} e_{\mathrm{Z}}(\lambda_3, \lambda_4)\Big] d\lambda_3 d\lambda_4\Bigg\} \end{aligned} \tag{8}$$

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 for rotation about the $C^\alpha-C^\alpha$ virtual-bond angles that cannot be accounted for by the single-torsional contributions.

The integrals in eqs 7 and 8 were calculated by a 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 Ramachandran 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 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 $\lambda^{(2)}$, and it is related to it by a nonlinear relationship.[61] Thus, even if $\lambda^{(2)}$ is on the grid, $\lambda^{(1)}$ need not be, in this case. The angle $\lambda^{(1)}$ is, in turn, related by eq 5 to $\lambda^{(2)}$ of the preceding residue; therefore, the last angle also, generally, takes values outside the grid points. Clearly, in other cases when $\lambda^{(1)}$ is unrestricted and if, as implemented, the virtual-bond dihedral angles γ are from a grid of the same spacing as the angles λ, then, by eq 5, all points taken to evaluate the integrals that occur in eq 7 and eq 8 are on the grid.

For use in UNRES, we express the torsional and double-torisonal potentials as one- and two-dimensional Fourier

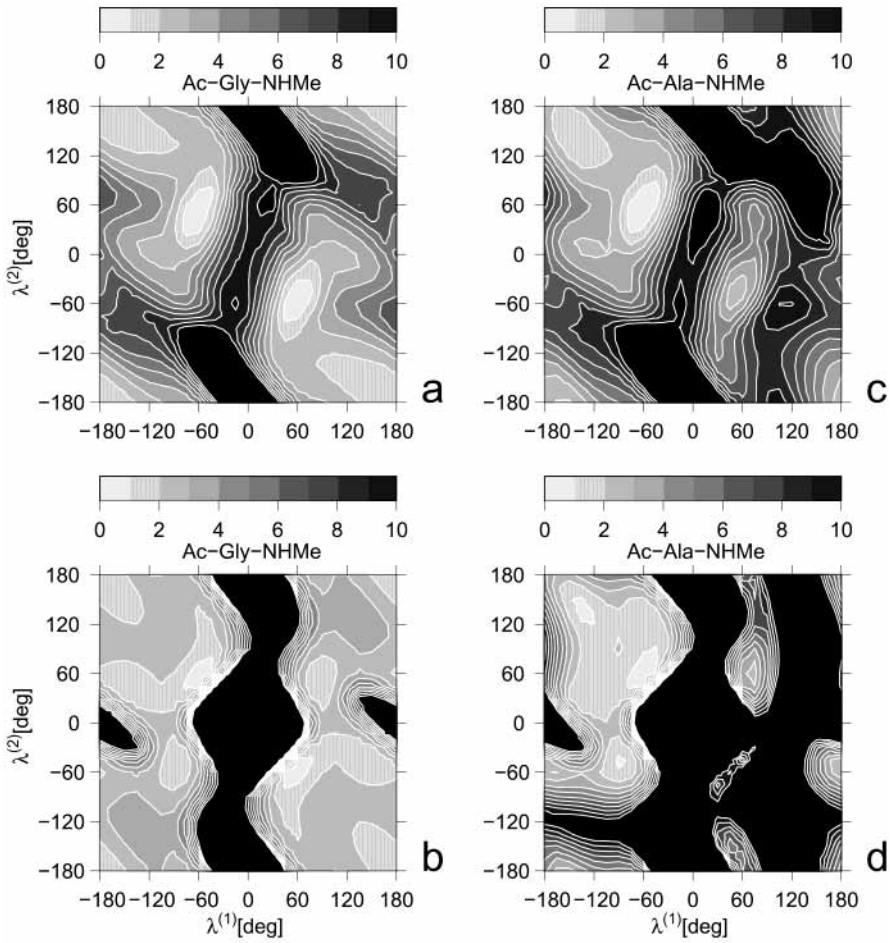


Figure 4. The Ramachandran maps of Ac−Gly−NHMe (a and b) and Ac−Ala-NHMe (c and d) obtained with the MP2/6-31G(d,p) ab initio method (a and c) and the ECEPP/3 force field (b and d) drawn with a 15° grid. Energies (kcal/mol) are expressed as relative values with respect to the global minimum of the surface, which corresponds to the C_{eq}^{7} conformation with ($\lambda^{(1)} = -60°$, $\lambda^{(2)} = 60°$).

series, respectively (eqs 9 and 10).

$$U_{XY}(\gamma) = a_0 + \sum_{i=1}^{n1} a_i \cos(i\gamma) + b_i \sin(i\gamma) \tag{9}$$

$$U_{XYZ}(\gamma_1, \gamma_2) = c_0 + \sum_{i=1}^{n2} c_i^{(1)} \cos(i\gamma_1) + d_i^{(1)} \sin(i\gamma_1) + \sum_{i=1}^{n2} c_i^{(2)} \cos(i\gamma_2) + d_i^{(2)} \sin(i\gamma_2) + \sum_{i=1}^{n2} \sum_{j=1}^{i-1} e_{ij} \cos[j\gamma_1 + (i-j)\gamma_2] + f_{ij} \cos[j\gamma_1 - (i-j)\gamma_2] + g_{ij} \sin[j\gamma_1 + (i-j)\gamma_2] + h_{ij} \sin[j\gamma_1 - (i-j)\gamma_2] \tag{10}$$

where $a_i - h_{ij}$ are coefficients and *n1* and *n2* are the orders of the expression; we used $n1 = 10$ and $n2 = 8$.

The coefficients in eqs 9 and 10 were calculated by fitting the U_{XY} and U_{XYZ} surfaces (obtained by numerical integration) to eqs 9 and 10, respectively. The target function consisted of the sum of the squares of the errors over the grid points in γ or 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 11).

$$F = \sum_{i=1}^{N} [U_i^{num} - U_i(X)]^2 + \alpha \int_{\gamma} \Delta U(\gamma) \exp(\Delta U(\gamma)) d\gamma \tag{11}$$

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 Ramachandran surfaces (eq 7 or 8), $U_i(X)$ denotes the value of *U* calculated by using the Fourier series (eq 9 or 10), $\Delta U(\gamma)$ is the difference between the *U* expressed by eq 9 or 10 and the value at that point obtained by linear interpolation between the grid points, and *a* is the weight of the "entropy" term; we used $\alpha = 0.2$. The value of *a* was set after several trial calculations to provide both smooth potentials and good fit to numerically calculated RFE values.

The target function of eq 11 was minimized with respect to the coefficients $a_0 - b_{10}$ (for single-torsional potentials; eq 9) or $c_0 - h_{8,8}$ (for double-torsional potentials; eq 10) by using the SUMSL procedure.[63] The "entropy" term (the last term in eq 11) 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. In our earlier work,[45,46] we introduced hyperbolic terms[45] or Lorentz-like terms[46] in addition to the Fourier-terms to the expressions for the torsional potentials in order to fit the torsional curves of X−Pro and Pro−X pairs, where the range of the values of the torsional potentials is considerably larger compared to systems not containing proline, to prevent the fitted potentials from oscillating between grid points and to ensure reasonable fits. This is no longer necessary when the maximum-entropy term is included in eq 11.

