Computational Analysis of Stability of the β-Sheet Structure

Daisuke Katagiri,* Takahiro Tsuchiya, Minoru Tsuda, Masayuki Hata, and Tyuji Hoshino

Graduate School of Pharmaceutical Sciences, Chiba University, 1-33, Yayoi-cho, Inage-ku, Chiba 263-8522, Japan

Received: March 12, 2002; In Final Form: June 5, 2002

Ab initio quantum chemical calculations were carried out on the main factor responsible for the formation of β-sheet structures in specific regions of proteins. Geometry optimization was performed at the Hartree−Fock level using a 3-21G** basis functional set. Total energies of α-helix and β-sheet structures were compared using several types of computational models. The computational models for the β-sheet structure were constructed by extracting a part of the protein from X-ray crystallographic data of HIV-1 or myosin, and the computational models for the α-helix structure were constructed by modifying the model for the β-sheet structure. The results of computations indicated that all helix-like structures were more stable than were the β-sheet structures under the condition of absence of any other molecules. On the other hand, all of the β-sheet structures became more stable than the helix-like structures when water molecules existed around them. This finding suggests that β-sheet structures are stabilized by the presence of water molecules and are primarily determined by local interaction among several neighboring residues.

1. Introduction

The amino acid sequence of proteins is determined by genetic information, which is described as nucleotide sequences in DNA or RNA. A sequence of amino acids specifies the secondary structure of the corresponding peptide, and an assemblage of the second structures determines the 3D structure of proteins. A protein with a specific 3D structure exhibits a particular function. Therefore, genetic information determines the function of proteins through the relation of nucleotide sequences → amino acid sequences → 3D structures → functions of proteins. Classification of protein structures has been studied in recent years for the purpose of predicting protein functions.[1] This approach is based on the fact that the structure of a protein is closely related to its function. Accordingly, it is important to know the 3D structures of proteins to determine the biological roles of protein.[2]

The mapping of the human genome was completed in early in 2001.[3] The 3D structures of almost 10 000 proteins have so for been determined, although more than 150 000 amino acid sequences have been registered in the Protein Data Bank (PDB).[4] Furthermore, more than 14 976 000 nucleotide sequences have been registered in GeneBank.[5] Thus, it is difficult to determine 3D structures of proteins in comparison with the determination of amino acid sequences by a biochemical experimental method. Prediction of 3D structures of proteins by theoretical method has become possible due to recent developments in computational equipment. The reliability of prediction of the 3D structure is, however, low when the amino sequence of the target protein has little homology with those of the proteins whose 3D structures were already known. Hence, another theoretical approach based on the folding of a protein has attracted much attention recently because of its applicability to proteins without helpful homology. Many molecular dynamics simulations have been performed to clarify the detailed process of protein folding. Because proteins consist of atoms, a quantum chemical computational method is also expected to be a feasible approach to provide a clear scheme of making up a protein's 3D structure. In the first step of analysis of protein folding, the process by which local secondary structures such as α-helix and β-sheet structures are formed must be examined in detail. The local secondary structure would initiate the subsequent formation of the entire protein. Therefore, this is one of the key processes of protein folding.

An important question is, what generates the distinction in local secondary structures in reflecting a specific amino acid sequence? One convincing explanation of the protein folding process is that a precursor protein is likely to form a helix-like secondary structure and then the protein formation converts into a specific structure through the transformation leading the energetic stabilization. On the basis of this explanation, a β-sheet is expected to be formed by transformation from a helix-like structure. As a matter of fact, structural change from helix structures to β-sheet structures was observed in the protein folding of β-lactoglobulin.[6] The results of a molecular dynamics study have also shown that the β-sheet structure is usually formed in the late stage of protein folding.[7] Another question is why α-helix structures transform to other secondary structures at a particular site of the amino acid sequence and turn or β-sheet structures appear at a particular site. The answers to these questions would clarify the process by which local secondary structures are formed. It is therefore important to investigate the origin of the β-sheet structure. From this point of view, theoretical analysis of the formation of the β-sheet structure is needed to understand the process of protein folding. In this study, HIV-1 protease and myosin were examined to try to determine the reason for the formation of a β-sheet structure in a particular region of a protein.

2. Method

To construct model cluster systems for quantum chemical calculations, we extracted Lys45−Lys55 from the X-ray crystallographic structure of HIV-1 protease (pdb code: 3hvp)[8] and Leu399−Leu404 from the X-ray crystallographic structure of myosin (pdb code: 1vom).[9] Each extracted region has a β-sheet structure as can be seen in Figure 1a,b.

As for HIV-1 protease, we constructed two kinds of computational models: "helix-like structure" and "β-sheet structure"

* Corresponding author. Phone: +81-43-290-2926. Fax: +81-43-290-2925. E-mail: katagiri@p.chiba-u.ac.jp.

