Molecular Dynamics Simulations of Three-Strand β-Sheet Folding

Hongwu Wang[†,‡] and Shen-Shu Sung*[,†,§]

Contribution from The Lerner Research Institute, The Cleveland Clinic Foundation, Cleveland, Ohio 44195, and Department of Physiology and Biophysics, Case Western Reserve University, Cleveland, Ohio 44106

Received July 7, 1999. Revised Manuscript Received November 15, 1999

Abstract: Traditionally, the empirical force field had great difficulties in simulating β-sheet folding. In the current study, we tested molecular dynamics simulations of β-sheet folding using a solvent-referenced potential. Three available β-sheet-forming synthetic peptides, TWIQNGSTKWYQNGSTKIYT, RGWSVQNGKYT-NNGKTTEGR, and VFITS^{D}PGKTYTEV^{D}PGOKILQ, were simulated at their experimental temperatures. From extended initial conformations, all three peptides folded into β-sheet conformations. The calculated ratios of the β-structure from the 100 ns simulations were 26.5%, 17.8%, and 28.5%, respectively, for the three peptides. From different initial conformations, folding into β-sheets was also observed. With the same energy functions, the alanine-based peptide folded into helical conformations, demonstrating the sequence dependence of folding. During simulations, the β-sheet folding is usually initiated by the fast formation of turns. The three-strand compact structures with favorable inter-strand side-chain interactions occur prior to backbone hydrogen bonding. The conversion of the compact structure to β-sheet is slow, and the peptide spends most of the time in these two states. The attractive side-chain interaction is mainly due to the solvent effect, especially the hydrophobic interactions. Without this solvent effect, β-sheet did not form in the simulations. For the first two sequences, the simulations suggest that the experimentally observed structure may include an ensemble of β-sheet structures. For the ^{D}P-containing peptide, one β-sheet structure with type II′ β-turns is much more stable than other structures.

Introduction

Secondary structure formation has been proposed as one of the early steps in the folding pathway by which a polypeptide chain folds into its native three-dimensional structure.[1−3] Understanding these initial steps is of primary importance to the study of the overall protein folding process as well as to the de novo design of proteins. Despite the importance of β-sheet structures as basic secondary structure elements in proteins, studies of the principles underlying their formation and stability have lagged behind studies of the α-helix, because it is difficult to find soluble β-sheet peptides, whereas short peptides that form monomeric α-helices in solution are readily available.[4,5] It is believed that the low solubility of the β-sheet-forming peptides or their high tendency to aggregate arises from the amphipathic character of the β-structure and the high content of hydrophobic residues, the ones with highest β-sheet propensities.[6−8] Until very recently, the scarce information available on the determinants of β-sheet stability has been obtained from systematic mutagenesis experiments using small engineered proteins that contain a solvent-exposed β-sheet.[9−12]

Despite the difficulties, some linear peptides have recently been shown to fold into monomeric β-hairpins in aqueous solution.[13−16] Furthermore, the design and structure elucidation of the model β-hairpins has demonstrated that the conformation of the turn plays a key role in directing β-hairpin structure, and achieved some rational control of the position and size of the β-hairpin loop.[17−19] Inspired by the observation that β-hairpins with two-residue loops in crystalline protein often have a type I′ or type II′ β-turn,[20,21] investigators have designed short

* To whom all correspondence should be addressed. Phone: (216) 444-0056. Fax: (216) 444-9263. E-mail: sungs@ccf.org.

† The Cleveland Clinic Foundation.

‡ Current address: Dup-1120-07c, Glexo Wellcom Inc., Five More Drive, P.O. Box 13398, Research Triangle Park, NC 27709-3398.

§ Case Western Reserve University.

(1) Kim, P. S.; Baldwin, R. L. *Annu. Rev. Biochem.* **1990**, *59*, 631−660.

(2) Dyson, H. J.; Wright, P. E. *Curr. Opin. Struct. Biol.* **1993**, *3*, 60−65.

(3) Matthews, C. R. *Annu. Rev. Biochem.* **1993**, *62*, 653−683.

(4) Scholtz, J. M.; Baldwin, R. L. *Annu. Rev. Biophys. Biomol. Struct.* **1992**, *21*, 95−118.

(5) Kemp, D. S.; Oslick, S. L.; Allen, T. J. *J. Am. Chem. Soc.* **1996**, *118*, 4249−4250.

(6) Dyson, H. J.; Wright, P. E. *Annu. Rev. Biophys. Biophys. Chem.* **1991**, *20*, 519−538.

(7) Quinn, T. P.; Tweedy, N. B.; Williams, R. W.; Richardson, J. S.; Richardson, D. C. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 8747−8751.