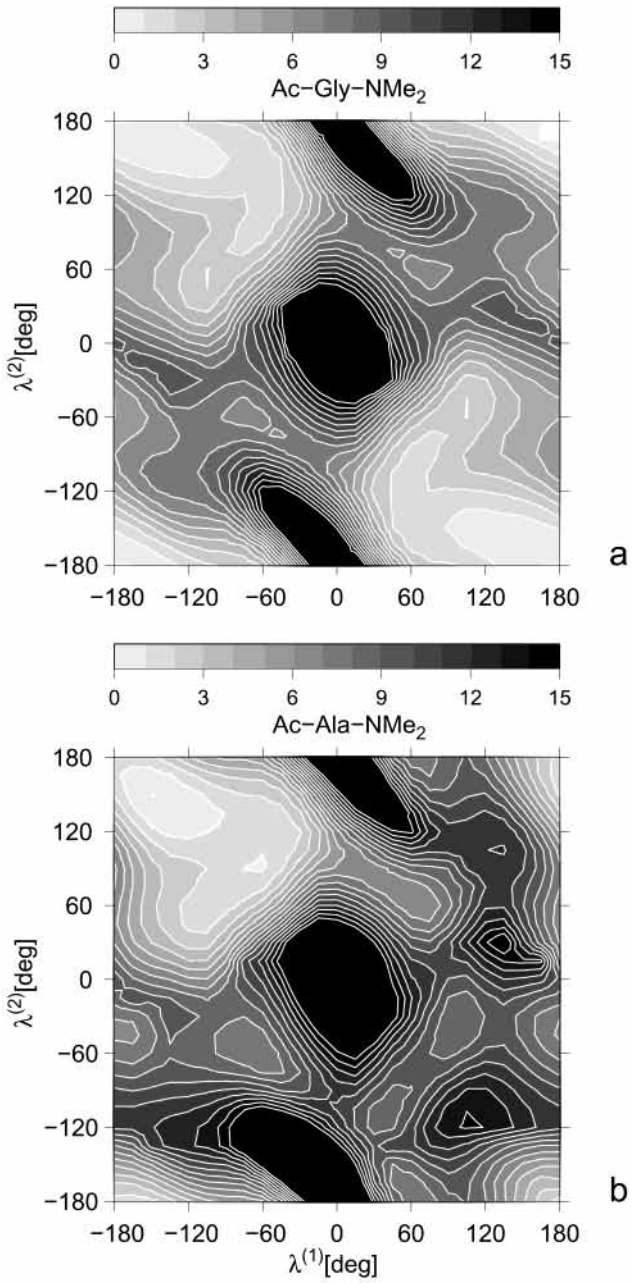


Figure 5. The Ramachandran maps of Ac−Gly−NMe_2 (a) and Ac−Ala−NMe_2 (b) obtained with the MP2/6-31G(d,p) ab initio method drawn with a 15° grid. Energies (kcal/mol) are expressed as relative values with respect to the global minimum of the surface, which corresponds to the E conformation with ($\lambda^{(1)} = -150°$, $\lambda^{(2)} = 150°$) for Ac−Ala−NHMe and ($\lambda^{(1)} = -180°$, $\lambda^{(2)} = 180°$) for Ac−Gly−NHMe.

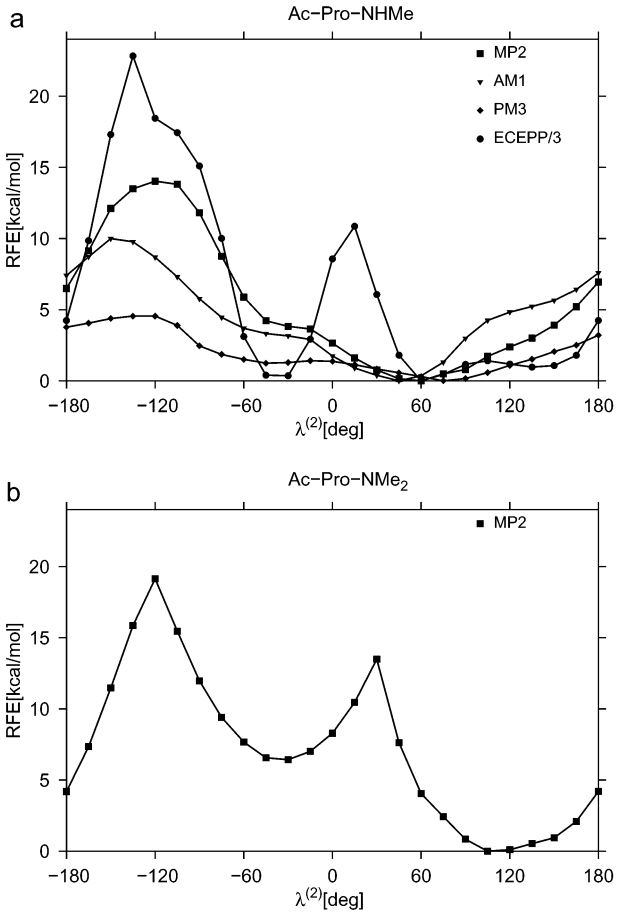


Figure 6. (a) Relative energy profiles (kcal/mol) of Ac−Pro−NHMe in the $\lambda^{(2)}$ angle, calculated with the MP2/6-31G(d,p) ab initio (squares), ECEPP/3 (circles), AM1 (triangles), and PM3 (diamonds) methods. (b) Relative energy profiles (kcal/mol) of Ac−Pro−NMe_2 in the $\lambda^{(2)}$ angle calculated with the MP2/6-31G(d,p) ab initio method.

3. Results and Discussion

3.1. Ramachandran Maps. Because the parameters of the torsional potentials were derived previously[46] from the Ramachandran maps calculated with the ECEPP/3 force field, we compare the Ramachandran maps obtained from the ab initio treatment in this study with those calculated previously with the ECEPP/3 force field. The Ramachandran maps of Ac−Gly−NHMe and Ac−Ala−NHMe (expressed as functions of the angles $\lambda^{(1)}$ and $\lambda^{(2)}$) calculated with the MP2/6-31G(d,p) ab initio method are shown in Figure 4, parts a and c, while the corresponding maps calculated with the ECEPP/3 force field are shown in Figure 4, parts b and d. It can be seen that the Ramachandran maps obtained for the Ac−Gly−NHMe and Ac−Ala−NHMe systems with the ab initio method share some features with the maps obtained with the ECEPP/3 force field. The first shared feature is the presence of a low-energy region in the upper left corner for Ac−Ala−NHMe and two such regions in the upper left and lower right corners for Ac−Gly−NHMe. In the notation of Zimmerman et al.,[2] these regions correspond to the C and E regions of the conformational space of terminally-blocked alanine and to the C, E, C*, and E* regions of the conformational space of terminally-blocked glycine. The minimum centered at ($\lambda^{(1)} = -60°$, $\lambda^{(2)} = 60°$) or ($\phi = -84°$, $\psi = 80°$) for Ac−Ala−NHMe corresponds to the C_{eq}^7 conformation; this is the global minimum of the energy surface for this molecule. The second minimum in the C + E region (with ($\lambda^{(1)} = -135°$, $\lambda^{(2)} = 135°$) or ($\phi = -156°$, $\psi = 154°$)) corresponds to the C_5 or extended conformation. For the Ac−Gly−NHMe surface, these minima have their counterparts with opposite signs of the angles.

