Origin of Intrinsic 3_{10}-Helix Versus Strand Stability in Homopolypeptides and Its Implications for the Accuracy of the Amber Force Field

ANNA JAGIELSKA, JEFFREY SKOLNICK
Center for the Study of Systems Biology, School of Biology, Georgia Institute of Technology, 250 14th Street, Atlanta, Georgia 30318

Received 22 May 2006; Revised 9 July 2006; Accepted 16 July 2006
DOI 10.1002/jcc.20616
Published online 6 March 2007 in Wiley InterScience (www.interscience.wiley.com).

Abstract: Current all-atom force fields often fail to recognize the native structure of a protein as the lowest free energy minimum. One possible cause could be the mathematical form of the potential based on the assumption that the conformation of a residue is independent of its neighbors. Here, using quantum mechanical (QM) methods (MP2/6-31g**//HF/6-31g** and MP2/cc-pVDZ//cc-pVDZ//HF/cc-pVDZ), the intrinsic correctness of the gas phase terms (without solvation) of the Amber ff03 and ff99 potentials are examined by testing their ability to reproduce the relative 3_{10}-helix versus extended structure stabilities in the gas phase for 1–7-residue alanine, valine, leucine, and isoleucine homopolypeptides. The 3_{10}-helix versus extended state stability strongly depends on chain length and less on the amino acid identity. The helical conformation becomes lower in energy than the extended conformation for all tested peptides longer than two residues, and its stability increases with the increase of chain length. The ff03 potential better describes the 3_{10}-helix versus extended state energy than ff99 and also reproduces the curvature of the relative helix-extended state energies. Therefore, the mathematical form of the Amber potential is sufficient to describe the local effect of 3_{10}-helix versus extended structure stabilization in the gas phase. However, the energy curves are shifted and the backbone geometries differ compared with the QM results. This may cause significant geometric discrepancies between native and predicted structures. Therefore, extant molecular mechanics force fields, such as Amber, need refinement of their parameters to correctly describe helix-extended state energetics and geometry of major conformations.

 J Comput Chem 28: 1648–1657, 2007

Key words: helix versus strand stability; polypeptides; force field accuracy; quantum mechanical calculations; Amber potential

Introduction

Current all-atom force fields in many cases fail to recognize the native structure of a protein as the lowest free energy minimum, and they lack a correlation between energy and the "native-likeness" of decoy structures, a feature which is necessary for successful refinement of low-resolution models to high resolution [$\sim$2 Å root-mean-square-deviation (RMSD) to the native structure].[1,2] The source of this inaccuracy may be: (a) the mathematical form of the potential, (b) lack of important physical interactions in the potential, e.g., polarization or correlation effects (many body interactions),[3] or (c) inadequate parameterization of the force field, which does not include sufficient information about the global shape of the energy landscape. Most all-atom force fields rely on the Flory *isolated-pair hypothesis*,[4] assuming that the conformation of each residue is energetically insensitive to the conformation of its neighbors. The backbone conformation independence is the usual assumption in protein force-field development, where often only an alanine dipeptide[5,6] (Amber) or tetrapeptide[7] (OPLS-AA) is used as a representative conformation for the derivation of the backbone torsional parameters as well as for force field evaluation and refinement. Most protein force fields do not contain explicit correlation terms for the backbone conformation (torsional angles) of the neighboring residues. In contrast to the isolated-pair hypothesis, there are experimental[8] and statistical data from solved protein structures,[9,10] showing that the backbone conformation of a residue in the amino acid chain is

Correspondence to: J. Skolnick; e-mail: skolnick@gatech.edu

Contract/grant sponsor: National Institutes of Health (NIH); contract/grant number: RR-12255

correlated with the conformation and identity of the neighboring residues. These data reflect the influence of the environment of the protein, where the correlations of local effects with solvent and nonlocal effects are nonseparable. The question of whether each pair of φ, ψ backbone dihedral angles is independent of the conformation of its neighbors is crucial for understanding some of the driving forces in protein folding, especially, the interplay between the entropy lost and the formation of the local structure at the beginning of the folding process. Therefore, it is important to determine if local interactions between residues close in amino acid sequence alter the accessible conformational space, and whether force fields that do not include correlation terms between residues can reproduce these effects. Conceivably, ignoring some important correlations could result in inaccuracies of the relative conformational energies, thereby affecting the ability to identify the native structure as the lowest energy minimum. However, if the functional form of current molecular mechanics (MM) potentials is mathematically correct, then the failure to recognize the native structure as the global energy minimum could arise from incorrect details of the force-field parameters that should be fixable by reparameterization.

The goal of this work is to determine the dependence of the backbone conformation of a residue on its *local environment*, i.e., the number of neighboring residues, their identity and conformation, and to check whether extant force fields that ignore explicit correlation terms between the residues can reproduce this dependence. To investigate these issues, we use quantum mechanical (QM) calculations (MP2 and Hartree-Fock (HF) methods with the 6-31g**[11] and cc-pVDZ basis sets[12]) to capture local interactions in peptides, and do not consider nonlocal interactions (between residues that are distant in sequence but local in structure) or the role of solvent. We analyze the relative stability of helix versus extended conformations, its dependence on amino acid identity, and the origin of this stability, in particular, the role of hydrogen bonding. The calculations were carried out for poly(L-alanine) (pAla), poly(L-valine) (pVal), poly(L-leucine) (pLeu), and poly(L-isoleucine) (pIle) peptides consisting of 1–7 residues. To our knowledge, these are the largest polypeptide calculations at this level of theory (MP2/6-31g** and MP2/cc-pVDZ) including amino acids larger than alanine. The QM data are then compared with the results using the Amber ff99[5] and ff03[6] force fields to test if the gas phase part of these all-atom force fields are correct.

