Importance of the Single Amino Acid Potential in Water for Secondary and Tertiary Structures of Proteins

Michio Iwaoka,* Daisuke Yosida, and Naoki Kimura

Department of Chemistry, School of Science, Tokai University, Kitakaname, Hiratsuka-shi, Kanagawa 259-1292, Japan

Received: April 8, 2006; In Final Form: May 25, 2006

The structure of a protein molecule is considered to be primarily determined by the inter-amino-acid nonbonded interactions, such as hydrogen bonds. However, the conformational space of the polypeptide chain should be simultaneously restricted by the intrinsic conformational preferences of the individual amino acids. We present here precise single amino acid potential (SAAP) surfaces for glycine (For-Gly-NH_2) and alanine (For-Ala-NH_2) in water ($\epsilon = 78.39$) and ether ($\epsilon = 4.335$), which were calculated at the HF/6-31+G(d,p) level applying the self-consistent isodensity polarizable continuum model (SCIPCM) reaction field with geometry optimization in the corresponding solvents. The obtained Ramachandran potential surfaces in water showed distinct potential wells in the α- and β-regions. The profiles were in almost perfect agreement with the Ramachandran plots of glycine and alanine residues in folded proteins, suggesting the Boltzmann distributions on the SAAP surfaces. Molecular simulations of polyalanines (For-Ala_n-NH_2; $n = 3-5$) by using the SAAP force field equipped with the SCIPCM potentials revealed that the polyalanines readily form 3_{10}-helical structures in water but not in vacuo. In ether (hydrophobic environments), the helical structures were relatively stable, but the most stable structure was assigned to a different one. These results indicated that the intrinsic conformational preferences of the individual amino acids (i.e., the SAAPs) in water are of significant importance not only for describing conformations of a polypeptide chain in the random coil state but also for understanding the folding to the secondary and tertiary structures.

Introduction

The structure of a protein molecule is controlled by the conformational space defined not only by the inter-amino-acid interactions but also by the intrinsic conformational preferences of the individual amino acids.[1] Physicochemical properties of the former nonbonded interactions, such as hydrogen bonds, electrostatic interactions, and van der Waals forces, have been elaborately studied for decades.[2] However, our understanding about the latter factor may be a little behind because the precise profile of a single amino acid potential (SAAP)[3] surface is difficult to be elucidated by experiment and even by theoretical calculation, especially in water and hydrophobic environments.

Conformations of amino acids are usually described by use of the main-chain ϕ and ψ dihedral angles (i.e., Ramachandran plots):[4] It was originally demonstrated that the two-dimensional conformational space can be divided into the sterically allowed, which covers most structures of amino acid residues in folded proteins, and not-allowed areas on the basis of molecular modeling. In the meantime, conformational preferences of individual amino acids were analyzed based on the frequencies found in the secondary structures of proteins.[5] For example, glycine (Gly) is more favorable in turns than in α-helices and β-strands. Alanine (Ala) is more favorable in α-helices than in β-strands and turns. Experimental assays using model polypeptides verified the tendencies.[6] However, the preferences thus obtained should be reflected by the mean force derived from several factors, which consist of the intrinsic conformational preferences of the individual amino acids and the interactions with surrounding amino acid residues and solvent molecules.

In the random coil (or non-native unfolded) state of a polypeptide chain, each amino acid residue would simply follow the intrinsic conformational preference in water because the inter-amino-acid interactions are small. Dobson and co-workers[7] proposed that the statistical structure of an amino acid in the random coil state is expressed by use of the distribution of the Ramachandran plots. This approach, in combination with experimentally available NMR coupling constants, successfully described the conformations of polypeptides in the random coil and partially folded states in aqueous solution,[8] demonstrating that Ramachandran plots have significance not only for protein structures in the native folded state but also for those in the unfolded state. Recent spectroscopic investigations[9] indeed suggested the existence of a significant amount of secondary structures in the denatured state. Nevertheless, the accurate profiles of the SAAP surfaces were still unclear according only to these experimental approaches.

Quantum chemical calculation is another straightforward approach to the intrinsic conformational preferences of amino acids. According to ab initio calculation results, glycine and alanine residues had strong preferences in vacuo to adopt so-called C5 and C7 conformations,[10] which have an intramolecular hydrogen bond between the two successive peptide linkages. Thus, the SAAPs in vacuo should be significantly different from those deducible from the Ramachandran plots. Meanwhile, the SAAPs in water,[11−14] which were obtained by applying the self-consistent reaction field (SCRF) method with various polarizable continuum models (PCM) of a solvent,[15−17] showed distinct

* Author to whom correspondence should be addressed. Phone: +81-463-58-1211. Fax: +81-463-50-2094. E-mail: miwaoka@tokai.ac.jp.

10.1021/jp062196g CCC: $33.50
Published on Web 06/29/2006

potential wells in the α-helix and β-sheet regions. The result suggested the importance of polar solvent effects on the conformational preferences of amino acids and also the similarity between the SAAPs in water and the distributions of the Ramachandran plots, reasonably supporting the aforementioned Dobson and co-workers' proposal[7] as well as Sakae and Okamoto's recent approach[18] to optimization of protein force-field parameters with the Ramachandran plots. These findings have also encouraged one of the authors to develop the new all-atom protein force field (the SAAP force field),[3] in which the total potential energy (E^{TOTAL}) was formulated by use of the single amino acid potential term (E^{SAAP}), the inter-amino-acid interaction term (E^{INTER}), and the other term (E^{OTHERS}).

The PCM methods[19] implemented in the Gaussian03 program package[20] have been greatly improved in the definition of the solute cavity and the treatment of the solute electrons lying outside the cavity. With reasonable accuracy and computation cost, the methods are now widely applied to conformational analysis for various amino acids in water.[21] Perczel and co-workers[13] calculated the fully optimized Ramachandran potential surface of For-Ala-NH_2 in water at the B3LYP/PCM/6-31+G(d) level. Wang and Duan[14] obtained similar potential surfaces of Ace-Ala-NHMe in ether and water at the MP2/cc-pVTZ//MP2/6-31G(d,p) level. However, the whole potential surfaces for glycine in ether and water have not yet been reported by applying the improved PCM methods.[19] The self-consistent isodensity polarizable continuum model (SCIPCM)[17] is one variation of the PCM methods in which the solute cavity is defined by the electron isodensity surface of a solute molecule and the free energy including the solute−solvent interaction is minimized in the SCF process. The method has been successfully employed for calculation of the vibrational spectra of formaldehyde in acetonitrile,[22] the conformational space of glucuronic acid in water,[23] etc.[24] However, it has not yet been applied to the conformational analysis of peptide molecules. Accordingly, it is still a continued interest of research to elucidate accurate profiles of the SAAP surfaces in water and hydrophobic environments for various amino acids.