There are a number of important differences between the MP2/6-31G(d,p) and ECEPP/3 Ramachandran surfaces. First, the ab initio surfaces of Ac−Gly−NHMe have no energy minima corresponding to the α_R and α_L conformations, in contrast to the ECEPP/3 surfaces, on which the α_R and the α_L minima appear at ($\lambda^{(1)} = \pm 85°$, $\lambda^{(2)} = \pm 45°$) or ($\phi = \pm 74°$, $\psi = \pm 32°$), with the "−" sign pertaining to the α_R and the "+" sign to the α_L conformation. The ab initio Ramachandran map of Ac−Ala−NHMe contains a shallow minimum corresponding to the α_L conformation with ($\lambda^{(1)} = 70°$, $\lambda^{(2)} = 55°$) or ($\phi = 56°$, $\psi = 44°$) (see Figure 4c) but no minimum

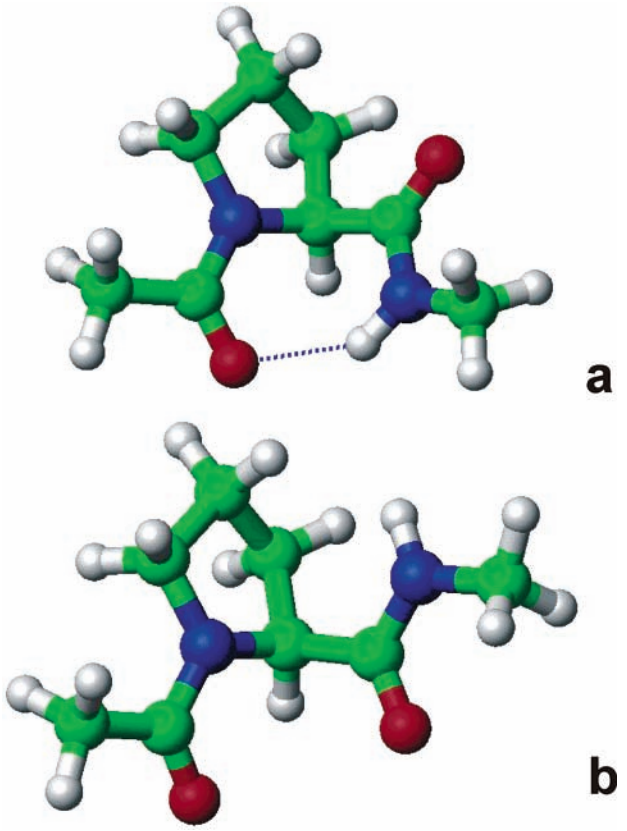


Figure 7. Characteristic conformations of the Ac−Pro−NHMe peptide optimized at the RHF/6-31G(d,p) level: energy maximum (b) with unfavorable interactions between the carbonyl oxygen atoms; hydrogen-bonded energy minimum (a). The dashed line in (a) represents the hydrogen bond.

corresponding to the α_R conformation. The same was observed by other workers for maps of formyl-blocked peptides.[24,25,27] It should be noted that the α_R conformation also appears as an energy minimum in the Ac−Ala−NHMe map calculated with the AMBER or CHARMM force fields when the distance-dependent dielectric constant or the dielectric constant $\epsilon = 4$ is used.[8] However, this minimum disappears for $\epsilon = 1$ (see Figures 2 and 7 in ref 8). The α_R and α_L minima also appeared on the Ramachandran surfaces of formyl-blocked glycine and alanine in the ab initio RHF study of Iwaoka et al.[26] that included solvation at the mean-field PCM and I−PCM level; they were not present when solvation was not included. Because $\epsilon > 1$ accounts partially for the effect of environment at the most rudimentary bulk-dielectric level, this effect can be attributed to the favorable electrostatic interactions of the peptide in the α_R or the α_L conformation with the solvent; in fact, the α_R and α_L conformations have the largest dipole moment.

Recently Vargas et al.[27] reported an extensive study of the conformational space of Ac−Ala−NHMe carried out at various levels of theory. They constructed the energy map using the DFT method at the BLYP/TZVP+ level, and subsequently, they carried out energy minimization at the MP2/aug-cc-pVDZ level for the minima found in the DFT-generated Ramachandran map; they also carried out single-point energy calculations at an even higher level of theory. Their Ramachandran map obtained by the DFT method (Figure 2 in ref 27) is similar to ours, and exactly the same six energy minima are obtained in both cases in the same order of energy: C_{eq}^7, C_5, C_{eq}^7, β_2 (in our case, it occurs at ($\lambda^{(1)} = -135°$, $\lambda^{(2)} = 15°$), or ($\phi = -145°$, $\psi = 47°$)), α_L, and α' (in our case it occurs at ($\lambda^{(1)} = -180°$, $\lambda^{(2)} = -30°$) or ($\phi = -169°$, $\psi = -32°$)). Approximate relative energies of these conformations obtained in our work (i.e., taken from the grid point closest to the position of the respective minimum) are 0, 1.32, 2.50, 2.83, 4.42, and 5.86 kcal/mol, and they agree with the energy relations between these conformations obtained in other studies at a high level of theory (Table 1 in ref 27). In both ref 27 and in our work, two other characteristic conformations reported in other studies (the α_L and β with ($\phi = -58°$, $\psi = 134°$) or ($\lambda^{(1)} = -35.4°$, $\lambda^{(2)} = 115.6°$)) are not energy minima.

The second difference between the MP2/6-31G(d,p) and ECEPP/3 maps is the lower energy barriers in the ab initio maps, which appears as flattening of most of the sterically forbidden H and H* regions. This difference arises mostly from the fact that the ECEPP/3 maps are obtained with rigid valence geometry, whereas the MP2/6-31G(d,p) energy surfaces are diabatic energy surfaces (i.e., the valence geometry and all torsions exept ϕ and ψ were relaxed). The same difference between adiabatic and diabatic Ramachandran surfaces was observed by Roterman et al. in their comparative study of empirical force fields.[8] Another consequence of relaxation of degrees of freedom other than $\lambda^{(1)}$ and $\lambda^{(2)}$ is the fact that the C_{ax}^7 conformation of Ac−Ala−NHMe, which is effectively forbidden with the ECEPP/3 force field,[2,8] is only 2.7 kcal/mol higher in energy than C_{eq}^7 (the global minimum) with the MP2/6-31G(d,p) calculations, and the C* region is much wider than that in the case of the ECEPP/3 force field. On the other hand, it should be noted that the energy difference between the C_{eq}^7 and C_{ax}^7 is only 1 kcal/mol with the AMBER force field when using the distance-dependent dielectric constant[6,8] and about 2 kcal/mol with $\epsilon = 1$ or $\epsilon = 4$.[8] This suggests that the AMBER force field has too soft a valence-angle-bending potential, as remarked by Roterman et al.[8] With the CHARMM force field, the energy difference between C_{eq}^7 and C_{eq}^7 conformations is about 3 kcal/mol, which is in agreement with the present results.