10.1021/jp025757m CCC: $22.00
Published on Web 08/03/2002

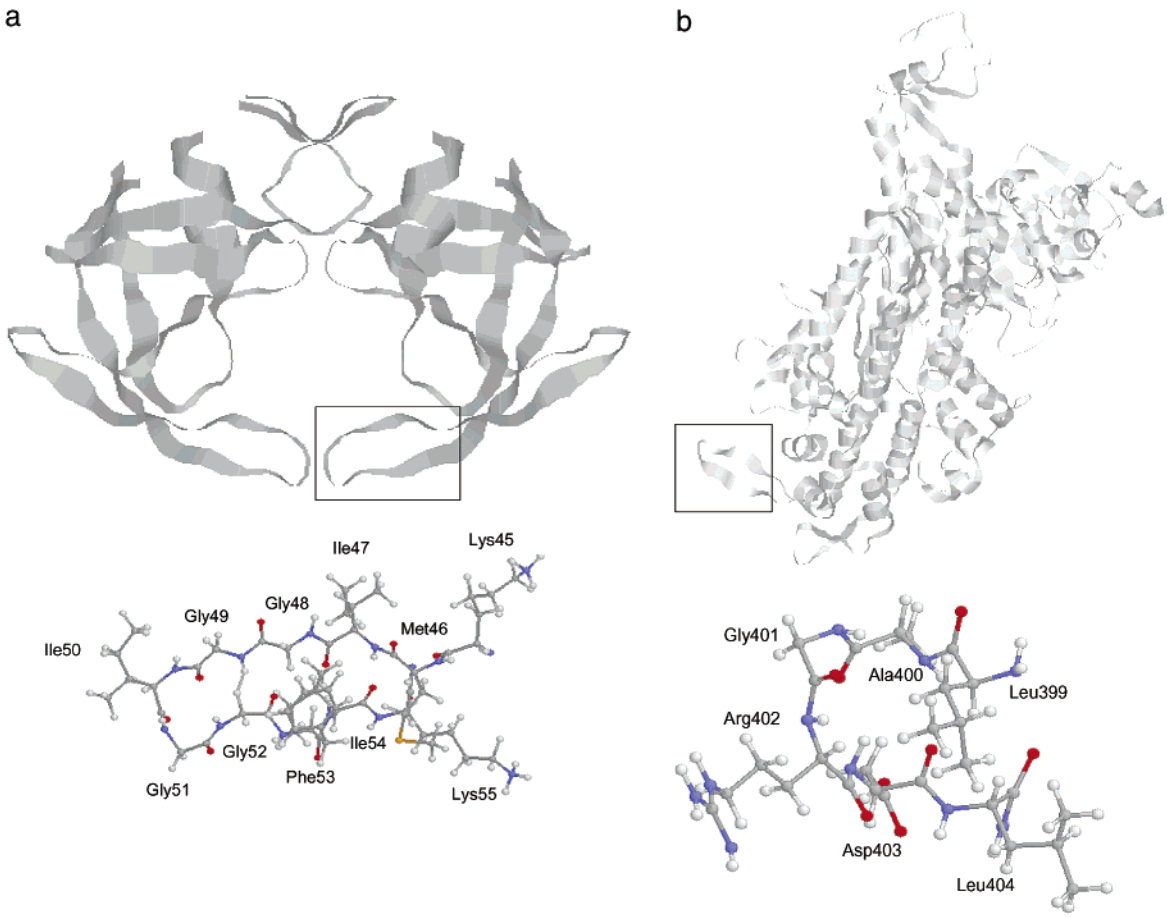

Figure 1. (a) Structure of the whole HIV-1 protease protein (pdb code: 3hvp) and the region extracted for the computational model (Lys45−Lys55). The extracted region is indicated by a square. (b) Structure of the whole myosin protein (pdb code: 1vom) and the region extracted for the computational model (Leu399−Leu400). The extracted region is indicated by a square.

models. Atomic geometries were optimized for both the helix-like and β-sheet structures, and then their total energies were compared. For the computational models representing the helix-like structure, a part (6 residues: Lys45−Ile50) of the above extracted region (11 residues) was modified so as to have the complete helix structure, and 1, 2, 3, 4, or 5 sequential residues from Gly51 to Lys55 were added while the helix structure was maintained. Thus, computations were performed using five models whose residue lengths were 7, 8, 9, 10, and 11. For the models representing β-sheet structure, five similar models with residue lengths of 7, 8, 9, 10, and 11 were constructed using the X-ray crystallographic structure without modification. Further, we picked up two models, one with a residue length of 8 (Lys45−Gly52) and one with a residue length of 11 (Lys45−Lys55). Gly52 is located on the first site of making the turn structure in the X-ray crystallographic data, and Lys55 is located on the last site of the β-sheet structure. Computations were also performed using these models to which was added one or four water molecules, because it would be interesting for the formation of β-sheet structure to consider the effect of hydrogen bonds provided by water molecules. The 8-residue (Lys45−Gly52) model included one water molecule and the 11-residue (Lys45−Lys55) model included four water molecules, and the total energies of the "helix-like structure" and "β-sheet structure" were compared. In the β-sheet structure, the additional water molecules settled at the location of the crystal water observed in the X-ray experiment on HIV-1 protease (pdb code: 7hvp). In the helix-like structure, each water molecule settled to connect with atoms that had hydrogen bonds with the corresponding water molecule in the β-sheet structure.

In the case of myosin, we constructed cluster models of two sizes: a 4-residue model (Leu399−Arg402) and a 6-residue model (Leu399−Leu404). Arg402 is located on the first site of the turn structure and Leu404 is located on the last site of the β-sheet structure in the X-ray crystallographic data. Two kinds of computational models, helix-like and β-sheet structure models, were prepared for both sizes of clusters. For the helix-like structure, the entire Leu399−Arg402 or Leu399−Leu404 part was modified to have the complete helix structure. On the other hand, the computational models of the β-sheet structure for Leu399−Arg402 or Leu399−Leu404 were constructed using the X-ray crystallographic structure without modification. Atomic geometries were optimized for all computational models, and the total energies of the helix-like and β-sheet structures were compared. Computations were also performed in the presence of water molecules around these two models. The model representing the initial turn structure (Leu399−Arg402) included two water molecules, and the model representing the final β-sheet structure (Leu399−Leu404) included four water molecules. After the geometry optimization, the total energies of the helix-like structure and the β-sheet structure were again compared. In the β-sheet structure, the water molecules were settled at the location of the crystal waters in the X-ray crystallographic data of myosin (pdb code: 1vom). In the helix-like structure, each water molecule connected with atoms that had hydrogen bonds with the corresponding waters in the β-sheet structure.