In this paper, we present precise potential surfaces for glycine (For-Gly-NH_2; **1**) and alanine (For-Ala-NH_2; **2**) in water ($\epsilon = 78.39$) and in ether ($\epsilon = 4.335$, hydrophobic environments) calculated at the HF/6-31G+(d,p) level applying the SCIPCM method with geometry optimization in the corresponding solvents. The obtained two-dimensional SAAP surfaces were compared with those obtained by the PCM and isodensity PCM (IPCM) methods and also with the distributions of the Ramachandran plots of glycine and alanine residues in proteins. The SCIPCM potentials were further utilized for molecular simulations of oligopeptides (For-Ala_n-NH_2; $n = 3-5$) in water and ether in the framework of the SAAP force field.[3] Implications of the SAAPs in water for the stability and folding of secondary and tertiary structures of proteins are discussed.

Methods

Ab Initio Calculation. All ab initio quantum chemical calculations were carried out by using the Gaussian03 program package[20] installed on a high-performance HPS SCC System computer. Single amino acid potential functions were calculated as follows by use of the sets of structures[12] previously obtained at the HF/6-31G(d) level for For-Gly-NH_2 (**1**) and For-Ala-NH_2 (**2**) with resolutions of 15° for the main-chain dihedral angles (ϕ and ψ).

To calculate the SAAP functions in water, we applied three different PCM methods for **1** and **2**. In the general PCM[19] and SCIPCM[17] reaction field calculations, each structure was reoptimized in water ($\epsilon = 78.39$) at the HF/6-31+G(d,p) level, fixing the ϕ and ψ dihedral angles at the appropriate values. However, since geometry optimization was not feasible in the IPCM[17] reaction field calculation, each structure was first reoptimized at the HF/6-31+G(d,p) level in vacuo, fixing the two dihedral angles, and then the single-point electronic energy was calculated in water by applying the IPCM method at the HF/6-31+G(d,p) level. The two-dimensional SAAP functions of **1** and **2** in ether ($\epsilon = 4.335$) were also calculated by applying the SCIPCM method in the same manner as in water.

In the SCIPCM calculation, a value of 0.0004 au (default) was employed as the electron isodensity to define the surface of the solute cavity, which was integrated by 2048 surface points (64×32 for ϕ and θ polar angles upon a single origin). The number of the surface points was chosen after extensive examination to obtain reasonable potential surfaces within a limited computation time. Having obtained suitable parameters for the ϕ and θ polar angles in the SCIPCM calculations, we applied similar calculation conditions (i.e., 2048 points on the cavity surface at the default isodensity of 0.001 au) in the IPCM calculation in water. Default parameters were employed without modification in the simple PCM calculation in water. Two-dimensional Ramachandran potential surfaces were drawn by using the Gnuplot program, in which the lattice-point data (24×24 for ϕ and ψ main-chain dihedral angles) were interpolated with spline functions.

The potential energy minima were subsequently sought on the SCIPCM potential surfaces obtained in water and ether as well as on the potential surfaces obtained in vacuo at the HF/6-31+G(d,p) level. The resulting stable structures (i.e., the partially optimized local energy minima) were fully optimized with a TIGHT option at the same calculation level without restrictions of the two dihedral angles. Frequency calculation was finally performed to ensure the characteristics of the obtained structures as potential minima. The nomenclature for the potential minima that is used throughout the present work is quoted from the previous paper,[12] but that used by Perczel's group[13] is also indicated in parentheses.

Protein Data Bank Analysis. According to the method described previously,[12] the values of main-chain dihedral angles (ϕ and ψ) were calculated for all nonterminal glycine and alanine residues involved in 1085 heterogeneous protein structures (resolution ≤ 3.0 Å) that were retrieved from Protein Data Bank (PDB).[25] The 10 836 glycine and 11 871 alanine residues were utilized for drawing the Ramachandran plots. The relative conformational energies of individual glycine or alanine residues in folded proteins were estimated by the use of the computed Ramachandran potential surfaces of **1** or **2**, respectively, and were statistically analyzed based on the densities (i.e., probabilities) of the plots in the specific energy region on the potential surface. To determine the area with an arbitrary conformational energy range, the potential surface was divided into 360×360 grids, for each of which the conformational energy was assigned by the linear interpolation method using the surrounding lattice points.

Simulation Using the SAAP Force Field. A simple Metropolis Monte Carlo program previously coded for the SAAP force field[3] was employed for molecular simulations of For-Ala_n-NH_2 ($n = 3-5$) in water, ether (hydrophobic environments), and vacuo with equipment of the newly developed SCIPCM potential functions for alanine. In each step, the ϕ and ψ dihedral angles of all alanine residues were randomly altered with maximum changes of 2−5°. Through the use of new sets

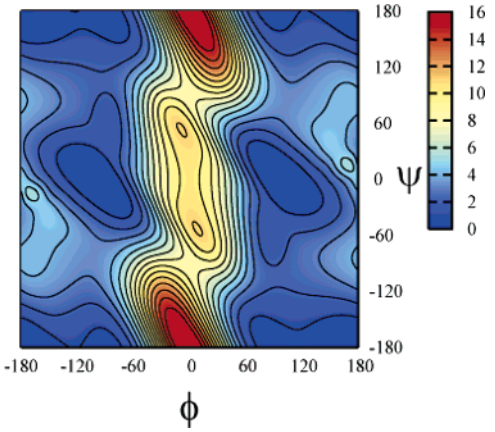

Figure 1. The $\phi-\psi$ potential surface of For-Gly-NH_2 (**1**) in water ($\epsilon = 78.39$) by applying the SCIPCM method at the HF/6-31+G(d,p) level. The structure was relaxed in water at the same calculation level. Contours are drawn with an interval of 1 kcal/mol.

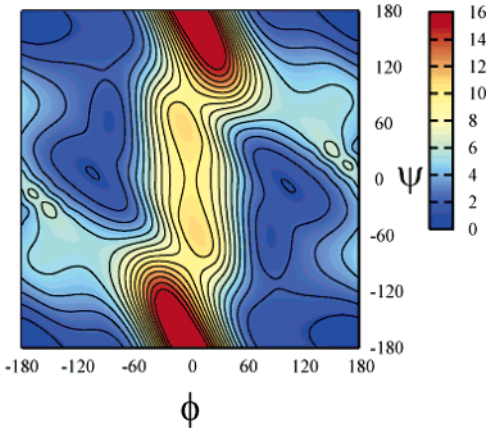

Figure 2. The $\phi-\psi$ potential surface of For-Gly-NH_2 (**1**) in ether ($\epsilon = 4.335$) by applying the SCIPCM method at the HF/6-31+G(d,p) level. The structure was relaxed in ether at the same calculation level. Contours are drawn with an interval of 1 kcal/mol.

of ϕ and ψ dihedral angles, the test structure was constructed, and the total energy E^{TOTAL}, which consists of single amino acid potential energies (E^{SAAP}) and interresidue nonbonded interactions ($E^{INTER} \approx E^{ES'} + E^{LJ'}$), was calculated and compared with the energy of the original structure. The high-order interaction term (E^{OTHERS}) was neglected in a current version of the SAAP force field.[3] The potential energy of each amino acid residue $E(\phi,\psi)$, the Cartesian coordinates, and the atomic charges were linearly interpolated at any point on the $\phi-\psi$ potential surface by the use of the corresponding values of the surrounding lattice points.