(8) Munoz, V.; Serrano, L. *Proteins* **1994**, *20*, 301−311.

(9) Kim, C. A.; Berg, J. M. *Nature* **1993**, *362*, 267−270.

(10) Smith, C. K.; Withka, J. M.; Regan, L. *Biochemistry* **1994**, *33*, 3, 5510−5517.

(11) Minor, D. L., Jr.; Kim, P. S. *Nature* **1994**, *367*, 660−663.

(12) Blasie, C. A.; Berg, J. M. *Biochemistry* **1997**, *36*, 6218−6222.

(13) Blanco, F. J.; Rivas, G.; Serrano, L. *Nature Struct. Biol.* **1994**, *1*, 584−590.

(14) Ramirez-Alvarado, M.; Blanco, F. J.; Serrano, L. *Nature Struct. Biol.* **1996**, *3*, 604−612.

(15) Searle, M. S.; Zerella, R.; Williams, D. H.; Packman, L. C. *Protein Eng.* **1996**, *9*, 559−565.

(16) Munoz, V.; Thompson, P. A.; Hofrichter, J.; Eaton, W. A. *Nature* **1997**, *390*, 196−199.

(17) de Alba, E.; Jimenez, M. A.; Rico, M. *J. Am. Chem. Soc.* **1997**, *119*, 175−183.

(18) Haque, T. S.; Gellman, S. H. *J. Am. Chem. Soc.* **1997**, *119*, 2303−2304.

(19) Ramirez-Alvarado, M.; Blanco, F. J.; Niemann, H.; Serrano, L. *J. Mol. Biol.* **1997**, *273*, 898−912.

(20) Hutchinson, E. G.; Thornton, J. M. *Protein Sci.* **1994**, *3*, 2207−2216.

(21) Gunasekaran, K.; Ramakrishnan, C.; Balaram, P. *Protein Eng.* **1997**, *10*, 1131−1141.

10.1021/ja992359x CCC: $19.00 © 2000 American Chemical Society
Published on Web 02/17/2000

peptides containing GS, NG, or ^{D}PG (^{D}P stands for D-proline) segments to adopt β-hairpin or β-sheet structures with two-residue turns at these locations.

On the basis of the experimental study and statistical analysis of amino acid β-turn and β-strand propensities, cross-strand side-chain interaction preferences,[22] and the success in design of soluble β-hairpins, several groups have designed peptides that can form a three-strand antiparallel β-sheet motif in aqueous solution. The lengths of these peptides were kept as short as possible while allowing the strands to be long enough to accommodate the interstrand hydrogen bonds and the side-chain interactions needed to stabilize the structure. Different β-turn and β-strand sequences were employed in these synthetic peptides. Kortemme et al.[23] have succeeded in designing a 20-residue peptide, RGWSVQNGKYTNNGKTTEGR, which folds into a stable three-stand antiparallel β-sheet with two NG β-turns. Schenck and Gellman[24] used two ^{D}PG segments as the two β-turn sequences in their 20-residue three-strand antiparallel β-sheet-forming peptide, VFITS^{D}PGKTYTEV^{D}PGOKILQ. De Alba et al.[25] have designed another 20-reisdue peptide, TWIQNG-STKWYQNGSTKIYT, which forms a three-strand antiparallel β-sheet in aqueous solution with two GS β-turns. In addition, Sharman and Searle[26] have described a 24-residue peptide, which forms a β-sheet in 50% methanol solution. These peptides, each of small size and distinctive structure, provide ideal models in refining existing molecular dynamics (MD) protocols, as well as testing recent theoretical approaches to protein folding.

The folding of secondary structures captures much of the basic physics of protein folding, but the extremely fast kinetics in secondary structure folding presents a great challenge in the experimental study of this phenomenon. These events, many of which occur in less than a millisecond, cannot be directly observed by most of the current kinetic methods. Only very recently, new experimental techniques have been put forward, which allow for the observation of protein folding, including secondary structure formation, in a time scale from nanoseconds to microseconds.[27−29] Computer simulations, on the other hand, can provide a detailed picture of these early events.[30,31]

We have previously reported peptide folding simulations using Monte Carlo and MD methods at experimentally relevant temperatures.[32−35] To make long folding simulations accessible with available computing power, water molecules were not included explicitly in our simulations; instead, the average solvent effect was included by using the solvent-referenced potential. The early simulations could not afford to include many solvent molecules.[36,37] Recently, with much greater computing power, MD simulations of peptide folding with explicit water molecules have become available.[38,39] However, most of these simulations were carried out on helical structures.[36−39] Traditionally, the empirical force fields had great difficulties in simulating the folding of β-sheet structures. To our knowledge, there has not been a β-sheet folding simulation with atom-based models. The previous success in the simulation of β-hairpin folding[34,35] prompted us to go one step further toward the folding simulation of the β-sheet. In the current study, the solvent-referenced potential we used previously was tested on the three available β-sheet-forming synthetic peptides.