It is interesting to compare the ab initio maps of Ac−Gly−NMe_2 and Ac−Ala−NMe_2 with those of the methylamide-blocked residues of Figure 4. The dimethylamide maps are shown in Figure 5. It can be seen that the presence of an additional methyl group in the terminal amide increases the area of the high (steric) energy H and H* regions and eliminates the C^7 conformations. The last observation is understandable, because no amide proton capable of forming a 1,7-hydrogen bond (between the carbonyl oxygen atom of the acetyl group and the amide hydrogen of the *N*-methylamide group) can be formed in these systems.

The conformational energy curves of Ac−Pro−NHMe calculated with the different methods are shown in Figure 6a, and the ab initio energy curves for Ac−Pro−NMe_2 are shown in Figure 6b. It can be seen that the ab initio energy profile of the Ac−Pro−NHMe system does not exhibit a maximum at $\lambda^{(2)} = 10°$ ($\psi = 25°$) or a minimum at $-35°$ ($\psi = -56°$), which are present in the ECEPP/3 energy profile. The only ab initio minimum for Ac−Pro−NHMe occurs at $\lambda^2 = 55°$ ($\psi = 75°$) and corresponds to a C_{eq}^7 hydrogen-bonded conformation; for Ac−Pro−NMe_2, it is shifted to $\lambda^2 = 115°$ ($\psi = 139°$), because of the absence of the amide hydrogen. A maximum occurs for both Ac−Pro−NHMe and Ac−Pro−NMe_2 at $\lambda^2 = -120°$ ($\psi = -113°$); it is caused by an unfavorable electrostatic interaction between the two carbonyl oxygen atoms. The conformations corresponding to energy maxima and minima (calculated with the ab initio method) are shown in Figure 7. For Ac−Pro−NMe_2, there is an additional maximum at $\lambda^{(2)} = 30°$ ($\psi = 47°$),

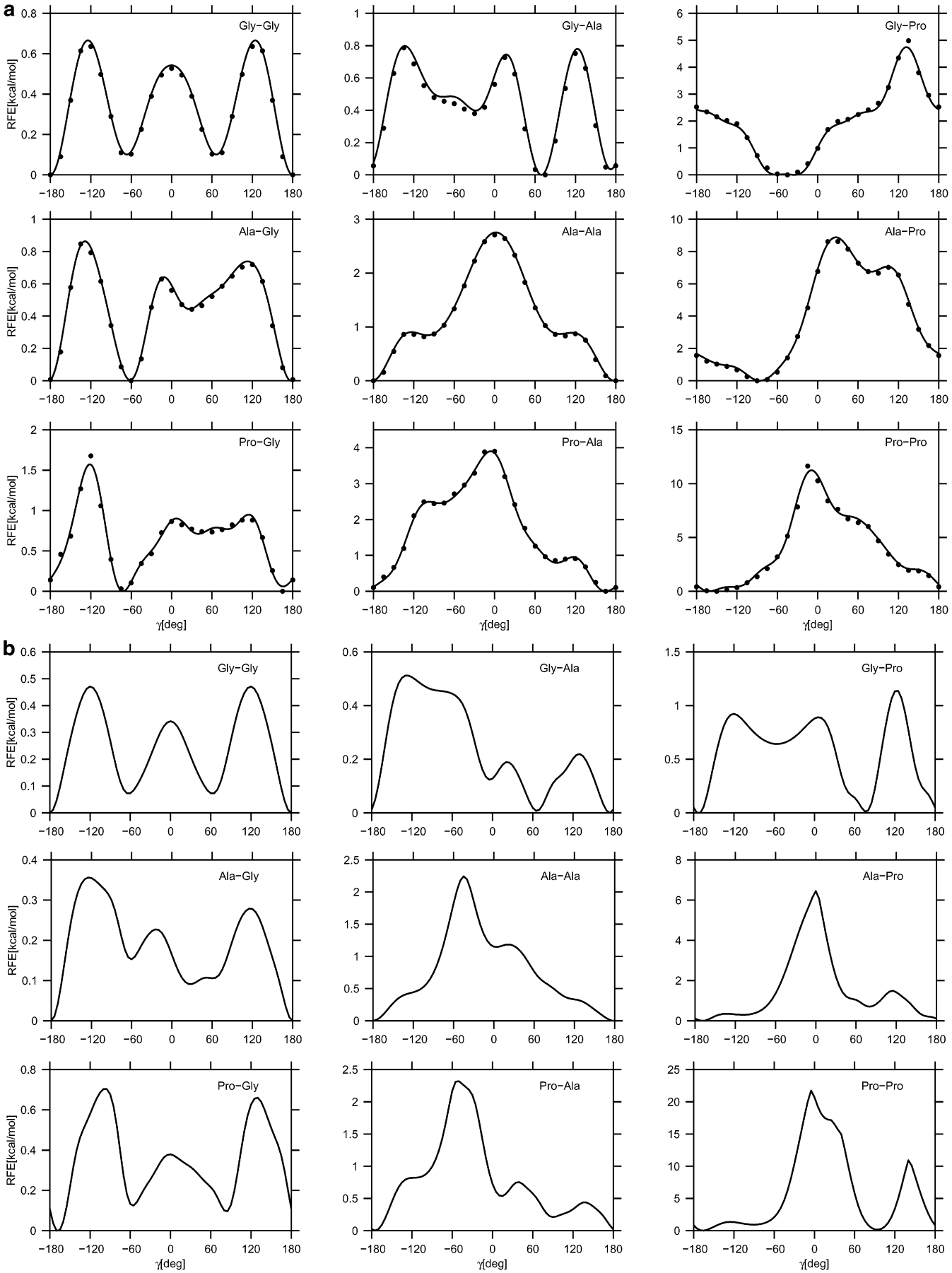
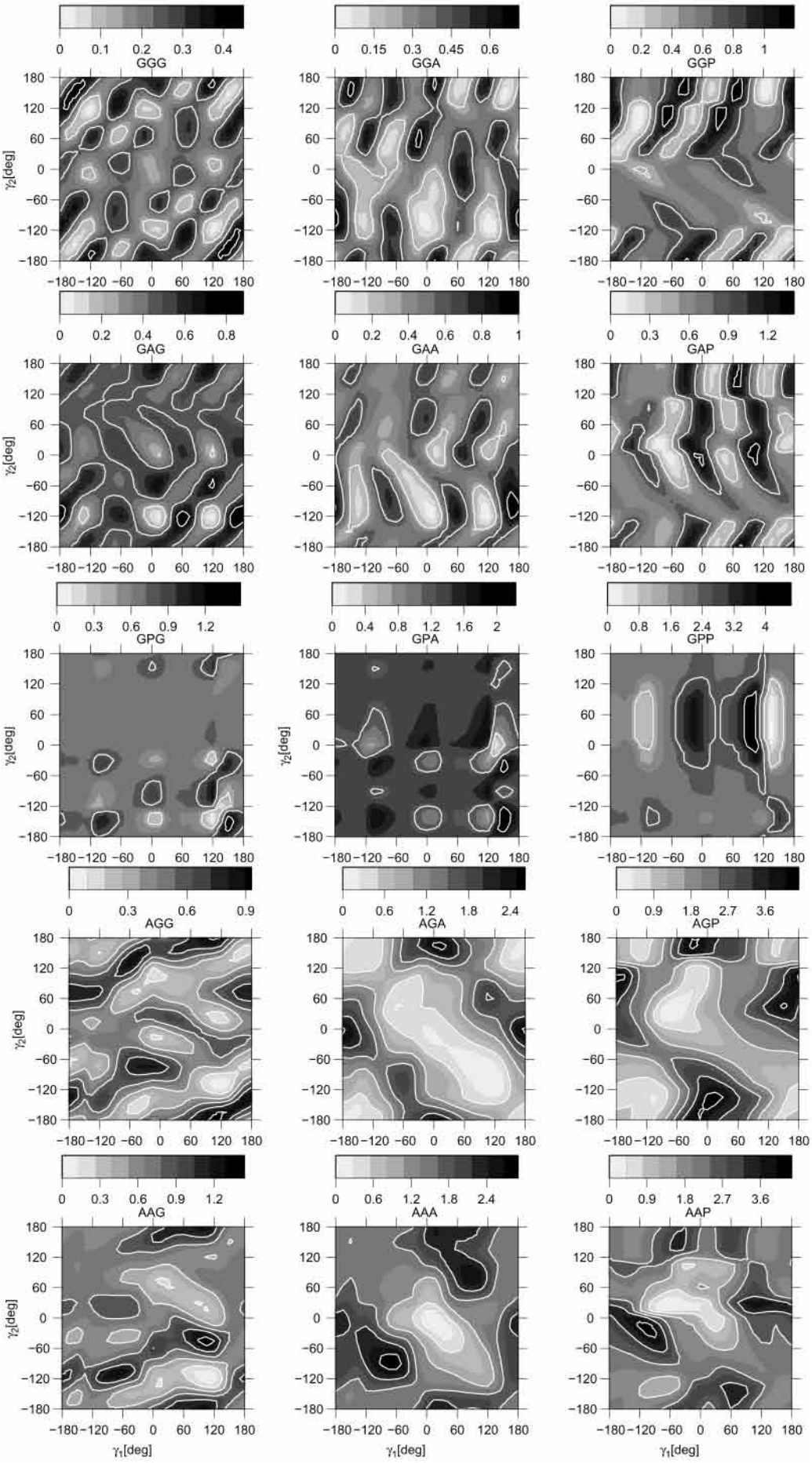