In the simulations with the SAAP force field in water and ether, fixed values of the dielectric constants ($\epsilon = 78.39$ and 4.335, respectively) were used for calculation of the electrostatic interaction term ($E^{ES'}$), although they should slightly change depending on the surrounding environment and the temperature. The dielectric constant was set to 1.0 for the simulation in vacuo. The Monte Carlo simulation was performed at 310 K for 20 000 steps to search the stable structures. The resulting energy minima were then optimized at 0 K for 20 000 steps to obtain the local energy minimum structures. All possible combinations of the stable conformers of **2** calculated in the appropriate solvent were applied for the initial structures. For example, since seven energy minima were characterized for alanine in water, 343 ($= 7^3$) initial structures were employed in the simulation of For-Ala_3-NH_2 in water. According to this systematic process, the stable structures of polyalanines were extensively searched in the whole conformational space.

Results

SCIPCM Potentials of For-Gly-NH_2 (1). Two-dimensional potential surfaces of **1** obtained by the SCIPCM method in water ($\epsilon = 78.39$) and in ether ($\epsilon = 4.335$) are shown in Figures 1 and 2, respectively. The low-energy regions are indicated in blue, and the high-energy regions are indicated in red. The $\phi-\psi$ potential surfaces of **1** are center-symmetric because the C_α atom of glycine is not chiral.

On the SCIPCM potential surface in water (Figure 1), there are two wide potential wells around $(\phi, \psi) = (-180°, -180°)$ and $(-105°, 0°)$. The former, which corresponds to the extended C5 (or β_L[13]) region with a five-membered hydrogen-bonded ring, extends to the PII (or ϵ_L[13]) region for a poly-Pro II helix [$(\phi, \psi) = (-90°, -180°)$]. The latter, which corresponds to the B region (a bridge region), spreads to the α_R (or α_L[13]) region for a right-handed α helix [$(\phi, \psi) = (-75°, -30°)$]. The two wells, which are energetically almost equal, are connected to each other by a wide shoal passing the so-called C7 (or γ_L[13]) region [$(\phi, \psi) = (-90°, 60°)$] with a seven-membered hydrogen-bonded ring.

Similarly, there are two wide wells on the SCIPCM potential surface of **1** in ether (Figure 2). However, the two wells are no longer equivalent energetically. Changes of the potential surface from water to ether can be rationalized by the decrease of the polar solvent effect, leading to stabilization of the conformer having a hydrogen bond: The regions corresponding to C5 and C7 conformers are relatively stabilized on the potential surface in ether.

TABLE 1: Characteristics of Potential Energy Minima Obtained for For-Gly-NH_2 (1)[a]

symbols[b]	ϕ (deg)	ψ (deg)	ΔE (kcal/mol)[c]	ν_{min} (cm^{-1})[d]
In Water (HF/SCIPCM/6-31+G(d,p))				
B (α_R)[e]	−94.4	−3.8	0.00	38
C5	−180.0	−180.0	0.13	51
PII	−82.7	171.4	0.69	36
C7	−87.2	56.1	1.40	38
In Ether (HF/SCIPCM/6-31+G(d,p))				
C5	−180.0	−180.0	0.00	52
B	−104.5	7.2	0.97	39
C7	−86.6	61.7	1.05	44
In Vacuo (HF/6-31+G(d,p))				
C5	−180.0	−180.0	0.00	53
C7	−85.4	66.7	0.60	45

[a] The geometry was fully optimized at the corresponding calculation level. [b] Symbols were quoted from ref 12. [c] Relative electronic energies without calibration for the zero-point vibrational energies. [d] The lowest wavenumbers obtained from the vibrational frequency calculation. [e] The potential minimum was located between the B and α_R regions.

Table 1 lists the ϕ and ψ dihedral angles and the relative energies (ΔE) for the local energy minima of **1** found in water and ether along with those in vacuo for comparison. All structures were fully optimized without restrictions of the dihedral angles and were characterized by vibrational frequency calculation. In water, four local energy minima were identified. The most stable one was assigned to a B conformer, although the location on the potential surface was slightly shifted toward the α_R region. The other minima were assigned to C5, PII, and C7 conformers with $\Delta E = 0.13$, 0.69, and 1.40 kcal/mol, respectively. On the PCM and IPCM potentials of **1** in water reported previously,[12] seven and eight stable conformers had been predicted. However, most of the high-energy conformers

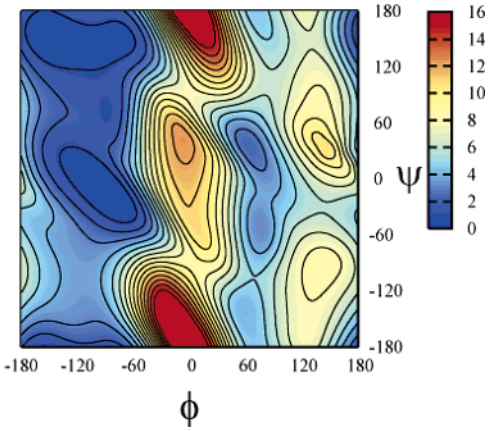

Figure 3. The $\phi-\psi$ potential surface of For-Ala-NH_2 (**2**) in water ($\epsilon = 78.39$) by applying the SCIPCM method at the HF/6-31+G(d,p) level. The structure was relaxed in water at the same calculation level. Contours are drawn with an interval of 1 kcal/mol.

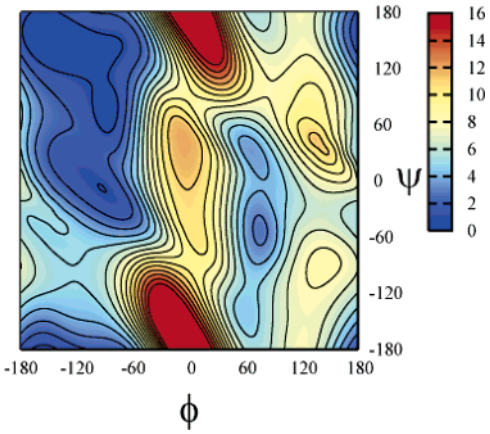

Figure 4. The $\phi-\psi$ potential surface of For-Ala-NH_2 (**2**) in ether ($\epsilon = 4.335$) by applying the SCIPCM method at the HF/6-31+G(d,p) level. The structure was relaxed in ether at the same calculation level. Contours are drawn with an interval of 1 kcal/mol.

disappeared on the SCIPCM potential (Figure 1) probably because the structures have been optimized in water. Importance of the geometry optimization process in the corresponding solvent has been pointed out by Perczel.[13]

However, three stable structures (i.e., C5, B, and C7 conformers) were identified on the SCIPCM potential of **1** in ether (Figure 2). When the solvent was changed from water to ether, the PII conformer became a transient structure, and the C5 and C7 conformers, which have an intramolecular hydrogen bond, were stabilized by ~1 kcal/mol with respect to the B conformer, which does not have a hydrogen bond. On the previous IPCM potential in ether,[12] two more conformers had been found in the high-energy regions, but they were not present on the optimized SCIPCM potential in ether. In vacuo at the HF/6-31+G(d,p) level, C5 and C7 conformers were characterized as the potential minima, suggesting that only the structures with an intramolecular hydrogen bond are dominant in vacuo for **1**. The result was consistent with the previous calculation results for For-Gly-NH_2 (**1**) and Ace-Gly-NHMe in vacuo at various levels of theory.[10]

SCIPCM Potentials of For-Ala-NH_2 (2). Figures 3 and 4 show two-dimensional potential surfaces of **2** obtained by the SCIPCM method in water ($\epsilon = 78.39$) and in ether ($\epsilon = 4.335$), respectively. Since the C_α atom of alanine is chiral in contrast to glycine, the $\phi-\psi$ potential surfaces of **2** become unsymmetrical with respect to the origin.