Methods

QM Calculations

Relative Stabilities of the Helical and Extended Conformations of the Peptides

We performed calculations for peptides consisting of 1, 3, 5, and 7 amino acids, for pAla, pVal, pLeu, and pIle, in the gas phase. Polypeptides were blocked with Ac and Nme groups, at the N- and C-terminal, respectively. Two basis sets were used, 6-31g**[11] and cc-pVDZ,[12] to check whether the results are robust. The geometry of each molecule was fully optimized using the HF method, starting from (a) extended ($\varphi = 180^\circ$, $\psi = 180^\circ$) and (b) α-helical ($\varphi = -60^\circ$, $\psi = -40^\circ$) conformations. The geometries of single-residue peptides in the helical conformation were optimized with the backbone dihedral angles constrained to $\varphi = -71^\circ$ and $\psi = -18^\circ$, because helical conformations are not minima for single residues (after geometry optimization, longer peptides adopted the 3_{10}-helix conformation; therefore, we froze single-residue peptides in the 3_{10}-helix conformation and used average reported values for the 3_{10}-helix[13,14]). The side chains were always positioned in their lowest-energy conformation. The energies of the optimized structures were calculated using the Möller-Plesset up to second order (MP2) method. The helical versus extended conformation stabilization energy was calculated as the energy difference between these conformations in a peptide consisting of the same number of residues. The accuracy of the procedure was estimated by comparing the relative energies of the minima of alanine dipeptide, calculated using this method with the high level QM calculations (MP4/cc-pVTZ) of Beachy et al.[15] The average error of the energies was 0.15 and 0.11 kcal/mol for the 6-31g** and cc-pVDZ basis sets, respectively. Thus, the accuracy and results for these basis sets were similar.

Pseudopeptides

To analyze the nature of the increasing stabilization of the helix with the increase of the number of residues, particularly, whether hydrogen bonding is the only effect stabilizing the helix, we performed similar calculations (single point MP2/6-31g** energy calculations) as for natural polypeptides, but replaced the peptide groups (NH—CO) with a CH=CH group in the geometry-optimized structures of the natural peptides. The backbone dihedral angles were kept at the optimized geometries of the real peptides, and the remaining geometry was reoptimized at the HF/6-31g** level to avoid nonphysical bond lengths and angles in the CH=CH groups. This reoptimization did not significantly change the short distances between the hydrogen atoms of the side chains and the backbone, which we identified as the major cause of the helix destabilization for pVal, pLeu, and pIle. Since the resulting pseudopeptides are isosteric with real peptides but cannot form hydrogen bonds,[16] any energy difference between the helix and extended conformations observed for these pseudopeptides would indicate that another effect, besides hydrogen bonds, e.g. excluded volume, also participates in the stabilization (or destabilization) of the helix.

MM Calculations

Relative Stabilities of the Helical and Extended Conformations of the Peptides

To check the ability of the all-atom force fields to reproduce the calculated QM relative stabilities of the helix versus extended structures in the gas phase, we performed gas phase calculations for all four polypeptides in the ff99[5] and ff03[6] Amber force fields, using the Amber8 package.[17] The structures were fully optimized, starting from (a) extended ($\varphi = 180^\circ$, $\psi = 180^\circ$) and (b) α-helical ($\varphi = -60^\circ$, $\psi = -40^\circ$) conformations. Only the geometries of single-residue peptides in the helical conformation

were optimized with the backbone dihedral angles constrained to $\varphi = -71^\circ$ and $\psi = -18^\circ$, as in the QM calculations, because helical conformations are not minima for a single residue. The cutoff for nonbonded and electrostatic interactions was 20 Å, the scaling factor for 1–4 electrostatic interactions was 1.2 and for the 1–4 van der Waals interactions the scaling factor was 2.0.

Results

Stability of the Helix versus Extended Conformations in Representative Homopolypeptides

The first issue addressed is whether in the gas phase the stability of the helix versus extended conformation differs for various amino acids. Previous studies analyzed the relative helix versus β-strand stability based on polyglycine[13] or pAla[18–20] peptides. The influence of single amino acid replacements in glycine pentapeptides[21] on helix stability were also studied, using mixed QM/semiempirical calculations (DFT/AM1). The relative stabilities of the most common secondary structure motifs of pAla and polyglycine (up to 8 residues) were studied by Perczel et al.[22] for fully optimized peptide geometries. At a medium level of QM (HF/6-311++g**) accuracy, they showed that the studied secondary structure motifs are intrinsically stable in the gas phase, and that the helical conformation (which in the gas phase is closer to a 3_{10}-helix than to an α-helix) is the most stable conformation for peptides longer than three-residues with its relative stability versus the extended conformation increasing with peptide length. Here, we apply a higher level of QM calculations (MP2), than in previous studies, and fully optimized geometries to investigate the dependence of helix versus extended structure stability on amino acid identity. We focus on alanine, leucine, isoleucine, and valine polypeptides. In native protein structures, alanine and leucine more frequently populate helices, whereas valine and isoleucine occur more often in an extended conformation.[9] Since these observations are in folded proteins, it is not apparent whether they reflect intrinsic amino acid preferences or reflect both nonlocal interactions (between separate secondary protein fragments that are local in space but not in sequence) and solvation effects. QM calculations of helix versus extended structure relative energies performed for isolated short peptides in the gas phase can determine to what extent the relative stability originates from the local interactions. Since we considered minimized peptide structures, only the energetic component of the helix versus extended conformation stability can be determined in this study.

With Increasing Chain Length, in Vacuum, Helices Become Much More Stable than Extended States

In Table1, the relative energies of helix versus extended structure (energies of extended conformers are always taken as the reference zero) for pAla, pVal, pLeu, and pIle, calculated at the HF and MP2 levels, with 6-31g** and cc-pVDZ basis sets are presented. The data for HF/6-31g** and MP2/6-31g** are also illustrated in Figures 1A and 1B, respectively. We used two different basis sets to check if the results are qualitatively independent of the basis set used. The relative helix versus β-strand energies for both basis sets are similar; the average* standard deviations between results obtained with 6-31g** and cc-pVDZ basis sets were 0.7, 0.3, 0.2, and 0.2 kcal/mol for pAla, pVal, pLeu, and pIle, respectively. For single-residue peptides, extended conformations related to a γ-turn ($\varphi \sim -60^\circ$, $\psi \sim 90^\circ$) and β-strand ($\varphi \sim -160^\circ$, $\psi \sim 160^\circ$) are the lowest energy minima, whereas the α-helix ($\varphi \sim -60^\circ$, $\psi \sim -40^\circ$) and 3_{10}-helix ($\varphi \sim -71^\circ$, $\psi \sim -18^\circ$) conformations are not. Rather, in agreement with previous work, they are higher in energy than extended conformations.[15,22] However, when the polypeptide chain becomes longer, the helical conformation becomes more and more stable for all four types of amino acids (Fig. 1, Table 1). In agreement with others,[13,18,20,22] helices in the gas phase adopt a 3_{10}-helix conformation (which we call "helical" for simplicity) rather than an α-helix, which, for shorter peptides (up to $\sim$20 residues),[18] is higher in energy and is not stable for lengths shorter than eight residues.[18] This is due to the formation of an additional hydrogen bond in a 3_{10}-helix compared with a canonical α-helix at the cost of breaking the parallel alignment of the hydrogen bonds. The stabilization of the helix is evident at the HF level but it is strongly enhanced by electron correlation (MP2 results) (Table 1). At the HF level, peptides longer than three residues become more stable in the helical rather than the extended conformation. After electron

Table 1. Relative Energies of the Helical versus Extended Structures (in kcal/mol), Calculated Using the Hartree-Fock (HF) and MP2 (MP2) Methods.