Figure 8. Torsional-potential curves for rotation about the $C^{\alpha}-C^{\alpha}$ virtual-bond axes ($U_{XY}(\gamma)$, where X and Y are Gly, Ala, or Pro; eq 7), obtained with the MP2/6-31G(d,p) ab initio method (a) and the ECEPP/3 force field (b). For the MP2/6-31G(d,p) profiles, the values calculated at points of the grid are shown as filled circles and the curves fitted with Fourier series are shown as solid lines.


GGG
GGA
GGP
GAG
GAA
GAP
GPG
GPA
GPP
AGG
AGA
AGP
AAG
AAA
AAP
γ2[deg]
γ1[deg]

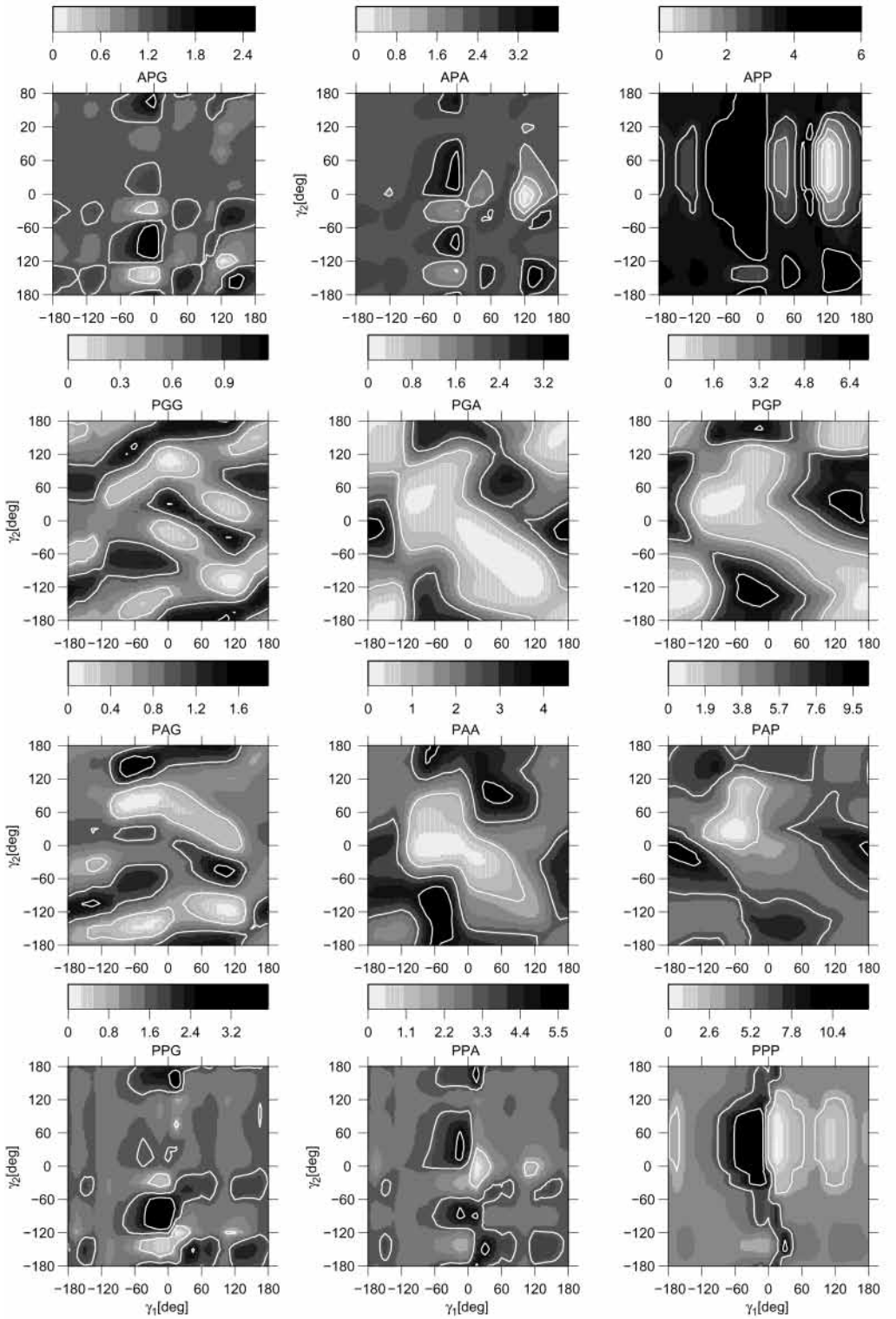


Figure 9. Surfaces of the double torsional potentials ($U_{XYZ}(\gamma_1,\gamma_2)$, where *X*, *Y*, and *Z* are Gly, Ala, or Pro; eq 8) obtained with the ab initio MP2/6-31G(d,p) method.

which arises from the repulsive interactions between the methyl groups of the acetyl and N,N′-dimethylamide parts of the molecule.