TABLE 2: Characteristics of Potential Energy Minima Obtained for For-Ala-NH_2 (2)[a]

symbols[b]	ϕ (deg)	ψ (deg)	ΔE (kcal/mol)[c]	ν_{min} (cm^{-1})[d]
In Water (HF/SCIPCM/6-31+G(d,p))				
α_R	−83.2	−19.0	0.00	38
C5	−151.4	155.0	0.17	45
PII	−75.3	149.6	0.60	45
$C7_{eq}$	−89.3	74.1	1.04	45
α_L	64.1	35.1	2.39	57
$C7_{ax}$	75.9	−50.4	3.20	51
α_D	59.6	−143.8	4.07	49
In Ether (HF/SCIPCM/6-31+G(d,p))				
C5	−152.7	156.2	0.00	46
$C7_{eq}$	−88.3	75.5	0.55	49
B (α_R)[e]	−92.8	−8.7	1.04	27
$C7_{ax}$	75.7	−52.9	2.79	52
α_L	67.1	30.5	3.53	50
α_D	60.0	−140.3	4.56	46
α¢	−157.0	−51.7	4.90	23
In Vacuo (HF/6-31+G(d,p))				
$C7_{eq}$	−86.1	76.6	0.00	52
C5	−155.7	160.3	0.15	49
B	−111.2	13.3	2.28	20
$C7_{ax}$	75.2	−55.0	2.53	48
α_L	70.0	24.2	4.81	37
α¢	−165.8	−41.9	5.68	49

[a] The geometry was fully optimized at the corresponding calculation level. [b] Symbols were quoted from ref 12. [c] Relative electronic energies without calibration for the zero-point vibrational energies. [d] The lowest wavenumbers obtained from the vibrational frequency calculation. [e] The energy minimum was located between the B and α_R regions.

There are four potential wells on the SCIPCM surface of **2** in water (Figure 3). The two wide wells are located in the left-hand side ($-180° \le \phi < 0°$) around the C5 (or β_L[13]) to PII (or ϵ_L[13]) region [(ϕ, ψ) = (−150° to −75°, 150°)] and the right-handed helical α_R (or α_L[13]) region [(ϕ, ψ) = (−90°, −15°)]. Similar to the potential surface for glycine (Figure 1), these wells are energetically almost equal and are connected by a wide shoal through the $C7_{eq}$ (or γ_L[13]) region [(ϕ, ψ) = (−90°, 75°)]. The other two wells, which are located on the right-hand side ($0° \le \phi < 180°$) around the left-handed helical α_L (or α_D[13]) to $C7_{ax}$ (or γ_D[13]) region [(ϕ, ψ) = (75°, 30° to −45°)] and the α_D (or ϵ_D[13]) region [(ϕ, ψ) = (60°, −150°)], are rather narrow and shallow. However, the wells around the α_R and α_D regions are a little destabilized in ether (Figure 4) and tend to merge to those around the $C7_{eq}$ and $C7_{ax}$ regions, respectively. The observation can be ascribed to the loss of the polar solvent effect as observed for glycine model **1**: The C5, $C7_{eq}$, and $C7_{ax}$ regions of **2**, which allow formation of an intramolecular hydrogen bond, become relatively stable in ether.

The potential energy minima of **2** characterized in water and ether are listed in Table 2 along with those characterized in vacuo. Seven conformers were located in water by full geometry optimization applying the SCIPCM method at the HF/6-31+G(d,p) level. The ϕ and ψ dihedral angles and the relative energies (ΔE) of the four stable conformers (i.e., α_R, C5, PII, and $C7_{eq}$), which were found in the left-hand side of Figure 3 ($-180° \le \phi < 0°$), were similar to those for glycine model **1** shown in Table 1, although the most stable B conformer of **1** [(ϕ, ψ) = (−94.4°, −3.8°)] has migrated to the α_R region [(ϕ, ψ) = (−83.2°, −19.0°)] for alanine model **2**. The other three minima (i.e., α_L, $C7_{ax}$, and α_D) were found in the right-hand side ($0° \le \phi < 180°$). These minima were relatively unstable ($\Delta E > 2$ kcal/mol) due to the steric repulsion from the side-chain CH_3 group. On the similar IPCM potential of **2** by use of the structures optimized in vacuo, 17 energy minima had been

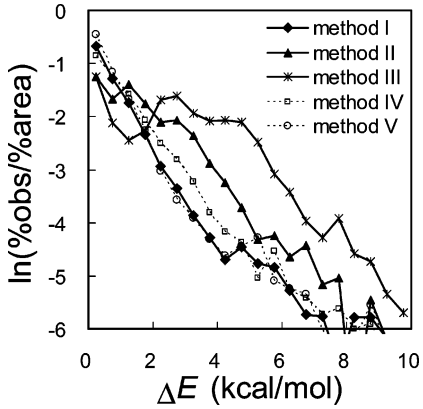

Figure 5. Density of the Ramachandran plots (% obs/% area) as a function of the relative energy (ΔE) for glycine. The calculation methods were HF/SCIPCM/6-31+G(d,p) in water (method I), HF/SCIPCM/6-31+G(d,p) in ether (method II), HF/6-31+G(d,p) in vacuo (method III), HF/IPCM/6-31+G(d,p)//HF/6-31+G(d,p) in water (method IV), and HF/PCM/6-31+G(d,p) in water (method V). The value of % obs/% area was calculated from the percentage of the Ramachandran plots observed in an arbitrary relative energy (ΔE) range on the SAAP surface divided by the percentage of the area.

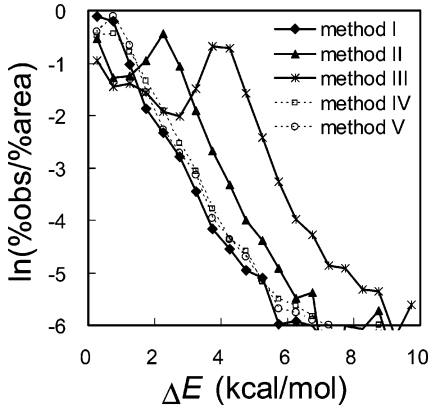

Figure 6. Density of the Ramachandran plots (% obs/% area) as a function of the relative energy (ΔE) for alanine. The calculation methods were HF/SCIPCM/6-31+G(d,p) in water (method I), HF/SCIPCM/6-31+G(d,p) in ether (method II), HF/6-31+G(d,p) in vacuo (method III), HF/IPCM/6-31+G(d,p)//HF/6-31+G(d,p) in water (method IV), and HF/PCM/6-31+G(d,p) in water (method V). The value of % obs/% area was calculated from the percentage of the Ramachandran plots observed in an arbitrary relative energy (ΔE) range on the SAAP surface divided by the percentage of the area.

predicted.[12] However, a recent elaborate PCM study at the HF/6-31+G(d) level[13] demonstrated that the number of the energy minima should decrease to 8 by applying geometry optimization in water. Accordingly, the number of the energy minima decreased to 7 on the SCIPCM potentials of **2** by applying geometry optimization in water (Figure 3 and Table 2).