	Relative energy							
	pAla		pVal		pLeu		pIle	
N	HF	MP2	HF	MP2	HF	MP2	HF	MP2
6-31g**[a]								
1[b]	3.1	2.5	3.3	3.2	3.1	3.1	2.8	2.8
3	0.4	−4.0	0.8	−2.4	0	−3.6	0.2	−3.7
5	−3.0	−11.9	−2.0	−10.5	−4.5	−13.4	−2.7	−12.2
7	−7.5	−20.9	−5.8	−20.7	−10.6	NA[c]	−6.5	NA[c]
cc-pVDZ[d]								
1[b]	3.0	2.1	3.4	3.2	3.2	3.0	2.9	2.7
3	0.2	−4.8	1.2	−2.6	0.2	−3.7	0.5	−3.8
5	−3.3	−13.2	−1.7	−11.2	−4.2	−13.9	−2.6	−12.9
7	−7.8	−22.6	−5.6	−21.5	−9.5	NA[c]	−6.4	NA[c]

N, number of residues; pAla, poly(L-alanine); pVal, poly(L-valine); pLeu, poly(L-leucine); pIle, poly(L-isoleucine).

[a]Calculations using the 6-31g** basis set and geometries fully optimized at the HF/6-31g** level.

[b]The geometry of single residue in helical conformation was optimized with backbone dihedral angles constrained to $\varphi = -71^\circ$ and $\psi = -18^\circ$.

[c]The system is too large to do the MP2 calculations with our available computational resources.

[d]Calculations using the cc-pVDZ basis set and geometries fully optimized at the HF/cc-pVDZ level.

* The standard deviations for each polypeptide were averaged over all standard deviations of relative energies calculated for each peptide length.

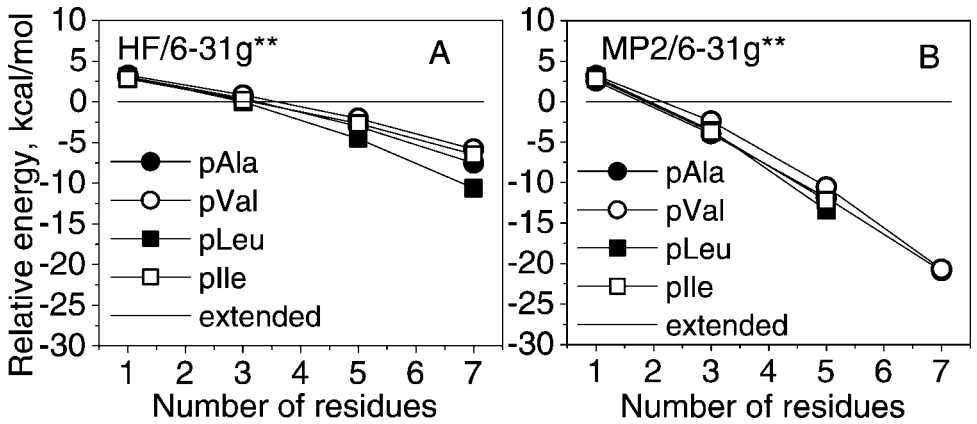

Figure 1. Relative energies of helical versus extended structures (kcal/mol), as a function of the number of residues for pAla, pVal, pLeu, and pIle, (A) at the HF/6-31g** level and (B) at the MP2/6-31g** level. The geometries were fully optimized at the HF/6-31g** level. The energy of the extended structures (extended) for each type of polypeptide is taken as the reference value.

correlation is included, even three-residue long helices are more stable than extended conformations. As shown in Table 1 and Figure 1, the stabilization of the helical versus extended conformation strongly depends on the number of residues in the chain, and in agreement with previous results for polyalanine and polyglycine,[13,18,22] cooperatively increases as the chain length increases. Next, we examine how this stability depends on the identity of amino acids and whether it is driven only by hydrogen bonding.

Differences in the Helix Stability of Different Amino Acids are Small

We next analyze how the relative stability of the helix versus β-strand depends on amino acid identity. As is evident from Figure 1 and Table 1, helical conformations of pVal and pIle are less stable than in the pAla helix, whereas pLeu is more stable compared with the helical pAla. A similar order of relative helix versus extended state stabilities was reported for glycine pentapeptides, where a single glycine was replaced by alanine, valine, leucine, or isoleucine.[21] Such differences in local secondary structure propensities are qualitatively similar to the differences in the statistically observed occurrence of analyzed amino acids in helices and β-strands in protein structures.[9] However, these differences do not exceed 1.6 (2.4) kcal/mol at the MP2/6-31g** (MP2/cc-pVDZ) level when compared with the helix stability of pAla and become only a small fraction of the total helix stabilization energy for longer peptides.

We investigated the reason for the observed differences in the helix versus β-strand stability by analyzing the average hydrogen bond lengths in helices (Table2, helix) and the average H—O backbone atom distances in the extended conformations (Table 2, β-strand). The smaller relative stability of the helix of pVal and pIle is a consequence of longer average hydrogen bonds, compared with pAla (by $\sim$0.1 and $\sim$0.2 Å for pIle and pVal, respectively), and destabilization of the helix (Table 2). The helical conformation of pLeu is more stable than in pAla. In this case, the difference in relative helix stability is not due to differences in helix structure (hydrogen bonds in the helix of pLeu are of similar length as in pAla; Table 2, helix). Rather, the destabilization of the extended conformation (which has longer H—O distances by $\sim$0.1 Å for pLeu compared with pAla) is the reason for the observed effective higher helix stabilization (Table 2, β-strand). Therefore, in the gas phase, amino acids differ in their helix/β-strand propensities, but these differences are small.