For comparison, we computed the energy profiles of Ac−Pro−NHMe, using the AMl and PM3 semiempirical methods. AMl and PM3 profiles of Ac−Gly−NHMe and Ac−Ala−NHMe were already computed and discussed in detail in the paper of Rodriguez et al.[20] The graphs are displayed in Figure 6a. It can be seen that the AMl method gives an energy profile qualitatively similar to that obtained with the MP2/6-31G(d,p) approach, with a smaller energy difference between the maximum and the minimum-energy conformation (10 kcal/mol compared to 14 kcal/mol). By contrast, the PM3 method gives no hydrogen-bonding minimum for Ac−Pro−NHMe in the region near $\lambda^{(2)} = 70°$ and three times lower energy difference between the maximum and minimum energy conformation, compared to the ab initio results. It should be noted that, in the case of Ac−Ala−NHMe,[20] the PM3 method does give the β conformation (called Π_L in ref 20) with ($\phi = -85.5°$, $\psi = 152.1°$) or ($\lambda^{(1)} = -63.5°$, $\lambda^{(2)} = 129.1°$) and not the

C_{eq}^{7} conformation as the global minimum on the Ramachandran surface; therefore, the angle ψ (the analogue of $\lambda^{(2)}$ in our representation) becomes more extended as in the PM3 energy surface of Ac−Pro−NHMe. The PM3 method, therefore, does not seem suitable to treat peptide and protein systems.

3.2. Torsional Potentials. The potentials of mean force calculated from the MP2/6-31G(d,p) ab initio and ECEPP/3 Ramachandran surfaces (eq 7) are shown in Figure 8, parts a and b, respectively. It can be seen that the profiles obtained from the two types of energy surfaces share some features. In particular, the Gly−Gly torsional potential has three minima at $\gamma = 180°$ and $\gamma = \pm 60°$, with quite similar energy maxima at $\gamma = \pm 120°$ between them (0.64 kcal/mol for the ab initio and 0.47 kcal/mol for the ECEPP/3 surface). The minimum at $\gamma = 180°$ corresponds to an extended conformation of the virtual-bond chain, while those at $\gamma = \pm 60°$ correspond to folded ones; after converting to all-atom chains, the last two conformations correspond to a type I′ or type III′ (for $\gamma = -60°$) or type I or type III (for $\gamma = 60°$) β-turn or, if the same value of the γ angle is repeated at least two times, to a left- or a right-handed α-helix. Conversely, the Ala−Ala torsional potential is different; the torsional potential calculated from the ECEPP/3 Ramachandran surface has a maximum at about $\gamma = -50°$, corresponding to the conformation of the C^{α} trace as in left-handed helices. This maximum in the torsional−potential curve calculated from the MP2/6-31G(d,p) Ramachandran surface is shifted to $\gamma = 0°$, and the torsional profile becomes almost symmetric, which means that the torsional potential discriminates against conformations with too small virtual-torsional angles (γ) and favors extended conformations, without giving preference to the right-handed helices. This difference between the torsional-potential profiles calculated from the ab initio and ECEPP/3 Ramachandran surfaces is a reflection of the fact that the E* and C* basins of the MP2/6-31G(d,p) Ramachandran map of Ac−Ala−NHMe are much wider and thus more comparable in size with their E and C counterparts than in the ECEPP/3 Ramachandran surface (Figure 4, parts c and d). Therefore, the energetic preference for the right-handed over left-handed α-helices must be accounted for by means other than the virtual-bond torsional energy terms in UNRES, presumably by the correlation contributions.

The Pro−Pro torsional potential calculated with the ECEPP/3 force field exhibits two maxima at about $\gamma = -5°$ and at about $\gamma = 120°$, while the potential obtained with the MP2/6-31G-(d,p) ab initio method exhibits only one maximum at about $\gamma = -15°$, with a height about two times smaller. This is also a direct consequence of the difference between the energy profiles of a terminally blocked proline residue calculated with the ECEPP/3 force field and that calculated with the ab initio method.

3.3. Double-Torsional Potentials. The double-torsional potentials calculated with the MP2/6-31G(d,p) method are shown in Figure 9. It can be seen that the range of energies of the double-torsional potentials is similar to that of the single-torsional potentials, which suggests that they cannot be ignored in the UNRES force field. It is remarkable that the Ala−Ala−Ala double-torsional potential has a minimum at ($\gamma_1 = 15°$, $\gamma_2 = -15°$), corresponding to a γ-turn. Generally, the region centered at $\gamma_2 = -\gamma_1$ forms large basins in the double-torsional-potential surfaces of the AGA, AGP, AAA, AAP, PGA, PGP, PAA, and PAP triplets. This is understandable, because for opposite-sign values of γ_1 and γ_2, all possible 1,7-hydrogen bonds between the backbone atoms of the three consecutive residues can be formed simultaneously. For the AAA triplet, the minimum appears toward the lower right corner, because in this case, the central residue of the γ-turn is in the C and not the C* region. It can also be seen that, for the triplets mentioned above, the right-handed α-helical conformations (with $\gamma_1 = \gamma_2 \approx 48°$) lie on the slope in the energy surface; therefore, the α-helical conformations are not stabilized by virtual-bond double-torsional potentials. Because right-handed α-helices have been found to be the lowest-energy conformations in ab initio calculations of oligoalanine peptides[13,14] by empirical energy calculations[64,65] and by experiment,[66−68] they are therefore stabilized by the correlation terms, i.e., by the coupling between backbone-local and backbone electrostatic interactions and not by backbone-local interaction alone. This conforms with classical helix-coil transition theories, in which it is assumed that the formation of backbone hydrogen bonds is a necessary factor for both initiating helix formation and stabilizing a formed helix.[69]

DeWitte and Shakhnovich[70] determined the distribution of pairs of consecutive virtual-bond dihedral angles from the PDB.[47] However, these distributions are total distributions of pairs of virtual-bond angles (the pseudo-dihedrals in their terminology), without removing contributions arising from single virtual-bond angles (cf., eq 8). Moreover, their surfaces contain contributions from all interactions and not just from backbone local interactions (cf., the discussion in the Methods section). Therefore, their results cannot be compared directly to ours.