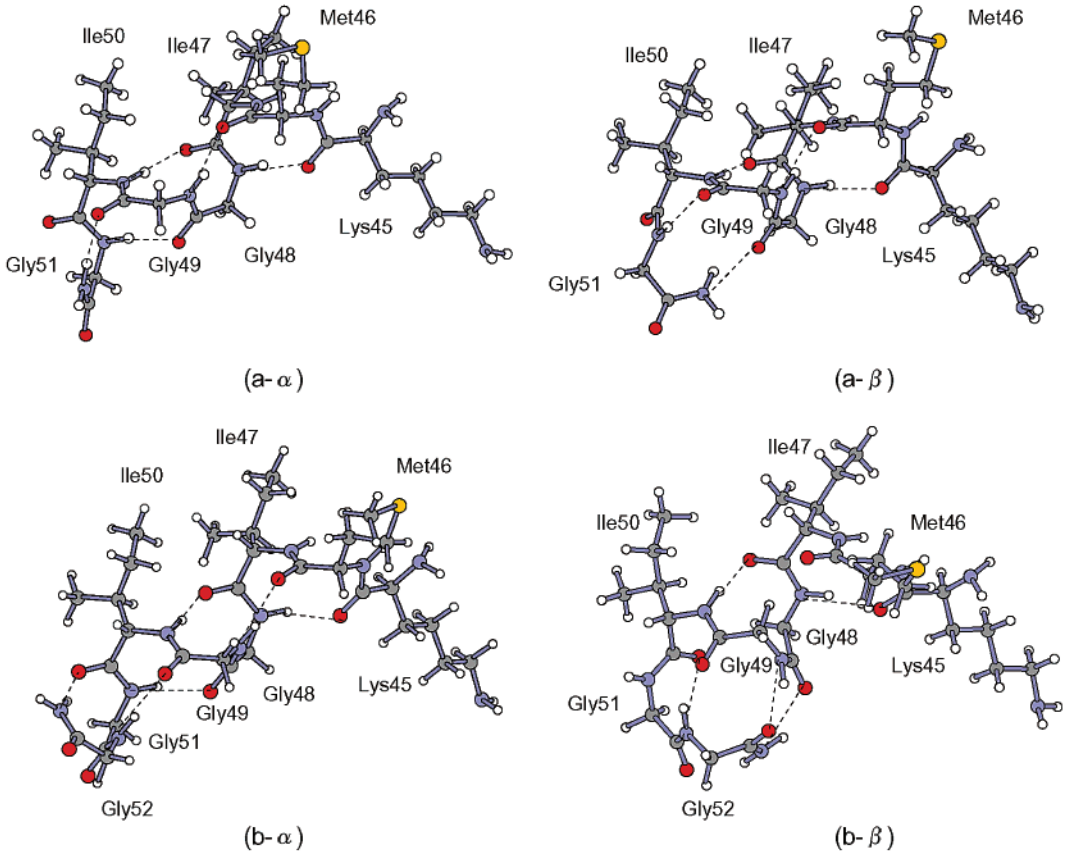

Figure 2. Optimized atomic configurations of the region extracted from HIV-1 protease, obtained in case of the absence of water molecules (7-residue peptide and 8-residue peptide).

The N-terminal side of each model terminated with NH_2, and the C-terminal side terminated with $CONH_2$.

The Schrödinger equation for each model was solved by the ab initio molecular orbital (MO) method. Geometry optimization and energy evaluation were carried out using the Hartree−Fock level. The basis functional set used was 3-21G**. The program package used was Gaussian 98.[10]

3. Results

Atomic geometry optimization was achieved for all sizes of the computational models representing helix-like and β-sheet structures both in the case of HIV-1 protease and in the case of myosin. All atoms were allowed to move freely during the optimization, and the comparison of total energies of the helix and sheet structures was carried out with the respective optimized atomic geometry.

3.1. HIV-1 Protease. *3.1.1. In the Absence of Water Molecules.* The helix-like structure (a-α) and the β-sheet structure (a-β) of Lys45−Gly51 and the helix-like structure (b-α) and the β-sheet structure (b-β) of Lys45−Gly52 are shown in Figure 2.

In the Lys45−Gly51 model, both (a-α) and (a-β) structures have hydrogen bonds between CO of Lys45 and NH of Gly48, between CO of Met46 and NH of Gly49, and between CO of Ile47 and NH of Ile50. That is, a helix structure had been formed in Lys45−Ile50. For Gly51, (a-α) appeared to have hydrogen bonds between NH of Gly51 and CO of Gly48 and between NH in $CONH_2$ of Gly51 and CO of Gly49. Accordingly, (a-α) retains a complete helix structure. On the other hand, (a-β) has hydrogen bonds between NH of Gly51 and CO of Gly49 and between NH in $CONH_2$ of Gly51 and CO of Gly48, indicating the formation of a turn structure. The potential energy difference between (a-α) and (a-β) is 0.50 kcal/mol. The potential energy of (a-α) is lower than that of (a-β). Both of these structures have 5 hydrogen bonds.

In the Lys45−Gly52 model, the optimized structures of (b-α) and (b-β) retain the conformation that has been obtained by adding one residue each (Gly52) to (a-α) and (a-β), respectively. (b-α) shows a complete helix structure with 6 hydrogen bonds, between Lys45 and Gly48, Met46 and Gly49, Ile47 and Ile50, Gly48 and Gly51, Gly49 and Gly52, and Ile50 and Gly52. In contrast, (b-β) loses 3 hydrogen bonds, one between Met46 and Gly49, Gly48 and Gly51, and Ile50 and Gly52 and acquires 2 other hydrogen bonds, between CO of Gly48 and NH of Gly52 and one between CO of Gly52 and NH of Gly49. Thus, (b-β) has a turn-β-sheet structure while retaining a part of the helix structure. (b-α) is more stable than (b-β), and the potential energy difference between them is 8.16 kcal/mol.

The helix-like structure (c-α) and the β-sheet structure (c-β) of Lys45−Phe53, the helix-like structure (d-α) and the β-sheet structure (d-β) of Lys45−Ile54, and the helix-like structure (e-α) and the β-sheet structure (e-β) of Lys45−Lys55 are shown in Figure 3.

For the Lys45−Phe53 model, (c-α) and (c-β) were obtained by adding one residue each to (b-α) and (b-β). The helix-like structure (c-α) has 7 hydrogen bonds (Lys45−Gly48, Met46−Gly49, Ile47−Ile50, Gly48−Gly51, Gly49−Gly52, Ile50−Phe53, Gly51−Phe53), and a part of the helix structure is tightly twisted, whereas (c-β) has 4 hydrogen bonds (Lys45−Gly48, Ile47−Phe53, Gly49−Gly52, Gly52−Gly49) and shows a turn-β-sheet structure while retaining part of the helix structure. The potential energy difference between (c-α) and (c-β) is 11.29 kcal/mol, and (c-α) is more stable than (c-β). For the Lys45−Ile54 model, (d-α) and (d-β) were expanded by adding one