On the SCIPCM potential of **2** in ether (Figure 4), 7 energy minima were identified (Table 2). When the solvent was changed from water to ether, the PII conformer disappeared, the α_R conformer migrated to the B region, and the α' (or δ_D[13]) conformer [(ϕ, ψ) = (−157.0°, −51.7°)] became a new local energy minimum. The C5, $C7_{eq}$, and $C7_{ax}$ conformers with a hydrogen bond were relatively stabilized in ether by ∼1 kcal/mol with respect to the B (α_R), α_L, and α_D conformers.

In vacuo, six stable conformers were found at the HF/6-31+G(d,p) level of theory. The α_D conformer, which was a distinct energy minimum in water and ether, was merged to the $C7_{ax}$ conformer in vacuo. The three conformers with a hydrogen bond (i.e., C5, $C7_{eq}$, and $C7_{ax}$) were further stabilized in vacuo with respect to the other conformers. Similar potential minima have been reported in the literature for For-Ala-NH_2 (**2**) and Ace-Ala-NHMe at various levels of theory.[10]

Comparison with the Ramachandran Plots. Figure 5 shows the correlation plots between the relative energy (ΔE) on the $\phi-\psi$ potential surface of **1** and the natural logarithm of the density (or probability) of Ramachandran plots (i.e., ln(% obs/% area)) for 10 836 glycine residues found in 1085 heterogeneous protein structures. Similarly, the correlations between ΔE on the $\phi-\psi$ potential surfaces of **2** and ln(%obs/%area) for 11 871 alanine residues are shown in Figure 6. The SAAPs obtained in five different calculation methods have been applied in the figures: the optimized SCIPCM potential surfaces in water (Figures 1 and 3; method I) and in ether (Figures 2 and 4; method II), the potential surface in vacuo optimized at the HF/6-31+G(d,p) level (method III), the single-point IPCM potential surface in water at the HF/6-31+G(d,p) level (method IV), and the optimized PCM potential surface in water at the HF/6-31+G(d,p) level (method V). While similar IPCM and PCM potential surfaces of **1** and **2** had previously been utilized in this type of analysis,[12] we employed here the corresponding potentials recalculated by applying more appropriate calculation methods (see the Methods section for details).

For both glycine (Figure 5) and alanine (Figure 6), almost linear correlations were obtained for the SAAP functions in water (methods I, IV, and V), indicating that the lower the potential energies of the single amino acids in water, the more structures are observed in proteins. Thus, it is suggested that the structures of glycine and alanine in proteins *statistically* follow Boltzmann (or random) distributions upon the SAAP surfaces in water. The temperatures of the presumable Boltzmann distributions for glycine and alanine in proteins were both estimated to be about 500 K based on the SCIPCM potentials of **1** and **2** in water. However, the correlation curves peaked significantly for the SAAP functions in ether (method II) and in vacuo (method III).

Molecular Simulations of For-Ala_n-NH_2 by Using the SAAP Force Field. All possible stable structures were exhaustively searched for For-Ala_n-NH_2 (n = 3−5) in water, ether, and vacuo by use of the SAAP force field,[3] in which the improved SCIPCM potential functions were employed. By simple Metropolis Monte Carlo simulation at 310 K, a large number of stable structures were characterized. Table 3 summarizes the relative energies of the obtained helical and extended structures that have more than three continuous α_R or C5 conformations along the sequence.

For the alanine trimer (n = 3) in water, the $\alpha_R\alpha_R\alpha_R$ conformation, which corresponds to a 3_{10}-helix structure with two intramolecular hydrogen bonds, was the most stable, while the C5C5C5 conformation, which corresponds to an extended β-strand structure, was slightly disfavored energetically (ΔE = 1.8 kcal/mol). However, the most stable structure was assigned to $C7_{eq}\alpha_R$C5 in ether, and the $\alpha_R\alpha_R\alpha_R$ and C5C5C5 conformations were found to have relative energies of 0.5 and 0.4 kcal/mol, respectively. Both helical and strand structures were, however, not stable any more in vacuo probably due to strong electrostatic interactions.

Similar simulation results were observed for the alanine tetramer (n = 4) and pentamer (n = 5). The 3_{10}-helical and β-strand structures were relatively stable in water and ether, but they were no longer stable in vacuo. It is also notable that the 3_{10}-helical structures were remarkably stable in water for all polyalanines (ΔE = 0.0−1.5 kcal/mol). The result is in complete agreement with the recent experimental observations that polyalanines form 3_{10}-helical structures in water,[26] sup-

TABLE 3: Relative Energies of Helical and Extended Structures Obtained for For-Ala$_n$-NH$_2$ by the Molecular Simulations at 310 K Applying the SAAP Force Field[a]

		ΔE (kcal/mol)[b]		
structures	conformations	in water ($\epsilon = 78.39$)	in ether ($\epsilon = 4.335$)	in vacuo ($\epsilon = 1.0$)
		$n = 3$[c]		
3_{10}-helix	$\alpha_R\alpha_R\alpha_R$	0.0 (2)	0.5 (2)	not stable
β-strand	C5C5C5	1.8 (0)	0.4 (0)	not stable
		$n = 4$[d]		
3_{10}-helix	$\alpha_R\alpha_R\alpha_R\alpha_R$	0.0 (1)	0.5 (3)	8.6 (2)
	$\alpha_R\alpha_R\alpha_R$Y	0.1 (0) (Y = C5)	0.5 (3) (Y = B)	not stable
	Y$\alpha_R\alpha_R\alpha_R$	0.7 (1) (Y = C5)	0.7 (2) (Y = C7$_{eq}$)	7.3 (2) (Y = C7$_{eq}$)
β-strand	C5C5C5C5	not stable	2.0 (0)	not stable
	C5C5C5Y	3.2 (0) (Y = α_R)	2.5 (0) (Y = C7$_{eq}$)	not stable
	YC5C5C5	2.8 (0) (Y = α_R)	2.3 (0) (Y = α_R)	not stable
		$n = 5$[e]		
3_{10}-helix	$\alpha_R\alpha_R\alpha_R\alpha_R\alpha_R$	0.2 (1)	1.1 (4)	not stable
	$\alpha_R\alpha_R\alpha_R\alpha_R$Y	0.2 (0) (Y = C5)	1.1 (4) (Y = B)	not stable
	Y$\alpha_R\alpha_R\alpha_R\alpha_R$	0.9 (1) (Y = C5)	1.3 (3) (Y = C7$_{eq}$)	13.2 (3) (Y = α_D)
	$\alpha_R\alpha_R\alpha_R$YZ	0.0 (0) (Y = B, Z = C5)	1.2 (1) (Y = B, Z = B)	not stable
	Y$\alpha_R\alpha_R\alpha_R$Z	1.1 (1) (Y = C5, Z = B)	1.3 (3) (Y = C7$_{eq}$, Z = B)	13.5 (3) (Y = α_D, Z = B)
	YZ$\alpha_R\alpha_R\alpha_R$	1.5 (1) (Y = α_R, Z = PII)	2.3 (2) (Y = C5, Z = α_D)	12.6 (3) (Y = β1, Z = α_D)
β-strand	C5C5C5C5C5	not stable	4.2 (0)	not stable
	C5C5C5C5Y	4.8 (0) (Y = α_R)	4.7 (0) (Y = C7$_{eq}$)	not stable
	YC5C5C5C5	4.6 (0) (Y = α_R)	4.5 (0) (Y = α_R)	not stable
	C5C5C5YZ	3.9 (1) (Y = α_R, Z = α_R)	3.9 (0) (Y = α_R, Z = C5)	not stable
	YC5C5C5Z	4.3 (0) (Y = α_R, Z = α_R)	5.0 (0) (Y = α_R, Z = C7$_{eq}$)	not stable
	YZC5C5C5	3.3 (0) (Y = α_R, Z = α_R)	3.6 (0) (Y = C7$_{eq}$, Z = α_R)	not stable