Relative Helix versus Extended Structure Stabilization Is Not Only Due to Hydrogen Bonding but Also Due to Backbone–Side Chain Interactions

The interactions that influence helix versus β-strand stability are now examined. Certainly, a significant part of helix stabilization originates from cooperative hydrogen bond formation[13,18] and increases with increasing helix length. The cooperative character of helix stabilization is reflected in the nonlinear stabilization of the energy versus the number of residues (Fig. 1). Is this the entire origin of differing helix versus strand stability? For amino acids with larger side chains, one could also expect excluded

Table 2. Average[a] Hydrogen-Bond Lengths in Helical Conformations (Helix), and Average[a] Distances between the Backbone Hydrogen and Oxygen Atoms (Belonging to the Same Residue) in the Extended Conformations (β-Strand) (Å) for Peptides Containing 1–7 Residues.[b]

	Helix				β-Strand			
N	pAla	pLeu	pIle	pVal	pAla	pLeu	pIle	pVal
1	–	–	–	–	2.21	2.30	2.25	2.25
3	2.32	2.33	2.42	2.53	2.20	2.30	2.25	2.25
5	2.31	2.32	2.47	2.52	2.19	2.28	2.25	2.24
7	2.29	2.31	2.49	2.46	2.18	2.28	2.25	2.24

N, number of residues; pAla, poly(L-alanine); pLeu, poly(L-leucine); pIle, poly(L-isoleucine); pVal, poly(L-valine).

[a]Averaging was over all the hydrogen bonds (in helical conformation) and H—O distances (in extended conformation) in the peptide of the given length.

[b]The geometries were optimized at the HF/6-31g** level.

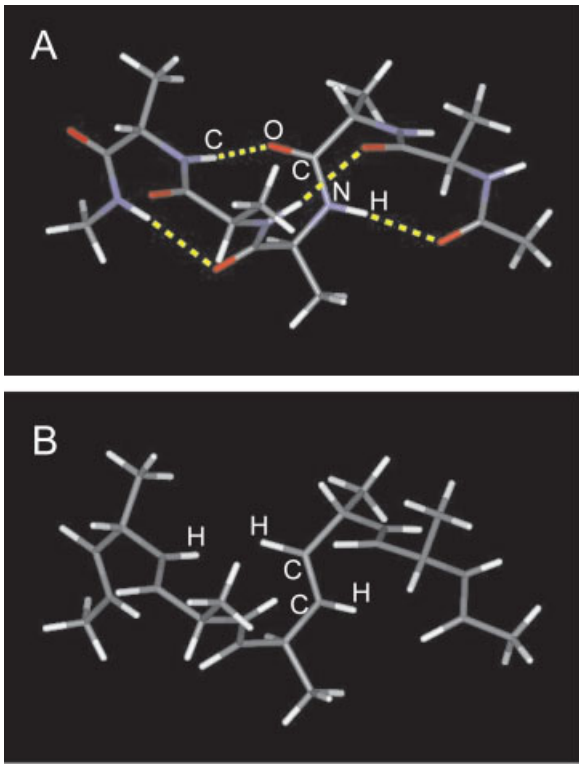

Figure 2. Natural peptide (A) and the related pseudopeptide (B) in the helical conformation. In the natural peptide, hydrogen bonds are formed (yellow lines); in pseudopeptides, the peptide groups NH—CO are replaced with CH=CH groups and hydrogen bonds cannot be formed. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

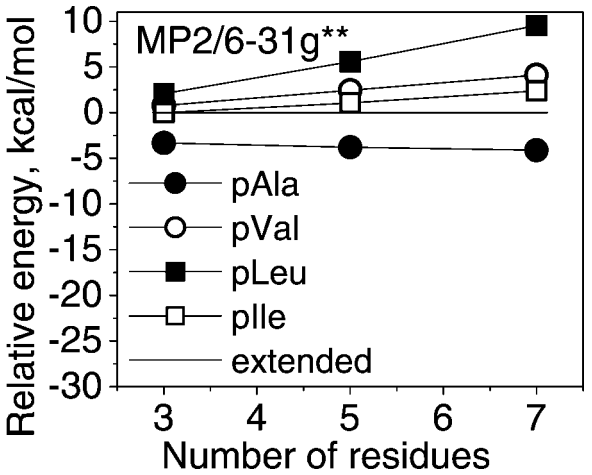

Figure 3. Pseudopeptides. Relative energies of helical versus extended structures (in kcal/mol), as a function of the number of residues for pseudo-pAla, pseudo-pVal, pseudo-pLeu, and pseudo-pIle at the MP2/6-31g** level. The energy of the extended structures (extended) for each type of polypeptide is taken as the reference.

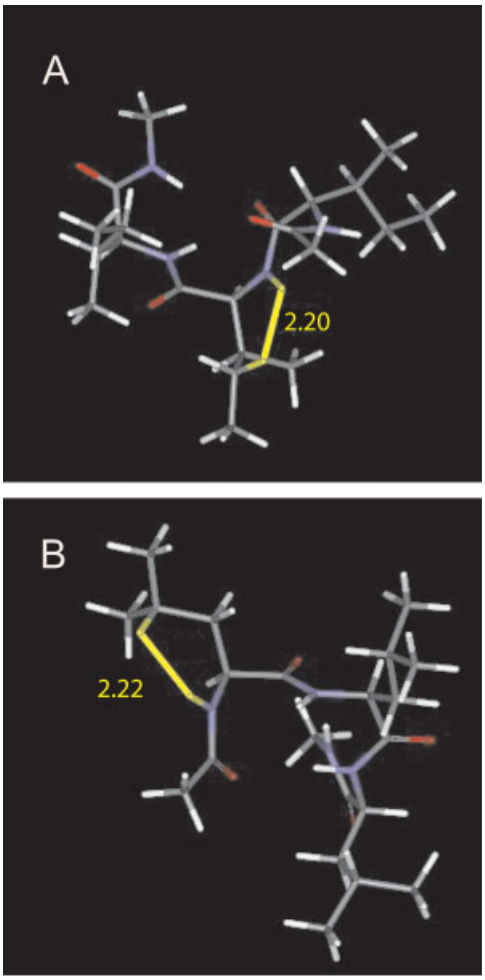

Figure 4. Hindrance between the backbone hydrogen atoms and the side chains (short distances [Å], marked in yellow) in (A) isoleucine and (B) leucine, in the helical conformation. The examples of the $(Ile)_3$ and $(Leu)_3$ are shown. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

volume effects, which could destabilize either the helical or extended structure. To analyze this issue, we performed QM calculations for pseudopeptides, where the peptide groups NH—CO were replaced by CH=CH groups in pAla, pVal,

Table 3. Pseudopeptides—The Relative Energies of the Helical versus Extended Structures (kcal/mol), Calculated Using the MP2/6-31g** Method and the Geometries Optimized at the HF/6-31g** Level.