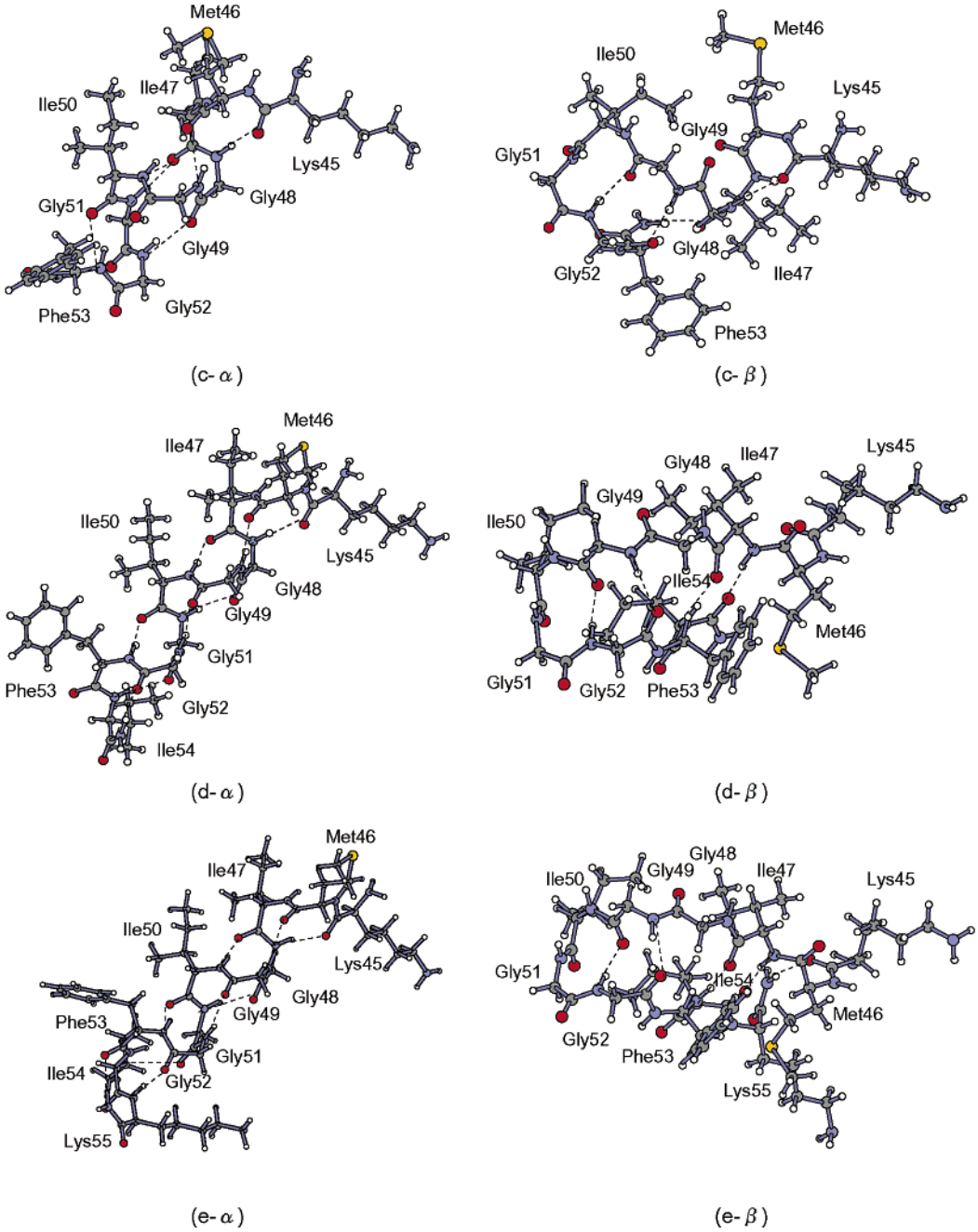

Figure 3. Optimized atomic configurations of the region extracted from HIV-1 protease, obtained in case of the absence of water molecules (9-residue peptide, 10-residue peptide, and 11-residue peptide).

residue to (c). (d-α) has a complete helix structure with 8 hydrogen bonds (Lys45−Gly48, Met46−Gly49, Ile47−Ile50, Gly48−Gly51, Gly49−Gly52, Ile50−Phe53, Gly51−Ile54, Gly53−Ile54), whereas (d-*β*) had completely lost the initial helix structure and has only 4 hydrogen bonds (Ile47−Ile54, Gly49−Gly52, Gly52−Gly49, Ile54−Ile47). The potential energy of (d-α) is lower than that of (d-*β*) by 14.01 kcal/mol.

For the Lys45−Lys55 model, (e-α) and (e-*β*) were expanded by adding one residue to (d). (e-α) has 9 hydrogen bonds (Lys45−Gly48, Met46−Gly49, Ile47−Ile50, Gly48−Gly51, Gly49−Gly52, Ile50−Phe53, Gly51−Ile54, Gly52−Lys55, Phe53−Lys55) and shows a complete helix structure, whereas (e-*β*) has 5 hydrogen bonds (Ile45−Lys55, Ile47−Ile54, Gly49−Gly52, Gly52−Gly49, Ile54−Ile47) and shows a turn-*β*-sheet structure with complete loss of the helix. The potential energy difference between (e-α) and (e-*β*) is 15.18 kcal/mol, and (e-α) is more stable than is (e-*β*).

3.1.2. In the Presence of Water Molecules. The helix-like structure (f-α) and the *β*-sheet structure (f-*β*) of Lys45−Gly52 contain water molecules in the model reaction system, as do (g-α) and (g-*β*) of Lys45−Lys55, as shown in Figure 4.

For the Lys45−Gly52 model, (f-α) had lost 1 hydrogen bond between Gly48 and Gly51, as was seen in (b-α), and instead had acquired 2 hydrogen bonds, one between CO of Gly48 and H of a water molecule and one between NH of Gly51 and O of the water molecule. (f-*β*) has no hydrogen-bond dissociation compared to that of (b-*β*) but has new hydrogen bonds between CO of Gly49 and H of a water molecule and between NH of Gly51 and O of the water molecule. Both these helix-like and *β*-sheet structures have 7 hydrogen bonds. The number of