Conclusions and Further Directions

In this work, we determined Ramachandran energy surfaces of terminally-blocked amino-acid residues representing the amino acid residues in a polypeptide chain using high-level theory. On the basis of these maps, we calculated the single- and double-torsional potentials of mean force for rotation about the $C^{\alpha}-C^{\alpha}$ virtual bonds and fitted them to Fourier series for use in the UNRES force field.[46] Comparison of the torsional potentials with those determined earlier from the ECEPP/3 Ramachandran surfaces showed that they share some common features but are also significantly different. The most important difference is the near symmetry of the Ala−Ala ab initio torsional energy curve. It, therefore, appears essential to have high quality potential-energy surfaces to parametrize a coarse-grain force field, and errors in the parent single-residue all-atom energy surfaces propagate to the mean-field surfaces by numerical integration. The double-torsional-potential surfaces have a range of energies comparable to those of the single-torsional potentials, and their introduction into UNRES, therefore, appears to be necessary.

Clearly, revisions of the torsional potentials and introduction of double-torsional potentials alone are insufficient to revise the UNRES force field. Because the average electrostatic (U_{pp} in eq 1) and correlation ($U^{(2)}_{corr}$ in eq 1) terms were also derived from empirical all-atom energy surfaces of model peptide systems,[46] they too need to be recalculated using ab initio energy surfaces instead of the ECEPP/3 surfaces. Introduction of the revised-torsional and double-torsional potentials, as well as of the revised correlation terms, involves the need to re-determine the weights of the energy terms (eq 1) by Z-score optimization;[57−59] we also plan to refine the internal coefficients in the energy terms (such as the Fourier coefficients in eq 9 and 10) by allowing them to undergo small variations, e.g., up to 20% of their original values, to obtain a more robust energy function. Work on all of these subjects is currently being carried out in our laboratory.

Acknowledgment. This work is dedicated to the memory of our friend and colleague, Andreas C. Albrecht, for his seminal

contributions to physical chemistry. This work was supported by grants from the National Institutes of Health (GMB14312), the National Science Foundation (MCB00B03722), the Fogarty Foundation (TW1064), and Grant DS 8372B4B0138B2 from the Polish State Committee for Scientific Research (KBN). Support was also received from the National Foundation for Cancer Research. This research was conducted by using the resources of (a) the Informatics Center of the Metropolitan Academic Network (IC MAN) in Gdańsk, (b) the Interdisciplinary Center of Mathematical and Computer Modeling (ICM) at the University of Warsaw, (c) our own 40-processor Linux cluster at the Faculty of Chemistry, University of Gdańsk, and (d) the National Science Foundation Terascale Computing System at the Pittsburgh Supercomputer Center.

Note Added after ASAP Posting. This article was released ASAP on 3/13/2003. Changes were made to the Figure 7 caption, and the correct version was posted on 3/18/2003.