Methods

The methods used in our study have been described in detail previously.[32,35] Here, we provide a brief outline of their main features. The basic idea is to use the average solvent effect as the reference state for energy calculation without explicitly including water molecules in the simulation. This solvent-referenced potential greatly increases the computational efficiency. It may also reduce the inaccuracy resulting from the cancellation of large energy terms calculated with a vacuum reference state and circumvent the difficulties caused by the multiple minima problem.[40] The competing effect of the interactions with the solvent usually reduces the strength of the interactions among protein atoms and lowers the energy barriers. The solvent-referenced potential represents a simplified and reduced interaction. It includes modifications of the van der Waals (VDW) and the electrostatic interactions, and the treatment of the solvent effect. We tested these treatments on the folding of protein secondary structures with the AMBER force field parameters.[41,42] In principle, the concept of the solvent reference can be applied to other force fields, although detailed treatments may be different.

In solutions, the intramolecular VDW interactions of a protein molecule are balanced by the intermolecular VDW interactions with solvent molecules. The long-range attractive VDW interactions provide a nearly uniform background potential and therefore can serve as the reference for the VDW energy calculation. The possible difference between the protein intramolecular VDW attraction and that with water is included in the solvent effect. The short-range repulsion represents the exclusive volume of each atom and needs to be calculated explicitly. On the basis of these considerations, a shifted truncation at the minimum energy distance is applied to the calculation of VDW interaction energy,[32] as shown in eq 1,

$$E_{\mathrm{VDW}}(r) = \begin{cases} \epsilon\left[\left(\frac{r^*}{r}\right)^{1/2} - 2\left(\frac{r^*}{r}\right)^{6} + 1\right] & r < r^* \\ 0 & r \geq r^* \end{cases} \qquad (1)$$

where r is the distance between two interacting atoms and r^* is the minimum energy distance for the given pair of atoms. This treatment is based on the mean field approximation in the VDW theory of the liquid−solid transition.[43,44] A similar treatment has been previously

(22) Wouters, M. A.; Curmi, P. M. *Proteins* **1995**, *22*, 119−131.
(23) Kortemme, T.; Ramirez-Alvarado, M.; Serrano, L. *Science* **1998**, *281*, 253−256.
(24) Schenck, H. D.; Gellman, S. H. *J. Am. Chem. Soc.* **1998**, *120*, 4869−4870.
(25) de Alba, E.; Santoro, J.; Rico, M.; Jimenez, M. A. *Protein Sci.* **1999**, *8*, 854−865.
(26) Sharman, G. J.; Searle, M. S. *J. Am. Chem. Soc.* **1998**, *120*, 5291−5300.
(27) Williams, S.; Causgrove, T. P.; Gilmanshin, R.; Fang, K. S.; Callender, R. H.; Woodruff, W. H.; Dyer, R. B. *Biochemistry* **1996**, *35*, 691−697.
(28) Ballew, R. M.; Sabelko, J.; Gruebele, M. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 5759−5764.
(29) Eaton, W. A.; Munoz, V.; Thompson, P. A.; Chan, C. K.; Hofrichter, J. *Curr. Opin. Struct. Biol.* **1997**, *7*, 10−14.
(30) van Gunsteren, W. F.; Luque, F. J.; Timms, D.; Torda, A. E. *Annu. Rev. Biophys. Biomol. Struct.* **1994**, *23*, 847−863.
(31) McCammon, J. A.; Harvey, S. C. *Dynamics of Proteins and Nucleic Acids*; Cambridge University Press: Cambridge, England, 1987; pp 1−137.
(32) Sung, S. S. *Biophys. J.* **1994**, *66*, 1796−1803.
(33) Sung, S. S.; Wu, X. W. *Proteins* **1996**, *25*, 202−214.
(34) Sung, S. S. *Biophys. J.* **1999**, *76*, 164−175.
(35) Wang, H.; Varady, J.; Ng, L.; Sung, S. S. *Proteins* **1999**, *37*, *325*−*333*.
(36) Brooks, B. R. *Chem. Scr.* **1989**, *29A*, 165−169.
(37) Caflisch, A.; Karplus, M. In *The protein folding problem and tertiary structure prediction*; Merz, K. M., LeGrant, S. M., Eds.; Birkhauser: Boston, 1994; pp 193−230.
(38) Daura, X.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E. *J. Mol. Biol.* **1998**, *280*, 925−932.
(39) Duan, Y.; Kollman, P. A. *Science* **1998**, *282*, 740−744.
(40) Ripoll, D. R.; Scheraga, H. A. *Biopolymers* **1988**, *27*, 1283−1303.
(41) 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.
(42) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(43) Widom, B. *Science* **1967**, *157*, 375−382.
(44) Chandler, D.; Weeks, J. D.; Andersen, H. C. *Science* **1983**, *220*, 787−794.