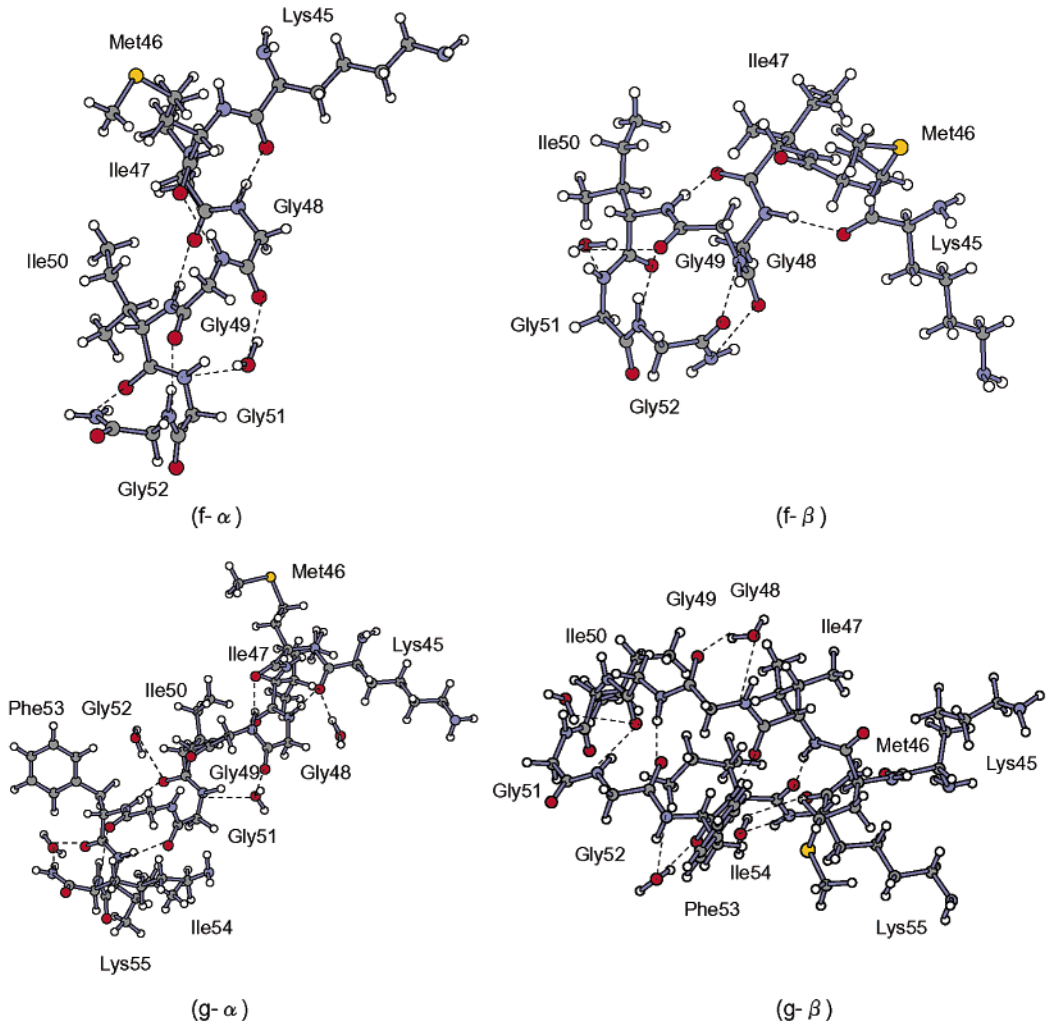

Figure 4. Optimized atomic configurations of the region extracted from HIV-1 protease, obtained in case of the presence of water molecules (8-residue peptide and 11-residue peptide).

hydrogen bonds in (b-α) and (b-β) were 6 and 5, respectively. Accordingly, the water molecule is of great assistance in the formation of hydrogen bonds in the β-sheet structure, making use of the atom that did not participate in the formation of hydrogen bonds in (b-β). It cannot be said that (f-β) is energetically more stable than (f-α). The potential energy difference between (f-α) and (f-β) is 1.06 kcal/mol.

For the Lys45−Lys55 model, (g-α) has 13 hydrogen bonds. The hydrogen bonds between Gly48−Gly51 and Phe53−Lys55 in (e-α) had disappeared and, in turn, there had appeared 6 new hydrogen bonds in (g-α). (g-β) also has 13 hydrogen bonds. No hydrogen bond had disappeared from (e-β) to (g-β), but 8 new hydrogen bonds had appeared in (g-β). The potential energy of (g-β) is also lower than that of (g-α), and the potential energy difference is 2.07 kcal/mol.

3.2. Myosin. *3.2.1. In the Absence of Water Molecules.* The helix-like structure, (h-α) and (i-α), and the β-sheet structure, (h-β) and (i-β), of Leu399−Arg402 and Leu399−Leu404 are shown in Figure 5.

For the Leu399−Arg402 model, (h-α) has a helix structure with 2 hydrogen bonds (Leu399−Arg402 and Ala400−Arg402), and (h-β) has a turn structure with 1 hydrogen bond (Leu399−Arg402). The potential energy of (h-α) is lower than that of (h-β) by 2.01 kcal/mol.

For the Leu399−Leu404 model, (i-α) has additional 3 more hydrogen bonds than (h-α) (Gly401−Leu404, Arg402−Leu404, OH−CO in Asp403) and shows a helix structure. (i-β) also has 3 more hydrogen bonds than (h-β) (Leu399−Gly401, Arg402−Leu404, Asp403−Leu404) and shows a sheet structure. (i-α) has a total of 5 hydrogen bonds, whereas (i-β) has 4 hydrogen bonds. The potential energy of (i-α) is lower than that of (i-β) by 4.45 kcal/mol.

3.2.2. In the Presence of Water Molecules. Figure 6 shows the helix-like structure (j-α) and the β-sheet structure (j-β) of Leu399−Arg402, which contains water molecules in the model system, as well as (k-α) and (k-β) of Leu399−Leu404.

For the Leu399−Arg402 model, (j-α) has 4 hydrogen bonds and a partially loosened helix structure. (j-β) also has 4 hydrogen bonds and shows a turn structure. The potential energy of (j-β) is lower than that of (j-α) by 9.16 kcal/mol. For the Leu399−Leu404 model, (k-α) has a helix structure with 5 more hydrogen bonds than (j-α). In contrast, (k-β) has a sheet structure with 6 more hydrogen bonds than (j-β). Accordingly, both of these structures have 10 hydrogen bonds. The potential energy difference between (k-α) and (k-β) is 6.21 kcal/mol, and (k-β) is more stable.