[a] Only the most stable structures with more than three continuous α_R or C5 conformations along the sequence are listed. [b] Relative potential energies with respect to that of the global energy minimum structure. The number of hydrogen bonds is indicated in parentheses. [c] The global energy minimum structures were $\alpha_R\alpha_R\alpha_R$ in water, C7$_{eq}\alpha_R$C5 in ether, and $\alpha_D\alpha_R\alpha_R$ in vacuo. [d] The global energy minimum structures were $\alpha_R\alpha_R$BC5 in water, C7$_{eq}$C7$_{ax}\alpha_R$C5 in ether, and β1$\alpha_D\alpha_R$C7$_{eq}$ in vacuo. [e] The global energy minimum structures were $\alpha_R\alpha_R\alpha_R$BC5 in water, C7$_{eq}\alpha_D\alpha_R$BC7$_{ax}$ in ether, and C7$_{eq}\alpha_D\alpha_R$BC7$_{ax}$ in vacuo.

porting the reliability of the SAAP force-field[3] approach as well as the SCIPCM potential surfaces obtained here. However, the β-strand structures were not so stable in water and ether (i.e., hydrophobic environments) as the helical structures.

Discussion

Accuracy of the SAAP Functions. To obtain SAAP functions for glycine and alanine in water, we applied three types of continuous dielectric models of a solvent: PCM, IPCM, and SCIPCM. These three methods would have their own advantageous features based on the definition of a solute cavity and the calculation of the solute−solvent interaction energy. In the previous PCM method,[16] the solute cavity was defined by the superimposed spherical surfaces of individual atoms or groups of a solute molecule. The current PCM method,[19] however, utilizes the solvent-excluding surface, which is defined by a probe sphere rolling on solute atoms, to smooth the cavity. However, the cavity is defined by the electron isodensity surface of a solute molecule in IPCM and SCIPCM.[17] The three potential surfaces obtained for glycine in water were very similar to each other irrespective of the types of the PCM reaction fields. Similarly, the potential surfaces for alanine in water were similar to each other. These similarities were also deducible from the linear correlation plots shown in Figures 5 and 6. Thus, the accuracy of the SAAP surfaces does not change significantly within the types of the continuous dielectric models of a solvent employed in this work.

However, calculations using the three PCM methods required significantly different computation times. In principle, the computation time should increase from PCM to SCIPCM. However, since geometry optimization was applied in the PCM (method V in Figures 5 and 6) and SCIPCM (method I) calculations, the IPCM (method IV) calculation became the least expensive in computation cost. The PCM method with geometry optimization (method V), which was previously shown to be useful by Perczel and co-workers,[13] was found to be a little more expensive. The SCIPCM method with geometry optimization (method I) was the most expensive among the three PCM methods applied here. It is important to note that despite the significant change of computation cost the PCM, IPCM, and SCIPCM calculations resulted in quite similar SAAP functions for both glycine and alanine in water. This in turn would suggest either that the obtained SAAP functions are already sufficient in accuracy or that this is a limit of the PCM approximation.

Coincidence between the SAAPs in water and the *statistical* conformations of the amino acids in proteins (i.e., Ramachandran plots) was previously pointed out experimentally[7,8] and theoretically.[12] The SCIPCM functions obtained here for glycine and alanine in water (Figures 1 and 3, respectively) strongly support their presence. If the observed coincidence was merely due to an artificial factor of the continuous solvent models, then it must be very curious to find such a *perfect* agreement between the SAAPs in water and the distribution of the Ramachandran plots as seen in Figures 5 and 6. This, in other words, indicates that the coincidence may be an inevitable consequence of some laws underlying in nature. It would be reasonable to consider that the SAAPs in water obtained by the PCM methods are reliable. Implications of the quantitative reflection of the Ramachandran plots on the SAAPs in water are discussed below in relation to protein secondary and tertiary structures.

Implications for Secondary Structures of Proteins. The SAAP functions obtained for glycine and alanine in water (Figures 1 and 3) showed wide potential wells in the α-helix and β-sheet regions. This implies that the amino acid residues have intrinsic preferences to readily adopt the conformations corresponding to the secondary structures in water. The preferences can be extended to a polypeptide molecule as well. Thus, the conformational space of a polypeptide chain would be

significantly biased to secondary structures in water. This entropic (or kinetic) effect must be important for rapid formation of the secondary structures in the early stage of protein folding,[27] as reasonably reproduced by molecular simulations of polyalanines by applying the obtained SCIPCM potentials to the SAAP force field.

Table 3 clearly shows that helical (3_{10}) structures are remarkably stable in water for short polyalanines, suggesting that the helical structures are major components of the structure ensemble in the random coil state in water.[26] Since interresidue hydrogen bonds within the polypeptide chain should be very weak in water ($\epsilon \approx 80$), it is unlikely that the polar hydrogen bond plays an energetically important role in the formation of the secondary structures in water. Instead, the conformational space of a polypeptide must be of primary importance for the formation.[3]

The above mechanism of secondary structure formation might somewhat conflict with the general view that distant inter-amino-acid hydrogen bonds are essential for the formation of α-helices and β-sheets. In our opinion, such hydrogen bonds merely characterize the structures of α-helices and β-sheets, and it must be the intrinsic conformational preferences of the individual amino acids (the SAAPs) that effectively guide the polypeptide chain to rapid formation of the secondary structures in water. Once the secondary structures formed and the hydrophobic collapse occurred, then the hydrogen bonds would become stronger and stabilize the secondary structures energetically.