	Relative energy			
N	pAla	pVal	pLeu	pIle
3	−3.3	0.8	2.1	0.0
5	−3.8	2.4	5.6	1.1
7	−4.1	4.1	9.5	2.4

N, number of residues; pAla, pseudo-poly(L-alanine); pVal, pseudo-poly (L-valine); pLeu, pseudo-poly(L-leucine); pIle, pseudo-poly(L-isoleucine).

Table 4. Average[a] φ, ψ Backbone Dihedral Angles (degree) and the RMSD (Å), with Respect to the QM Structures, Over C_α Atoms (RMSD) for the Helical and Extended Conformations of the Peptides Containing 1–7 Residues in Geometries Fully Optimized at the HF/6-31g** Level (HF) and Using the Amber ff03 (ff03) Force Field.

	pAla		pLeu		pIle		pVal	
N	HF	ff03	HF	ff03	HF	ff03	HF	ff03
Helical conformation[b]								
3								
φ	−69.7	−60.0	−72.2	−61.5	−75.8	−61.3	−76.2	−58.3
ψ	−16.4	−17.8	−14.5	−16.6	−8.8	−15.2	−8.6	−14.6
RMSD	0.2		0.2		0.3		0.2	
5								
φ	−66.9	−68.8	−68.5	−62.5	−72.4	−82.6	−72.4	−54.1
ψ	−19.3	−30.8	−17.3	−18.8	−10.9	−17.4	−10.9	−18.0
RMSD	0.9		0.3		1.0		0.4	
7								
φ	−65.6	−64.4	−67.0	−65.4	−71.1	−84.1	−70.1	−53.3
ψ	−20.3	−36.3	−19.5	−29.9	−11.9	−22.0	−12.5	−18.6
RMSD	1.2		1.0		1.8		0.6	
err[c]								
φ	8.8 (8.1)		7.0 (3.9)		22.1 (18.8)		17.5 (5.1)	
ψ	12.7 (5.9)		7.9 (5.4)		9.3 (8.1)		6.4 (2.8)	
Extended conformation								
1								
φ	−157.4	−159.9	−128.8	−139.6	−127.6	−140.3	−130.4	−147.2
ψ	159.4	166.3	149.7	157.7	158.3	157.8	159.0	157.5
RMSD[d]	0.1		0.1		0.1		0.1	
3								
φ	−157.2	−159.7	−128.5	−138.3	−130.8	−137.7	−129.4	−146.7
ψ	160.5	168.2	152.9	159.7	156.9	161.3	159.2	156.7
RMSD	0.1		0.2		0.2		0.2	
5								
φ	−157.0	−159.8	−129.3	−138.0	−128.6	−137.3	−130.4	−146.6
ψ	160.9	168.6	153.2	160.0	159.5	162.2	158.0	156.5
RMSD	0.2		0.3		0.2		0.2	
7								
φ	−157.0	−159.9	−133.3	−138.0	−128.2	−137.4	−130.4	−146.6
ψ	161.2	168.7	150.2	160.1	159.8	162.4	157.8	156.5
RMSD	0.3		0.5		0.3		0.3	
err[c]								
φ	2.7 (0.4)		7.3 (3.6)		8.9 (2.6)		16.4 (1.2)	
ψ	7.6 (0.5)		8.3 (2.8)		3.1 (1.7)		1.6 (0.8)	

N, number of residues; pAla, poly(L-alanine); pLeu, poly(L-leucine); pIle, poly(L-isoleucine); pVal, poly(L-valine).

[a]Averaging was over all the residues in the peptide of the given length.

[b]In helical conformation, 1-residue-long peptides are omitted because the helical conformation is not stable for the single residue and the structures were minimized with frozen dihedral angles.

[c]The average absolute differences in φ and ψ (degree) between the HF and ff03 results, and standard deviations from these averages (in parentheses [degree]).

[d]For single residues, the RMSD is over the backbone atoms: C, O, and N.

pLeu, and pIle (Fig. 2). This allowed us to exclude the hydrogen bond component from the interaction energy. Such pseudopeptides are isosteric with the real peptides (Fig. 2A), but cannot form hydrogen bonds[16] (Fig. 2B). The relative helix versus β-strand energies for pseudopeptides were calculated in the same way as for natural peptides (Fig. 3, Table 3). For pseudo-pAla, the curve is almost flat, which means that without hydrogen bonds the helices are not increasingly stabilized with increasing length. Therefore, the stabilization of the helix in real pAla comes entirely from the hydrogen bonds. In contrast to pAla, for all remaining amino acids, a growing stabilization of the β-strand is observed, which becomes more stable than the helix as the number of residues increases (Fig. 3, Table 3). This indicates that some steric (excluded volume) effects involving the large

side chains effectively destabilize the helical conformation. Our structural analyses of the helical and extended conformations of pVal, pLeu, and pIle (in natural peptides) reveal that the large methyl groups of the side chains push on the backbone atoms in both helical and extended conformations. In the extended conformation, this results in decreasing of the φ and ψ backbone dihedral angles, which are much smaller than for pAla ($\varphi = -157^\circ$, $\psi = 160^\circ$ for pAla; $\varphi = -130^\circ$, $\psi = 150^\circ$ for pLeu; $\varphi = -130^\circ$, $\psi = 160^\circ$ for pVal and pIle) (Table4, columns HF, Extended conformation). This effect occurs not only for longer peptides but even for a single residue (no side chain–side chain interactions are possible), indicating that it results from a backbone–side chain interaction, but not from side chain–side chain steric effects, as might be expected for large side chains. The backbone deformation is larger for pLeu than for pVal and pIle, compared with the backbone dihedral angles for pAla, because of the closer proximity of the γ-branched methyl groups than the β-branched methyl groups to the backbone (Table 4, columns HF, Extended conformation). The consequence of backbone deformation in the extended state is to increase the distances between the backbone H and O atoms, with a stronger effect for pLeu ($\sim$2.20 Å for pAla; $\sim$2.30 Å for pLeu; $\sim$2.25 Å for pVal and pIle) (Table 2, β-strand). The helical conformation of pLeu, pVal, and pIle is also affected by their large side chains, because of the hindrance between the hydrogen atoms of the large side chains and the hydrogen atom attached to the backbone nitrogen (the distances between these hydrogen atoms are as small as 2.2 Å; examples are shown in Fig. 4 for $(Ile)_3$ and $(Leu)_3$), which in all-atom force fields are usually in the repulsive region of the van der Waals interactions.[23] This hindrance is a cause of the growing stabilization of the β-strand over helical conformation for pLeu, pVal, and pIle pseudopeptides as the chain length increases (Fig. 3). As was the case for the extended conformation, in the helical state, steric effects also result from a backbone–side chain interaction within a single residue. In fact, the relative energy curves for pseudopeptides are almost linear (Fig. 3), which indicates that excluded volume effects for these peptides are not cooperative and are separable into single residue additive components. Therefore, for amino acids with bulky side chains, the helix versus extended structure stability is driven by two major competing effects: the strength of the hydrogen bonds that stabilize the helical conformation and excluded volume effects between the side chain and the backbone that effectively more strongly destabilize helices than extended conformations. Hydrogen bonds dominate over excluded volume effects, which results in the growing stabilization of the helix versus extended state with the increase in chain length.