4. Discussion

4.1. Role of Water Molecules in the Stability of the β-Sheet Structure. In the absence of water molecules, all of the helix-like structures were more stable than were the β-sheet structures consisting of the same residues. However, in the presence of

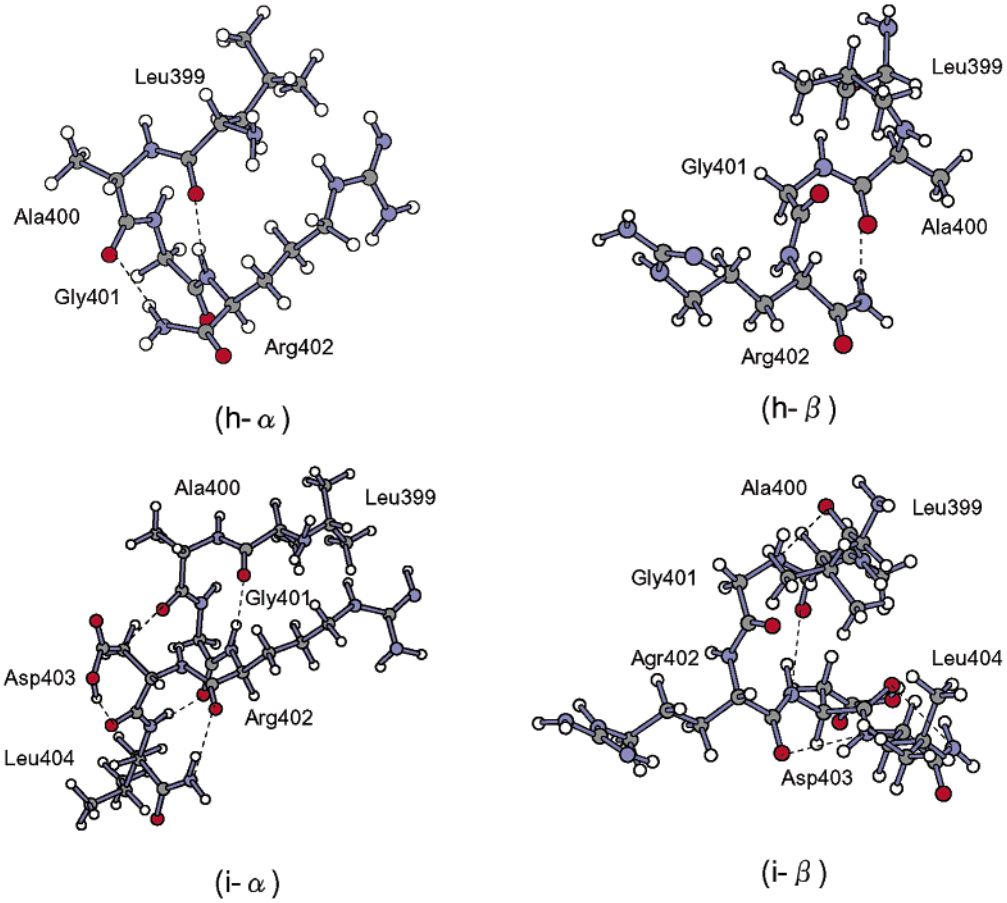

Figure 5. Optimized atomic configurations of the region extracted from myosin, obtained in case of the absence of water molecules (4-residue peptide and 6-residue peptide).

water molecules, the stability of all of the β-sheet structures derived from X-ray crystallographic data was greater than that of the helix-like structure (see Tables 1−4). The results of HIV-1 protease in the case of the absence of water molecules (see Table 1) showed that the energetic advantage of the helix-like structure to the β-sheet structure gradually increased with an increase in the difference between the numbers of hydrogen bonds in them and that the energetic difference became larger with an increase in the number of residues. These results suggest that hydrogen bonds contribute greatly to HIV-1 protease as a structure-decision factor to assist in the formation of a helix-like or β-sheet structure.

If the numbers of hydrogen bonds in the helix-like and β-sheet structures are the same, the β-sheet structure may become more stable. A good example of this was demonstrated in the case of the presence of water molecules (See Table 2.). The energetic advantage of the β-sheet structure gradually increased with an increase in the number of residues, that is, the stabilization energy caused by an increase in the number of hydrogen bonds made in water molecules. In myosin, the computational results in the case of the absence of water molecules showed that the helix-like structure became stabilized with an increase in the number of residues and, furthermore, that the difference between the number of hydrogen bonds in the helix-like and β-sheet structures changed (see Table 3.). This suggests that not only hydrogen bonds but also the effect of neighboring residues is greatly involved in the formation of the helix-like or β-sheet structure. Therefore, in the case of the presence of water molecules (see Table 4), the energetic advantage of the β-sheet structure to the helix-like structure decreased as the number of residues increased. In the present study, it appeared that formation of the β-sheet structure was determined not only by interaction with several residues in the neighborhood but also by interaction with solvent water molecules. Thus, interaction with water molecules seems to be indispensable for stabilizing the β-sheet structure. The advantage of interaction with water molecules is that water molecules assist in the formation of hydrogen bonds between NH or CO groups of the amino acid residue that had no hydrogen bond in case of the absence of solvent water molecules. The presence of solvent water molecules results in an increase in the number of hydrogen bonds in the β-sheet structure to a level comparable with that of the helix-like structure. Hydrogen bonds have been reported to stabilize the protein structure by about 2.87−7.17 kcal/mol per hydrogen bond.[11] This finding is consistent with the finding in this study, in which the average stabilization energy per hydrogen bond was 3.64 kcal/mol.

Attention should be given to the location of solvent water molecules. Remarkably effective water is seen in the models of an 8-residue peptide of HIV-1 protease (f-β) and a 4-residue peptide of myosin (j-β). For HIV-1 protease, only one water molecule stabilizes the β-sheet structure by 9.22 kcal/mol. This is the most largely contributing water for the stability of β-sheet structure among the waters existing in the HIV-1 protease models. Two water molecules in myosin stabilize the β-sheet structure by 11.16 kcal/mol, i.e., , 5.58 kcal/mol per a water molecule. Because the four water molecules of the 6-residue peptide of myosin (k-β) stabilize the β-sheet structure by 2.67 kcal/mol per a water molecule, the initial two water molecules incorporated in the 4-residue peptide model contribute the most. These remarkably effective water molecules are located at the