References and Notes

(1) Némethy, G.; Scheraga, H. A. *Q. Rev. Biophys.* **1977**, *10*, 239.
(2) Zimmerman, S. S.; Pottle, M. S.; Némethy, G.; Scheraga, H. A. *Macromolecules* **1977**, *10*, 1.
(3) Dill, K. A. *Biochemistry* **1990**, *29*, 7133.
(4) Ramachandran, G. N.; Ramakrishnan C.; Sasisekharan, V. *J. Mol. Biol.* 1963, *7*, 95.
(5) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
(6) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765.
(7) Roterman, I. K.; Gibson, K. D.; Scheraga, H. A. *J. Biomol. Struct. Dyn.* **1989**, *7*, 391.
(8) Roterman, I. K.; Lambert, M. H.; Gibson, K. D.; Scheraga, H. A. *J. Biomol. Struct. Dyn.* **1989**, *7*, 421.
(9) Schäfer, L.; Van Alsenoy, C.; Scarsdale, J. N. *J. Chem. Phys.* **1982**, *76*, 1439.
(10) Scarsdale, J. N.; van Alsenoy C.; Klimkowski, V. J.; Schäfer, L.; Momany, F. A. *J. Am. Chem. Soc.* **1983**, *105*, 3438.
(11) Böhm, H.-J.; Brode, S. *J. Am. Chem. Soc.* **1991**, *113*, 7192.
(12) Gould, I. R.; Kollman, P. A. *J. Phys. Chem.* **1992**, *96*, 9255.
(13) McAllister, M. A.; Perczel, A.; Császár, P.; Viviani, W.; Rivail, J. L.; Csizmadia, I. G. *J. Mol. Struct. THEOCHEM* **1993**, *107*, 161.
(14) Viviani, W.; Rivail, J. L.; Perczel, A.; Csizmadia, I. G. *J. Am. Chem. Soc.* **1993**, *115*, 8321.
(15) Gould, I. R.; Cornell, W. D.; Hillier, I. H. *J. Am. Chem. Soc.* **1994**, *116*, 9250.
(16) Kang, Y. K. *J. Phys. Chem.* **1996**, *100*, 11589.
(17) Beachy, M. D.; Chasman, D.; Murphy, R. B.; Halgren, T. A.; Friesner, R. A. *J. Am. Chem. Soc.* **1997**, *119*, 5908.
(18) Cornell, W. D.; Gould, I. R.; Kollman, P. A. *J. Mol. Struct. THEOCHEM* **1997**, *392*, 101.
(19) Kaschner, R.; Hohl, D. *J. Phys. Chem. A* **1998**, *102*, 5111.
(20) Rodríguez, A. M.; Baldoni, H. A.; Suvire, F.; Vázquez, R. N.; Zamarbide, G.; Enriz, R. D.; Farkas, Ö.; Perczel, A.; McAllister, M. A.; Torday, L. L.; Papp, J. G.; Csizmadia, I. G.; *J. Mol. Struct. THEOCHEM* **1998**, *455*, 275.
(21) Philipp, D. M.; Friesner, R. A. *J. Comput. Chem.* **1999**, *20*, 1468.
(22) Jhon, J. S.; Kang, Y. K. *J. Phys. Chem. A* **2002**, *103*, 5436.
(23) Hudáky, I.; Baldoni, H. A.; Perczel A. *J. Mol. Struct. THEOCHEM* **2002**, *582*, 233.
(24) Perczel, A.; Hudáky, P.; Csizmadia, I. G. *J. Mol. Struct. THEOCHEM* **2000**, *500*, 59.
(25) Yu, C.-H.; Norman, M. A.; Schäfer, L.; Ramek, M.; Peeters, A.; van Alsenoy, C. *J. Mol. Struct.* **2001**, *567*−*568*, 361.
(26) Iwaoka, M.; Okada, M.; Tomoda, S. *J. Mol. Struct. THEOCHEM* **2002**, *586*, 111.
(27) Vargas, R.; Garza, J.; Hay, B. P.; Dixon, D. A. *J. Phys. Chem. A* **2002**, *106*, 3213.
(28) Stewart, J. J. P. *J. Comput.-Aided Mol. Design* **1990**, *4*, 1.
(29) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J. Comput. Chem.* **1986**, *7*, 230.
(30) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E., III; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. A. *Comput. Phys. Commun.* **1995**, *91*, 1.
(31) Némethy, G.; Gibson, K. D.; Palmer, K. A.; Yoon, C. N.; Paterlini, G.; Zagari, A.; Rumsey, S.; Scheraga, H. A. *J. Phys. Chem.* **1992**, *96*, 6472.
(32) Miertús, S.; Scrocco, E.; Tomasi, J. *Chem. Phys.* **1981**, *55*, 117.
(33) Foresman, J. B.; Keith, T. A.; Wiberg, K. B.; Snoonian, J.; Frisch, M. J. *J. Phys. Chem.* **1996**, *100*, 16098.
(34) Levitt, M. *J. Mol. Biol.* **1976**, *104*, 59.
(35) Crippen, G. M.; Viswanadhan, V. N. *Int. J. Pept. Protein Res.* **1984**, *24*, 279.
(36) Crippen, G. M.; Snow, M. E. *Biopolymers* **1990**, *29*, 1479.
(37) Kolinski, A.; Skolnick, J. *J. Chem. Phys.* **1992**, *97*, 9412.
(38) Kolinski, A.; Godzik, A.; Skolnick, J. *J. Chem. Phys.* **1993**, *98*, 7420.
(39) Skolnick, J.; Kolinski, A.; Brooks, C. L. III.; Godzik, A.; Rey, A. *Curr. Biol.* **1993**, *3*, 414.
(40) Liwo, A.; Pincus, M. R.; Wawak, R. J.; Rackovsky, S.; Scheraga, H. A. *Protein Sci.* **1993**, *2*, 1715.
(41) Kolinski, A.; Skolnick, J. *Proteins*: *Struct., Funct., Genet.* **1994**, *18*, 338.
(42) Kolinski, A.; Milik, M.; Rycombel, J.; Skolnick, J. *J. Chem. Phys.* **1995**, *103*, 4312.
(43) Crippen, G. M. *J. Mol. Biol.* **1996**, *260*, 467.
(44) Skolnick, J.; Kolinski, A.; Ortiz, A. R. *J. Mol Biol.* **1997**, *265*, 217.
(45) Liwo, A.; Pincus, M. R.; Wawak, R. J.; Rackovsky, S.; Ołdziej, S.; Scheraga, H. A. *J. Comput. Chem.* **1997**, *18*, 874.
(46) Liwo, A.; Czaplewski, C.; Pillardy, J.; Scheraga, H. A. *J. Chem. Phys.* **2001**, *115*, 2323.
(47) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F. Jr.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J. Mol. Biol.* **1977**, *112*, 535.
(48) Liwo, A.; Ołdziej, S.; Pincus, M. R.; Wawak, R. J.; Rackovsky, S.; Scheraga, H. A. *J. Comput. Chem.* **1997**, *18*, 849.
(49) Liwo, A.; Kaźmierkiewicz, R.; Czaplewski, C.; Groth, M.; Ołdziej, S.; Wawak, R. J.; Rackovsky, S.; Pincus, M. R.; Scheraga, H. A. *J. Comput. Chem.* **1998**, *19*, 259.
(50) Liwo, A.; Pillardy, J.; Kaźmierkiewicz, R.; Wawak, R. J.; Groth, M.; Czaplewski, C.; Ołdziej, S.; Scheraga, H. A. *Theor. Chem. Acc.* **1999**, *101*, 16.
(51) Liwo, A.; Lee, J.; Ripoll, D. R.; Pillardy, J.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 5482.
(52) Lee, J.; Liwo, A.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 2025.
(53) Orengo, C. A.; Bray, J. E.; Hubbard, T.; Lo Conte, L.; Sillitoe, I., *Proteins*: *Struct., Funct., Genet.* **1999**, *3*, 149.
(54) Pillardy, J.; Czaplewski, C.; Liwo, A.; Lee, J.; Ripoll, D. R.; Kaźmierkiewicz, R.; Ołdziej, S.; Wedemeyer, W. J.; Gibson, K. D.; Arnautova, Y. A.; Saunders, J.; Ye, Y.-J.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 2329.
(55) Liwo, A.; Pillardy, J.; Czaplewski, C.; Lee, J.; Ripoll, D. R.; Groth, M.; Rodziewicz- Motowidło, S.; Kaźmierkiewicz, R.; Wawak, R. J.; Ołdziej, S.; Scheraga, H. A. "UNRES − a UNRES force field for energy-based prediction of protein structure − origin and significance of multibody terms", In *RECOMB 2000, Proceedings of the Fourth Annual International Conference on Computational Molecular Biology*. Shamir, R., Miyano, S., Istrail, S., Pevzner, P. and Waterman, M., Eds., ACM: New York, 2000; pp 193−200.
(56) Kubo, R.; *J. Phys. Soc. Jpn.* **1962**, *17*, 1100.
(57) Lee, J.; Ripoll, D. R.; Czaplewski, C.; Pillardy, J.; Wedemeyer, W. J.; Scheraga, H. A. *J. Phys. Chem. B* **2001**, *105*, 7291.
(58) Pillardy, J.; Czaplewski, C.; Liwo, A.; Wedemeyer, W. J.; Lee, J.; Ripoll, D. R.; Arłukowicz, P.; Ołdziej, S.; Arnautova, Y. A.; Scheraga, H. A. *J. Phys. Chem. B* **2001**, *105*, 7299.
(59) Liwo, A.; Arłukowicz, P.; Czaplewski, C.; Ołdziej, S.; Pillardy, J.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 1937.
(60) Nishikawa, K.; Momany, F. A.; Scheraga, H. A. *Macromolecules* **1974**, *7*, 797.
(61) Liwo, A.; Pincus, M. R.; Wawak, R. J.; Rackovsky, S.; Scheraga, H. A. *Protein Sci.* **1993**, *2*, 1697.
(62) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. Jr. *J. Comput. Chem.* **1993**, *14*, 1347
(63) Gay, D. M. *ACM Trans. Math. Software* **1983**, *9*, 503.
(64) Ooi, T.; Scott, R. A.; Vanderkooi, G.; Scheraga, H. A. *J. Chem. Phys.* **1967**, *46*, 4410.
(65) Yan, J. F.; Vanderkooi, G.; Scheraga, H. A. *J. Phys. Chem.* **1968**, *49*, 2713.
(66) Alter, J. E.; Taylor, G. T.; Scheraga, H. A. *Macromolecules* **1972**, *5*, 739.
(67) Maxfield, F. R.; Alter, J. E.; Taylor, G. T.; Scheraga, H. A. *Macromolecules* **1975**, *8*, 479.
(68) Scheraga, H. A. *Pure Appl. Chem.* **1978**, *50*, 315.
(69) Poland, D.; Scheraga, H. A. *Theory of Helix-Coil Transitions in Biopolymers*, Academic Press: New York, 1970.
(70) DeWitte, R. S.; Shakhnovich, E. I. *Protein Sci.* **1994**, *3*, 1570.