How Well Does the Amber Force Field Reproduce the QM Helix versus β-Strand Stabilities?

One of the questions we want to answer in this work is whether the gas phase contribution (no solvation included) to all-atom force fields is correct. We calculated the relative energies of the helical versus extended state for pAla, pVal, pLeu, and pIle using Amber ff99 and ff03 force fields without solvation (i.e., only the gas phase potential is tested). In Figures 5A–5D, the relative helix versus extended state energies for the force fields together with the QM curves are presented for comparison. The results for the ff03 force field are significantly closer to the QM results than the results for ff99 (Fig. 5, Table5). We therefore focus on the better performing force field, ff03. For pAla, the relative helix stability is reproduced with high accuracy; the average error over all energy points compared with the QM data is 0.40 kcal/mol. For the remaining amino acids, the average energy errors are significant, ranging from 2.45, 1.20, and 2.64 kcal/mol for pVal, pLeu, and pIle, respectively. For comparison, the average standard deviations of the QM results using MP2/6-31g** and MP2/cc-pVDZ methods were 0.7, 0.3, 0.2, and 0.2 kcal/mol for pAla, pVal, pLeu, and pIle, respectively. We also compare the curvature of the QM and MM (ff03) relative helix versus β-strand energy curves, recalculated per single residue (the relative energies for each N-residue peptide were divided by the number of residues, N) (Figs. 5A1–5D1). The per-residue-energy curves, which show the average stabilization of a residue in helix as a function of helix length, are with good accuracy parallel.[†] Therefore, the error of MM compared with QM results with regards to the helix stabilization per residue is almost constant and independent of helix length. It results mostly from the inaccuracy of the interaction energy between two residues forming H-bonds. Correcting this short-range interaction involving the formation of the first hydrogen bond, which should be achievable through reoptimization of the pairwise electrostatic interactions between residues involved in the hydrogen bond, will significantly reduce the error of entire MM stability curve and shift it toward the QM curve.[‡] Therefore, the gas-phase force fields based on the single-residue approximation (not containing correlation terms between residues) should be sufficient, after parameter refinement, to describe intrinsic helix versus extended state energies with good accuracy. We also compared the geometries of the peptides optimized using the HF/6-31g** method and the Amber ff03 force field (Table 4). The backbone dihedral angles of both helical and extended conformations are significantly shifted compared with the QM geometries, with larger differences for the helical conformation (Table 4, rows err). For comparison, the average standard deviations of backbone dihedral angles between structures obtained with 6-31g** and cc-pVDZ basis sets were less than 0.5° and 0.6° for helical and extended conformations, respectively, and the average errors in backbone dihedral angles of structures optimized at HF level compared with those optimized using the MP2[§] method were less than 2.2° and 1.6° for helical and extended structures, respectively. Therefore, differences between QM and MM structures are significantly larger than the inaccuracy of the QM cal-

[†]The single residues relative energies (Fig. 5) should be disregarded, because their geometries were optimized with constraints on backbone dihedral angles in helical conformation, as described in Methods.

[‡]We linearly transformed the MM full-stabilization energy curves, by adding of $N\Delta$ to each MM energy point, where N is number of residues and Δ is the average difference between MM and QM per-residue stabilization energy, which led to overlap of MM and QM stabilization energy curves with accuracy below 0.9, 0.4, and 0.2 kcal/mol for pLeu, pIle, and pVal, respectively (results not shown), which is close to the standard deviations of the QM results.

[§]To evaluate the accuracy of structures optimized at HF level, we compared the geometries of blocked tripeptides of Ala, Val, Leu, and Ile optimized at HF level with the geometries optimized at MP2 level.

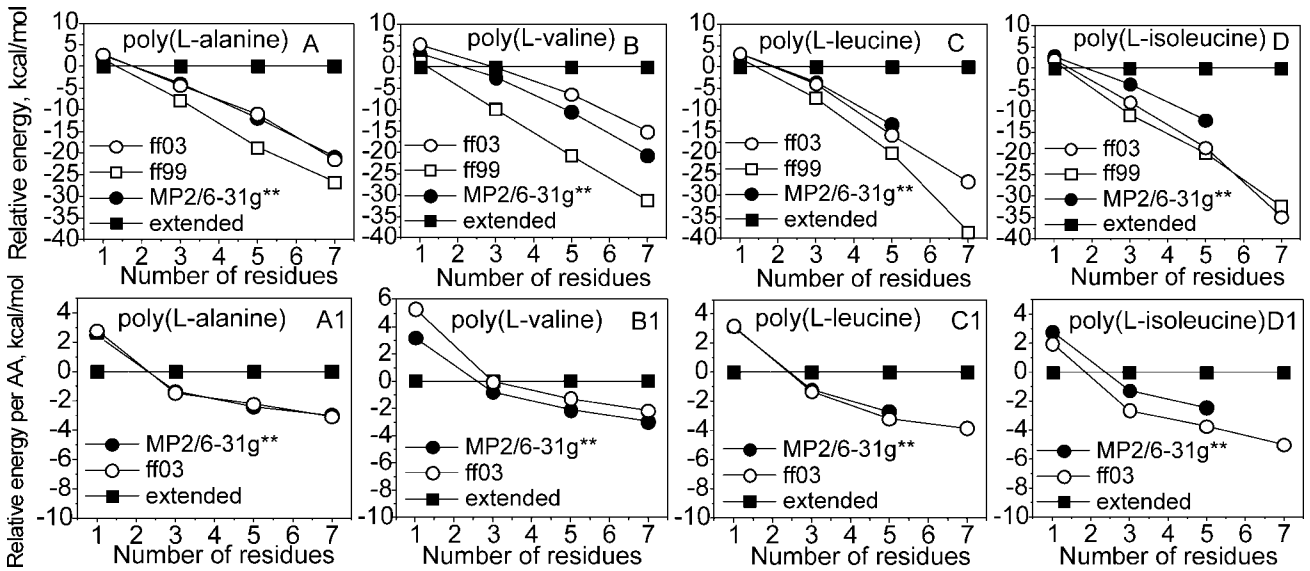

Figure 5. Relative energies of the helical versus extended structures for (A) pAla, (B) pVal, (C) pLeu, and (D) pIle, and the relative energies of the helical versus extended structures per residue (relative helix vs. extended state energy divided by the number of residues) for (A1) pAla, (B1) pVal, (C1) pLeu, and (D1) pIle (in kcal/mol), as a function of the number of residues, obtained using the MP/6-31g** method (filled circles), the ff03 force field (open circles), and the ff99 force field (open squares) (not shown in Figs. A1–D1). The energy of the extended structures (closed squares) for each type of polypeptide is taken as the reference.