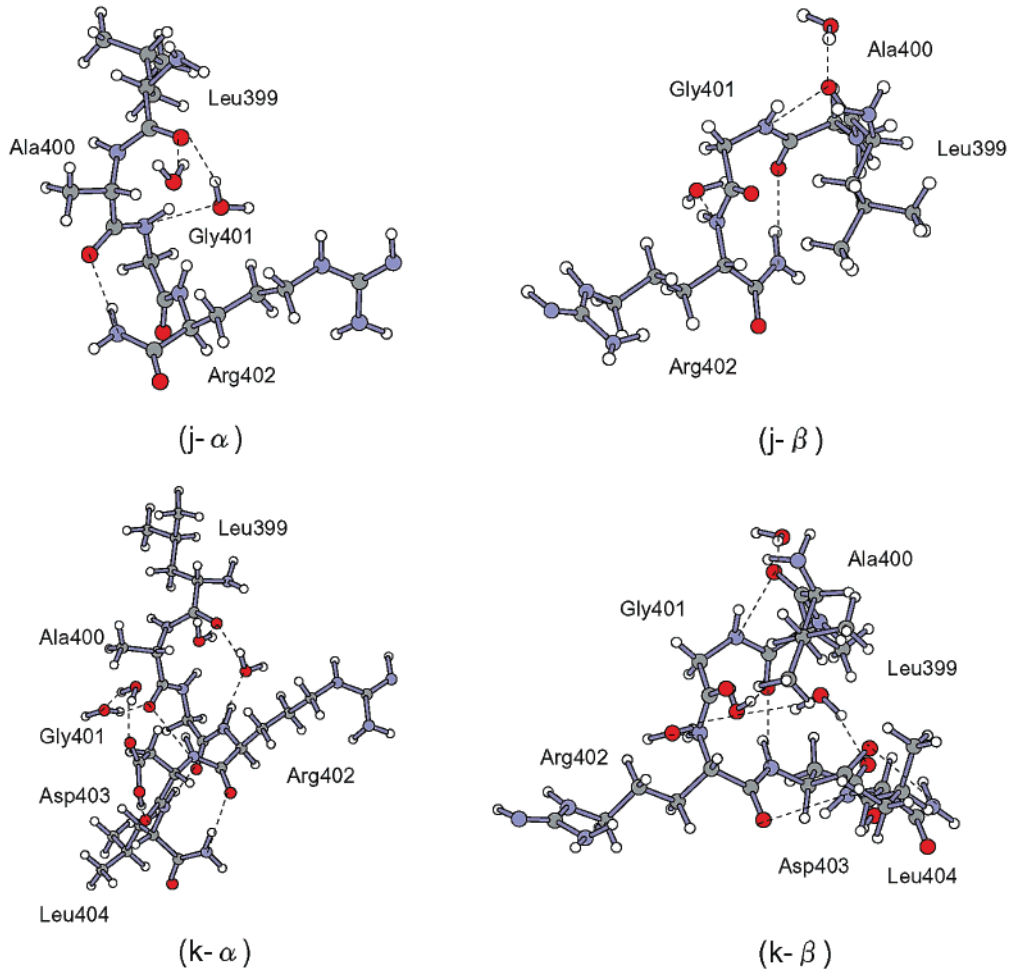

Figure 6. Optimized atomic configurations of the region extracted from myosin, obtained in case of the presence of water molecules (4-residue peptide and 6-residue peptide).

TABLE 1: Comparison of Potential Energies and Numbers of Hydrogen Bonds of the Helix-like Structure and β-Sheet Structure in the Case of the Absence of Water Molecules for HIV-1 Protease

	energy diff $E_{\text{helix-like}} - E_{\beta\text{-sheet}}$ (kcal/mol)	no. of hydrogen bonds: helix-like structure	no. of hydrogen bonds: β-sheet structure
7-residue peptide (Lys45−Gly51)	−0.50	5	5
8-residue peptide (Lys45−Gly52)	−8.16	6	5
9-residue peptide (Lys45−Phe53)	−11.29	7	4
10-residue peptide (Lys45−Ile54)	−14.01	8	4
11-residue peptide (Lys45−Lys55)	−15.18	9	5

TABLE 2: Comparison of Potential Energies and Numbers of Hydrogen Bonds of the Helix-like Structure and β-Sheet Structure in the Case of the Presence of Water Molecules for HIV-1 Protease

	energy diff $E_{\text{helix-like}} - E_{\beta\text{-sheet}}$ (kcal/mol)	no. of hydrogen bonds: helix-like structure	no. of hydrogen bonds: β-sheet structure
8-residue peptide (Lys45−Gly52)	1.06	7	7
11-residue peptide (Lys45−Lys55)	2.07	13	13

TABLE 3: Comparison of Potential Energies and Numbers of Hydrogen Bonds of the Helix-like Structure and β-Sheet Structure in the Case of the Absence of Water Molecules for Myosin

	energy diff $E_{\text{helix-like}} - E_{\beta\text{-sheet}}$ (kcal/mol)	no. of hydrogen bonds: helix-like structure	no. of hydrogen bonds: β-sheet structure
4-residue peptide (Leu399−Arg402)	-2.01	2	1
6-residue peptide (Leu399−Leu404)	-4.45	5	4

turn region. Thus, the presence of water molecules is thought to be particularly important for the formation of a turn structure.

4.2. Possibility of Transformation of Helix and β-Sheet Structures. The presence of water molecules resulted in an increase in the number of hydrogen bonds in the model systems and stabilized the β-sheet structure as described in section 3.1.2 and 3.2.2. Water molecules function to maintain the number of hydrogen bonds by recovering the reduction of hydrogen bonds due to conformational change of the "helix structure → β-sheet structure". Thus, it is reasonable to assume that water molecules play a role in initiation of the formation of a β-sheet structure by assisting the disintegration of the helix-like structure and the generation of a turn structure by means of creating new hydrogen bonds. The observation of water-bound structures indicated that water molecules efficiently made hydrogen bonds at the surface of the peptide. The configuration of a β-sheet structure is like a sheet of paper, and that of a helix-like structure is like a roll of a cylinder. Therefore, the surface area of a

TABLE 4: Comparison of Potential Energies and Numbers of Hydrogen Bonds of the Helix-like Structure and β-Sheet Structure in the Case of the Presence of Water Molecules for Myosin

	energy diff	no. of hydrogen bonds	
	$E_{\text{helix-like}} - E_{\beta\text{-sheet}}$ (kcal/mol)	helix-like structure	β-sheet structure
4-residue peptide (Leu399−Arg402)	9.16	4	4
6-residue peptide (Leu399−Leu404)	6.21	10	10

β-sheet is longer than that of a helix-like structure. The β-sheet structures had more atoms exposed to solvent than did the helix-like structures, which would result in an advantage of formation of hydrogen bonds by water molecules.