In our Monte Carlo simulations of polyalanines (Table 3), only 3_{10}-helical structures, not α-helical ones, have been characterized. The results are in accord with experimental observations.[26] The reason for the prominent stability of α-helices, instead of 3_{10}-helices, in folded proteins is not clear at the moment within the framework of the SAAP formalism. However, it should be noted that the observed 3_{10}-helical structures of polyalanines in water gradually deform from the normal 3_{10}-helix toward loose α-helical structures with the increasing length of the chain (Figures S1−S3 of the Supporting Information).

Implications for Tertiary Structures of Proteins. The conformation of each amino acid residue in a protein must be mainly determined by two factors: the intrinsic conformational preferences of the amino acid by itself and the weak nonbonded interactions from the surrounding amino acid residues and/or other molecules, such as prosthetic groups and solvent water. The Ramachandran plots for one protein reflect both factors that are present in the native structure. Therefore, each protein shows a different distribution pattern of the Ramachandran plots depending on the structure.

The correlation plots shown in Figures 5 and 6, however, revealed that the Ramachandran plots for a large number of protein structures exactly reflect the conformational surface of the single amino acid in water: The linear correlations quantitatively showed that the statistical conformations in proteins follow the random distribution on the SAAP surface (i.e., Boltzmann distribution). In general, protein folds in water. This implies that the conformations of each amino acid in the unfolded state (i.e., before protein folds) randomly obey the force field of the SAAP in water. Therefore, it is obvious that the SAAP functions in water are of significant importance not only for unfolded random coil structures of proteins but also for the native folded structures.

What do the linear correlations in Figures 5 and 6 imply for protein tertiary structures? Two interpretations may be hypothesized. First, the mean force of nonbonded interactions around glycine and alanine residues would become zero over a large number of protein structures. In other words, the protein structures look highly disordered *statistically*. This statistical view would be very unusual because the structures of proteins are believed to have been selected since the birth of life. Second, the force field that each amino acid residue feels in the unfolded state of a protein seems to be the same as the average force field that the amino acid *statistically* feels over a large number of folded protein structures. This dynamic feature is even more surprising because it might indicate that the folding pathways are *statistically* not biased to any directions on the SAAP surfaces, which may be, or may not be, in conflict with the normal hierarchical convergent scenarios.[28] These considerations are based on the SAAPs in water obtained only for glycine and alanine. The generality will be examined as soon as we obtain a complete set of SAAP parameters for all 20 amino acids of proteins.

Conclusions

We have obtained reliable SAAP functions for glycine and alanine in water and ether (i.e., hydrophobic environments) by applying the SCIPCM method. The potential surfaces in water have distinct potential wells in the α- and β-regions, suggesting that the SAAPs in water are important for the formation of protein secondary structures. Moreover, the SAAPs in water are in almost perfect agreement with the Ramachandran plots over a large number of nonredundant protein structures. This strongly suggested that the SAAPs are also important for understanding folded protein structures. Molecular simulations of For-Ala$_n$-NH$_2$ ($n = 3-5$) by using the SAAP force field[3] equipped with the SCIPCM potentials showed that helical (3_{10}) structures are significantly stable in water but not in vacuo. The results are well consistent with recent experimental observations.[26] In ether, the helical structures were relatively stable, but the most stable structure was assigned to a different one.

The almost perfect agreement between the SAAPs in water and the Ramachandran plots should have significant implications for protein structures and the folding pathways. Two possibilities have been postulated: The protein structures are highly disordered *statistically*, and the folding pathways are *statistically* not biased to any directions on the SAAP surfaces. These considerations may be very elusive at the moment. To better understand the scientific meanings of the observed coincidence, much elaborate study needs to be done from both experimental and theoretical points of view.

Acknowledgment. This work was supported by Grants-in-Aid for Scientific Research on Priority Areas (C) "Genomic Information Science" (Grant Nos. 15014211 and 16014223) and Scientific Research (B) (Grant No. 16350092) from the Ministry of Education, Culture, Sports, Science and Technology of Japan. This work was also supported in part by a Fellowship to Researchers from the Association for the Progress of New Chemistry of Japan and by the Research and Study Program of the Tokai University Educational System General Research Organization.

Supporting Information Available: Atomic coordinates of the potential energy minima obtained for **1** and **2** and molecular structures of the helical and extended conformations obtained for For-Ala$_n$-NH$_2$ ($n = 3-5$) by using the SAAP force field. This information is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) (a) Creighton, T. E. *Proteins: Structure and Molecular Principles*, 2nd ed.; Freeman: New York, 1993. (b) Lazaridis, T.; Karplus, M. *Biophys. Chem.* **2003**, *100*, 367−395. (c) Ferguson, N.; Fersht, A. R. *Curr. Opin. Struct. Biol.* **2003**, *13*, 75−81. (d) Dobson, C. M. *Nature* **2003**, *426*, 884−890.

(2) (a) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S. Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217. (b) Weiner, S. J.; Kollman, P. A., Case, D. A.; Singh, U. C., Ghio, C.; Alagona, G.; Profeta, S., Jr.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765−784. (c) Weiner, S. J.; Kollman, P. A., Nguyen, D. T.; Case, D. A. *J. Comput. Chem.* **1986**, *7*, 230−252. (d) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.

(3) Iwaoka, M.; Tomoda, S. *J. Comput. Chem.* **2003**, *24*, 1192−1200.

(4) Ramachandran, G. N.; Sasisekharan, V. *Adv. Protein Chem.* **1968**, *23*, 283−437.

(5) (a) Williams, R. W.; Chang, A.; Juretiæ, D.; Loughran, S. *Biochim. Biophys. Acta* **1987**, *916*, 200−204. (b) Wilmot, C. M.; Thornton, J. M. *J. Mol. Biol.* **1988**, *203*, 221−232.

(6) (a) O'Neil, K. T.; DeGrado, W. F. *Science* **1990**, *250*, 646−651. (b) Lyu, P. C.; Liff, M. I.; Marky, L. A.; Kallenbach, N. R. *Science* **1990**, *250*, 669−673. (c) Padmanabhan, S.; Marqusee, S.; Ridgeway, T.; Laue, T. M.; Baldwin, R. L. *Nature* **1990**, *344*, 268−270.

(7) (a) Smith, L. J.; Bolin, K. A.; Schwalbe, H.; MacArthur, M. W.; Thornton, J. M.; Dobson, C. M. *J. Mol. Biol.* **1996**, *255*, 494−506. (b) Smith, L. J.; Fiebig, K. M.; Schwalbe, H.; Dobson, C. M. *Folding Des.* **1996**, *1*, R95−R106.

(8) (a) Fiebig, K. M.; Schwalbe, H.; Buck, M.; Smith, L. J.; Dobson, C. M. *J. Phys. Chem.* **1996**, *100*, 2661−2666. (b) Bolin, K. A.; Pitkeathly, M.; Miranker, A.; Smith, L. J.; Dobson, C. M. *J. Mol. Biol.* **1996**, *261*, 443−453. (c) Penkett, C. J.; Redfield, C.; Dodd, I.; Hubbard, J.; McBay, D. L.; Mossakowska, D. E.; Smith, R. A. G.; Dobson, C. M.; Smith, L. J. *J. Mol. Biol.* **1997**, *274*, 152−159.