culation. The largest distortions occur for pIle and pVal (Table 4). The shift of the backbone torsional geometry is reflected in the RMSD over C_α atoms between the QM and ff03 minimized structures, which can be larger than 1 Å even for five-residue long peptides (pIle) (Table 4, rows RMSD). Such a shift of local geometry may result in significant geometry discrepancies between the native and the predicted protein structures.

Discussion

Previous studies on helix versus β-strand stability at the QM level were limited to polyglycine[13] and pAla[18–20] peptides. The influence of amino acid type on helix stability has not been systematically analyzed before; only single amino acids replacements in glycine pentapeptides[21] were studied, using mixed DFT/semiempirical calculations. Here, we investigate the dependence of 3_{10}-helix versus extended conformation stability in the gas phase on amino acid chain length (1–7 residues) and type. We focus on alanine, leucine, isoleucine, and valine homopolypeptides because in native protein structures these amino acids occur with different frequencies in helices and β-stands; alanine and leucine more frequently populate helices, whereas valine and isoleucine more often occur in an extended conformation.[9] We apply a higher level of QM calculations than in previ-

Table 5. The Relative Energies of the Helical versus Extended Structures (kcal/mol).[a]

	Relative energy											
	pAla			pVal			pLeu			pIle		
N[a]	MP2	ff03	ff99	MP2	ff03	ff99	MP2	ff03	ff99	MP2	ff03	ff99
1	2.5	2.8	1.7	3.2	5.3	1.3	3.1	3.2	2.2	2.8	2.0	1.7
3	−4.0	−4.4	−7.7	−2.4	−0.1	−9.8	−3.6	−3.9	−7.2	−3.7	−7.9	−10.9
5	−11.9	−10.9	−18.6	−10.5	−6.3	−20.7	−13.4	−15.8	−20.1	−12.2	−18.7	−20.0
7	−20.9	−21.5	−26.7	−20.7	−15.1	−31.3	NA[b]	−26.7	−38.5	NA[b]	−34.9	−32.4

N, number of residues; pAla, poly(L-alanine); pVal, poly(L-valine); pLeu, poly(L-leucine); pIle, poly(L-isoleucine).

[a]Comparison of the results calculated at the MP2/6-31g** (MP2) level and using Amber ff99 (ff99) and ff03 (ff03) force fields.

[b]The system is too large to do the MP2/6-31g** calculations with our available computational resources.

ous studies and use fully optimized geometries (MP2/6-31g**//HF-6-31g** and MP2/cc-pVDZ//HF/cc-pVDZ). Further, we analyze the influence of excluded volume effects on relative helix-strand stability in all types of homopolypeptides. We use calculations for pseudopeptides to separate the hydrogen-bond effect from excluded volume effects in helix versus strand stabilization energy. Finally, we analyze the intrinsic correctness of force fields based on "independent residue" approximation, using ff03 and ff99 Amber potentials, by testing their ability to reproduce the QM helix versus extended state stabilities.

The helical conformation becomes lower in energy than the extended one for peptides longer than two residues and its stability cooperatively increases with chain length increase (in agreement with previous studies for polyalanine and polyglycine[13,18,22]) for all tested amino acids. There is a weak dependence of helix versus β-strand stability on amino acid identity that does not exceed 1.6 (2.4) kcal/mol at the MP2/6-31g** (MP2/cc-pVDZ) level compared with pAla. The stability of a pVal helix is less than that of pAla and is caused by interactions of the side chain with the backbone atoms that destabilize the helix, whereas the stability of the pLeu helix is greater than that of pAla due to side chain–backbone interactions that destabilize the extended state. In both cases, excluded volume effects between the side chain and the backbone lie within a single residue. That is, side chain–side chain interactions do not influence the relative helix versus extended state stability. The local propensities of valine (leucine) to destabilize (stabilize) the helical conformation with respect to alanine can be part of the reason for the statistically observed preferences of valine (leucine) to be in an extended (helical) state in native protein structures. However, the local effect of amino acid identity on helix versus extended state energy is small; rather, the observed stability of extended fragments in proteins must involve tertiary interactions and solvent effects. The relative stabilization of the helical versus extended conformation of pAla mainly depends on the hydrogen-bond energy, whereas for the other considered amino acids with large side chains, it results from both hydrogen bonding and local excluded volume effects that occur between the side chain and backbone atoms. Furthermore, the helix versus extended state stability depends strongly on the chain length and less on the amino acid identity. Therefore, some features of the local neighborhood of a residue, i.e., then number and backbone conformation of flanking residues, influence its backbone conformational preferences in opposition to the isolated pair hypothesis. However, since hydrogen bonding is mainly electrostatic in origin[24] and excluded volume interactions are limited to a single residue, the stabilization of the helix versus extended state could be accounted for by single residue MM potentials.

In practice, we find that the Amber ff03 force field reproduces the QM-based helix versus extended structure energy more accurately than the ff99 force field. The dependence of the relative 3_{10}-helix stabilization on chain length is correctly captured in the ff03 force field, i.e., the curvatures of the per-residue helix stabilization are the same for QM and ff03 results; however, the per-residue energy curves are shifted with respect to the QM curves by a constant error for pVal, pLeu, and pIle. This indicates that ff03 force field correctly describes the long-distance electrostatic stabilization of a residue by the rest of the helix, but short-range interactions between two residues involved in formation of hydrogen bond are inaccurate. Improving the accuracy of short-range interactions between residues forming hydrogen bond should be possible by reoptimization of parameters within a framework of "independent residue" potentials. Another argument supporting our conclusion that "independent residue" potentials are sufficient to describe helix (at least for 3_{10}-helix) versus extended state stability with good accuracy is the very well reproduced relative energy curve for pAla. Although our conclusions are based on the results for the Amber ff03 potential, they should be true for any potential based on the single residue approximation; differences in the accuracy of the helix-strand stability should reflect insufficient optimization of particular force-field parameters, rather than its intrinsic inadequacy. There are differences in geometries of the minimized helical and extended MM conformations compared to QM structures. Such local geometric inaccuracies in the force field may result in significant geometry discrepancies between the native and the predicted protein structures. Therefore, we conclude that while the gas phase component of MM force fields does not require the inclusion of correlation terms to correctly describe the local effect of the 3_{10}-helix versus extended structure stabilization, they do require parameter refinement to correctly describe the helix versus extended state energetics and geometry of major conformations. Our QM results provide reference data for refining the relative helix-strand energy and geometry in force fields, up to accuracy of about 1 kcal/mol and 2°–3° for energy and dihedral backbone angles, respectively. Such work is now in progress.