The above explanation is compatible with previously reported results of theoretical and experimental works. Results of a simulation have shown that water molecules are present in the protein core up to the late stage of protein folding and that the β-sheet is formed at the late stage.[7] Results of another simulation have indicated that the β-sheet is formed by the turn structure and that the β hairpin is formed from the N terminus.[12] Kinetic measurements have also shown that solvent conditions are of significant importance for the formation of the β-sheet structure.[13] It is notable that a hydrogen bond appears between NH of Gly51 and a water molecule in the case of the β-sheet structure of HIV-1 protease in (f-β) and between NH of Arg402 and a water molecule in the case of myosin in (j-β). That is, water molecules coordinate to saturate the unbonded NH or CO of the main chain when amino acid residues have a nonhelix formation due to the formation of a turn structure. On the basis of these findings, we make the following speculation regarding the formation of a β-sheet structure. First, a whole protein is held to the coil-structure-like helix. Then a turn structure emerges in a part of the protein due to stabilization by water molecules. Finally, this stabilized part expands by incorporating adjacent amino acid residues to make a β-sheet structure in an environment that permits local interaction with the solvent.

4.3. Higher-Order Structure of Proteins and Contribution of Local Interaction. It is necessary in the field of drug development to understand the biological roles of proteins.[14,15] To understand the functions of proteins, information on the higher-order structures of proteins is indispensable. However, elucidation of the scheme for the formation of a higher-order structure is difficult. According to the results of previous studies, theoretical prediction of a higher-order structure requires elucidation of the involvement of various factors, such as the hydrophobic effect, van der Waals interaction, hydrogen-bond, entropy effect due to conformation change in a side chain, and enthalpy effect.[16−19] In this study, it was demonstrated that molecular orbital calculations are an effective means for determining whether the formation of a helix-like structure or a β-sheet structure is favorable at a particular site of a protein. Previously, some reports suggested that not only short-range interaction but also long-range interaction was important to decide the local secondary structure because there were some sequences that appeared in several proteins commonly but had different conformations in the respective protein.[20,21] For example, Ala63−Ser69 in 1AMP and Ala83−Ser89 in 1GKY are sequentially identical (Ala-Ser-Val-Lys-Gln-Val-Ser), but the structures are different (β-sheet in 1AMP and α-helix in 1GKY).[22,23] Despite the above fact, we still found a possibility to describe the difference of these structures from short-range interaction. The main chain of Ala63−Ser69 in 1AMP forms hydrogen bonds with other neighboring residues, whereas there are no neighboring residues that can form hydrogen bonds with the main chain of Ala83−Ser89 in 1GKY. Accordingly, the β-sheet structure of Ala63−Ser69 in 1AMP could be stabilized without the decrease of the number of hydrogen bonds in the conformation change of helix structure $\rightarrow$ β-sheet structure, but the helix structure of Ala83−Ser89 in 1GKY could not change. Because a whole protein is regarded as an assemblage of local parts, precise calculation of short-range interaction is helpful for predicting higher-order structures of the whole proteins, if the interaction between two local parts of a protein are also carefully examined.

Conclusion

The results obtained from the theoretical calculations in this study are as follows.

(1) A β-sheet structure is dominantly stabilized by local interaction among several residues in the neighborhood and water molecules for the solvation.

(2) The stability of a β-sheet structure can be estimated by the use of potential energy calculation under the condition of the presence of water molecules.

Acknowledgment. We thank the staff of the Research Center for Computational Science in Okazaki. Computations were also carried out using the DRIA system at the Graduate School of Pharmaceutical Science, Chiba University.

References and Notes

(1) Dietmann, S.; Holm, L. *Nat. Struct. Biol.* **2001**, *8*, 953.
(2) Baker, D.; Sali, A. *Science* **2001**, *294*, 93.
(3) Maggio, E. T.; Ramnarayan, K. *Trends Biotechnol.* **2001**, *19*, 266.
(4) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucl. Acids Res.* **2000**, *28*, 235.
(5) http://www.ncbi.nlm.nih.gov/Genbank/GenbankOverview.html.
(6) Kuwata, K.; Hoshino, M.; Era, S.; Batt, C. A.; Goto, Y. *J. Mol. Biol.* **1998**, *283*, 731.
(7) Sheinerman, F. B.; Brooks, C. L., III. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 1562.
(8) Swain, A. L.; Miller, M. M.; Green, J.; Rich, D. H.; Schneider, J.; Kent, S. B.; Wlodawer, A. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 8805.
(9) Smith, C. A.; Rayment, I. *Biochemistry* **1996**, *35*, 5404.
(10) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*, revision A.7; Gaussian, Inc.: Pittsburgh, PA, 1998.
(11) Donald, V.; Judith, G. V. *Biochemistry*, 2nd ed.; PUBLISHER: LOCATION OF PUBLISHER, 1995.
(12) Ferrara, P.; Caflisch, A. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 10780.
(13) Bachmann, A.; Kiefhaber, T. *J. Mol. Biol.* **2001**, *306*, 375.
(14) Rutka, J. T.; Taylor, M.; Mainprize, T.; Langlois, A.; Ivanchuk, S.; Mondal, S.; Dirks, P. *Neurosurgery* **2000**, *46*, 1034.
(15) Harris, T. *Med. Res. Rev.* **2000**, *20*, 203.
(16) Pace, N.; Shirley, B. A.; McNutt, M.; Gajiwala, K. *FASEB J.* **1996**, *10*, 75.
(17) Creamer, T. P. *Proteins* **2000**, *40* 443.
(18) Jaenicke, R. *J Biotechnol.* **2000**, *79* 193.
(19) Edgcomb, S. P.; Murphy, K. P. *Curr. Opin. Biotechnol.* **2000**, *11*, 62.
(20) Daniel, L.; Minor, Jr.; Peter, S, K. *Nature* **1996**, *380*, 730.
(21) Sucha, S. *Proteins* **1998**, *30*, 228.
(22) Chevrier, B.; Schalk, C.; D'Orchymont, H.; Rondeau, J, M.; Moras, D.; Tarnus, C. *Structure* **1994**, *2*, 283.
(23) Stehle, T.; Schulz, G. E. *J. Mol. Biol.* **1992**, *224*,1127.