(9) (a) Zhang, X.; Xu, Y.; Zhang, J.; Wu, J.; Shi, Y. *Biochemistry* **2005**, *44*, 8117−8125. (b) Dyer, R. B.; Maness, S. J.; Franzen, S.; Fesinmeyer, R. M.; Olsen, K. A.; Andersen, N. H. *Biochemistry* **2005**, *44*, 10406−10415. (c) Li, Y.; Picart, F.; Raleigh, D. P. *J. Mol. Biol.* **2005**, *349*, 839−846.

(10) (a) Head-Gordon, T.; Head-Gordon, M.; Frisch, M. J.; Brooks, C. L., III.; Pople, J. A. *J. Am. Chem. Soc.* **1991**, *113*, 5989−5997. (b) Endrédi, G.; Perczel, A.; Farkas, O.; McAllister, M. A.; Csonka, G. I.; Ladik, J.; Csizmadia, I. G. *J. Mol Struct. (THEOCHEM)* **1997**, *391*, 15−26. (c) Cornell, W. D.; Gould, I. R.; Kollman, P. A. *J. Mol. Struct. (THEOCHEM)* **1997**, *392*, 101−109. (d) Oldziej, S.; Kozlowska, U.; Liwo, A.; Scheraga, H. A. *J. Phys. Chem. A* **2003**, *107*, 8035−8046.

(11) Shang, H. S.; Head-Gordon, T. *J. Am. Chem. Soc.* **1994**, *116*, 1528−1532.

(12) Iwaoka, M.; Okada, M.; Tomoda, S. *J. Mol. Struct. (THEOCHEM)* **2002**, *586*, 111−124.

(13) Hudáky, I.; Hudáky, P.; Perczel, A. *J. Comput. Chem.* **2004**, *25*, 1522−1531.

(14) Wang, Z.-X.; Duan, Y. *J. Comput. Chem.* **2004**, *25*, 1699−1716.

(15) Onsager, L. *J. Am. Chem. Soc.* **1936**, *58*, 1486−1493.

(16) Miertus, S.; Scrocco, E.; Tomasi, J. *Chem. Phys.* **1981**, *55*, 117−129.

(17) Foresman, J. B.; Keith, T. A.; Wiberg, K. B.; Snoonian, J.; Frisch, M. J. *J. Phys. Chem.* **1996**, *100*, 16098−16104.

(18) (a) Sakae, Y.; Okamoto, Y. *Chem. Phys. Lett.* **2003**, *382*, 626−636. (b) Sakae, Y.; Okamoto, Y. *J. Theor. Comput. Chem.* **2004**, *3*, 339−358. (c) Sakae, Y.; Okamoto, Y. *J. Theor. Comput. Chem.* **2004**, *3*, 359−378.

(19) (a) Cossi, M.; Scalmani, G.; Rega, N.; Barone, V. *J. Chem. Phys.* **2002**, *117*, 43−54. (b). Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. *J. Comput. Chem.* **2003**, *24*, 669−681.

(20) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision B.04; Gaussian, Inc.: Wallingford, CT, 2004.

(21) (a) Benzi, C.; Improta, R.; Scalmani, G.; Barone, V. *J. Comput. Chem.* **2002**, *23*, 341−350. (b) Langella, E.; Rega, N.; Improta, R.; Crescenzi, O.; Barone, V. *J. Comput. Chem.* **2002**, *23*, 650−661. (c) Improta, R.; Mele, F.; Crescenzi, O.; Benzi, C.; Barone, V. *J. Am. Chem. Soc.* **2002**, *124*, 7857−7865. (d) Langella, E.; Improta, R.; Barone, V. *J. Am. Chem. Soc.* **2002**, *124*, 11531−11540. (e) Baldauf, C.; Günther, R.; Hofmann, H.-J. *J. Org. Chem.* **2004**, *69*, 6214−6220. (f) Jeon, I.-S.; Ahn, D.-S.; Park, S.-W.; Lee, S.; Kim, S. K. *Chem. Phys. Lett.* **2005**, *403*, 72−76. (g) Wan, J.; Xu, X.; Ren, Y.; Yang, G. *J. Phys. Chem. B* **2005**, *109*, 11088−11090.

(22) Begue, D.; Carbonniere, P.; Barone, V.; Pouchan, C. *Chem. Phys. Lett.* **2005**, *416*, 206−211.

(23) Nyerges, B.; Kovács, A. *J. Phys. Chem. A* **2005**, *109*, 892−897.

(24) (a) Ribeiro-Claro, P. J. A.; Vaz, P. D. *Chem. Phys. Lett.* **2004**, *390*, 358−361. (b) Kleinpeter, E.; Klod, S.; Rudorf, W.-D. *J. Org. Chem.* **2004**, *69*, 4317−4329. (c) Tejero, I.; Huertas, I.; González-Lafont, Á.; Lluch, J. M.; Marquet, J. *J. Org. Chem.* **2005**, *70*, 1718−1727. (d) Roos, G.; Proft, F. D.; Geerlings, P. *J. Phys. Chem. A* **2005**, *109*, 652−658. (e) Keresztury, G.; István, K.; Sundius, T. *J. Phys. Chem. A* **2005**, *109*, 7939−7945. (f) Vrèek, V.; Vrèek, I. V.; Siehl, H.-U. *J. Phys. Chem. A* **2006**, *110*, 1868−1874.

(25) Sussman, J. L.; Lin, D. W.; Jiang, J. S.; Manning, N. O.; Prilusky, J.; Ritter, O.; Abola, E. E. *Acta Crystallogr., Sect. D* **1998**, *54*, 1078−1084.

(26) (a) Miick, S. M.; Martinez, G. V.; Fiori, W. R.; Todd, A. P.; Millhauser, G. L. *Nature* **1992**, *359*, 653−655. (b) Millhauser, G. L.; Stenland, C. J.; Hanson, P.; Bolin, K. A.; van de Ven, F. J. M. *J. Mol. Biol.* **1997**, 267, 963−974. (c) Schweitzer-Stenner, R.; Eker, F.; Huang, Q.; Griebenow, K. *J. Am. Chem. Soc.* **2001**, *123*, 9628−9633.

(27) (a) Kuwajima, K. *Proteins: Struct., Funct., Genet.* **1989**, *6*, 87−103. (b) Christensen, H.; Pain, R. H. *Eur. Biophys. J.* **1991**, *19*, 221−229. (c) Qi, P. X.; Sosnick, T. R.; Englander, S. W. *Nat. Struct. Biol.* **1998**, *5*, 882−884. (d) Akiyama, S.; Takahashi, S.; Ishimori, K.; Morishima, I. *Nat. Struct. Biol.* **2000**, *7*, 514−520.

(28) (a) Karplus, M.; Weaver, D. L. *Biopolymers* **1979**, *18*, 1421−1437. (b) Harrison, S. C.; Durbin, R. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 4028−4030. (c) Kim, P. S.; Baldwin, R. L. *Annu. Rev. Biochem.* **1990**, *59*, 631−660.