Recent studies[13,18,25] argue that a significant part of cooperative effect in helix stabilization with increasing length is due to a substantial electron density redistribution (polarization) induced by a large helix macrodipole (with a stronger effect for an α-helix than for a 3_{10}-helix). The magnitude of this effect would limit the accuracy of pairwise nonpolarizable potentials, because the induced cooperativity cannot be reproduced by nonpolarizable pairwise force field. Our results show that nonpolarizable pairwise force fields such as Amber ff03 can well reproduce the curvature of per-residue helix stabilization. This suggests that for the 3_{10}-helix the polarizable component of cooperativity in helix stabilization must be small, compared with the total stabilization energy (at least for the lengths up to 7 residues that we analyzed). This observation, however, is limited to the lengths of the helices studied in this work.

Concluding Remarks

Using QM methods (MP2/6-31g**//HF/6-31g** and MP2/cc-pVDZ//cc-pVDZ), we analyzed the origin of intrinsic (gas phase) 3_{10}-helix versus extended state stability in homopolypeptides (alanine, leucine, valine, and isoleucine) and its implications for the accuracy of "single residue" potentials. The helix versus extended state stability depends strongly on chain length and only weakly on amino acid identity. The helical conformation (3_{10}-helix) becomes lower in energy than the extended conformation for all tested peptides longer than two residues, and its stability increases with the increase of chain length. The relative stabilization of the helical versus extended conformation of pAla mainly depends on the hydrogen-bond energy, whereas for the other con-

sidered amino acids with large side chains, it results from both hydrogen bonding and local excluded volume effects that occur between the side chain and backbone atoms.

Next, we examined the ability of the "single residue" potentials (using Amber ff03 and ff99 force fields as an example) to reproduce the QM results for relative helix versus extended structure stabilities in the gas phase. The ff03 Amber potential better describes the helix versus extended state energy than ff99 and also reproduces the curvature of the relative helix-extended state energies. Therefore, the mathematical form of the Amber potential (and other potentials based on the single residue approximation) is sufficient to describe the local effect of helix versus extended structure stabilization in the gas phase. However, the energy curves are shifted and the backbone geometries differ compared with the QM results. Therefore, we conclude that, while the gas phase component of MM force fields does not require the inclusion of correlation terms to correctly describe the local effect of the 3_{10}-helix versus extended structure stabilization, they do require parameter refinement to correctly describe the helix versus extended state energetics and geometry of major conformations. Our QM results provide reference data for refining the relative helix-strand energy and geometry in force fields, up to accuracy of about 1 kcal/mol and 2°–3° for energy and dihedral backbone angles, respectively. Such work is now in progress.

Acknowledgments

We thank L. Wroblewska, Georgia Institute of Technology, for helpful discussions. This research was partly conducted using the resources of the National Science Foundation Teragrid Project and Terascale Computing System at the Pittsburgh Supercomputer Center.

References

1. Jang, S.; Kim, E.; Shin, S.; Pak, Y. J Am Chem Soc 2003, 125, 14841.
2. Lee, M. C.; Duan, Y. Proteins 2004, 55, 620.
3. Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A.; Cao, Y. X.; Murphy, R. B.; Zhou, R.; Halgren, T. A. J Comput Chem 2002, 23, 1515.
4. Flory, P. J. Statistical Mechanics of Chain Molecules;Wiley: New York, 1969.
5. Wang, J.; Cieplak, P.; Kollman, P. A. J Comput Chem 2000, 21, 1049.
6. Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. A. J Comput Chem 2003, 24, 1999.
7. Kaminski, G.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. J Phys Chem B 2001, 105, 6474.
8. Marquesee, S.; Robbins, V. H.; Baldwin, R. L. Proc Natl Acad Sci USA 1989, 86, 5286.
9. Betancourt, M.; Skolnick, J. J Mol Biol 2004, 342, 635.
10. Hovmoller, S.; Zhou, T.; Ohison, T. Acta Crystallogr Sect D 2002, 58, 768.
11. Hehre, W. J.; Ditchfield, R.; Pople, J. A. J Chem Phys 1972, 56, 2257.
12. Woon, D. E.; Dunning, T. H., Jr. J Chem Phys 1993, 98, 1358.
13. Wu, Y.-D.; Zhao, Y.-L. J Am Chem Soc 2001, 123, 5313.
14. Toniolo, C.; Benedetti, E. Trends Biochem Sci 1991, 16, 350.
15. Beachy, M. D.; Chasman, D.; Murphy, R. B.; Halgren, T. A.; Friesner, R. A. J Am Chem Soc 1997, 119, 5908.
16. Baginski, M.; Piela, L.; Skolnick, J. J Comput Chem 1993, 14, 471.
17. Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Wang, B.; Pearlman, D. A.; Crowley, M.; Brozell, S.; Tsui, V.; Gohlke, H.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister, C.; Caldwell, J. W.; Ross, W. S.; Kollman, P. A.AMBER8, University of California, San Francisco, 2004.
18. Wieczorek, R.; Dannenberg, J. J. J Am Chem Soc 2004, 126, 14198.
19. Park, C.; Goddard, W. A., III. J Phys Chem B 2000, 104, 7784.
20. Liu, H.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Hermans, J.; Yang, W. Proteins 2001, 44, 484.
21. Wieczorek, R.; Dannenberg, J. J. J Am Chem Soc 2003, 125, 14065.
22. Perczel, A.; Jakli, I.; Csizmadia, I. G. Chem Eur J 2003, 9, 5332.
23. Arnautova, Y. A.; Jagielska, A.; Scheraga, H. A. J Phys Chem B 2003, 107, 7143.
24. Jeziorski, B.; van Hemert, M. Mol Phys 1976, 31, 713.
25. Morozov, A.; Tsemekhman, K.; Baker, D. J Phys Chem B 2006, 110, 4503.