applied to protein structure studies.[45] However, when a peptide molecule is big and compact, solvent molecules cannot readily fill in any internal space of the peptide. Completely truncating the attractive part of the VDW interaction may result in unreasonable empty spaces inside a compact structure. Therefore, a truncation at a larger distance is sometimes applied.[46] In the current study, the peptide molecule forms a three-strand β-sheet, which has a larger buried portion than a β-hairpin, and a truncation at $r = 1.25r^*$ is also tested.

Similarly, the electrostatic interaction of the in vacuo calculation needs modification to account for the solvent effect. As a widely applied approximation, different effective dielectric constant values as well as various forms of distance-dependent functions have been used in MD simulations.[47−49] In our previous studies, the electrostatic interaction was consistently scaled down by a factor of 2. A scaled distance-dependent function has been tested for α-helix folding,[32,33] and a constant value 2 has been tested for both α-helix and β-hairpin folding.[34,35] For simple secondary structure elements without charged side-chains, either a distance-dependent function or a constant value within a reasonable range can be used to simulate α-helix and β-sheet folding. In the current study, a dielectric constant 2 was used to account for the solvent effect on intramolecular electrostatic interactions among protein atoms.

However, in a real heterogeneous protein−solvent system, a single dielectric constant cannot exactly account for the complicated solvent effect on electrostatic interactions. For a system as simple as a peptide with charged side-chains, the calculated electrostatic interactions between charged side-chains and other peptide groups, such as the attraction between a lysine side-chain and the backbone carbonyl, are often so strong that it traps the structure in an unrealistic conformation. Experimentally, the charged side-chains tend to be surrounded by water molecules and a much higher dielectric constant, such as 80 for water, needs to be applied for the side-chain. Or, it may have a close interaction with a counterion and its charge is largely neutralized. Therefore, using a neutral side-chain could be a reasonable option. The AMBER force field[42] provides such an option that side-chain atomic charges are adjusted to neutralize the total change of the amino acid. In the current study, instead of using different values of the dielectric constant for charged side-chains, neutral side-chains for lysine, arginine, glutamate, and ornithine residues were used to reduce the unrealistic intramolecular interactions. Similar treatment has been used previously.[50]

In the implicit solvent approach, the solvent effect is often assumed to be proportional to the solvent-accessible surface area (SASA).[51,52] In the current study, the solvation free energy, ΔG, was calculated as the sum of the contribution from each atom according to its exposure to the solvent, as shown in eq 2. The contribution from an atom i is assumed to be the product of an atomic solvation parameter $\Delta\sigma_i$ and the solvent-accessible surface area A_i.[51,53]

$$\Delta G = \sum_i \Delta\sigma_i A_i \qquad (2)$$

MD simulations with surface area-based solvation energy have been carried out before.[53,54] On the basis of the free energy of transfer between water and octanol,[55] we obtained a set of solvation parameters in which the protein atoms were categorized into six types: C, O, N, O^-, N^+, and S.[33] In the current study, similar parameters with some modification were used. The atomic solvation parameters used were 25 cal mol^{-1} $Å^{-2}$ for C, −9 cal mol^{-1} $Å^{-2}$ for N and O, and 5 cal mol^{-1} $Å^{-2}$ for S. The atomic solvation parameters for O^- and N^+ were not used in this study.

In the current study, MD simulations were carried out using AMBER software.[41,42] The electrostatic interaction was cut off at 20 Å without using extra switching functions. The time step used in the simulations was 0.002 ps with the bond lengths involving hydrogen atoms fixed by the SHAKE procedure.[56] The AMBER software we obtained does not contain the SASA calculation. The algorithm of the TINKER software[57] for calculating the surface area and its derivative was used. In this algorithm, the derivative of the surface area was calculated analytically for evaluating the force resulting from the surface area change. The surface area-based calculation is more efficient than including water molecules explicitly, but it still increases computing time by about 15-fold compared with the calculations that do not include the solvent effect. However, the solvation energy change at each step is more than an order of magnitude smaller than the fastest changing energy term (usually, the bond vibration energies). To speed up the simulation, a larger time step (0.02 ps) was used to calculate the surface area and its derivative. The force resulting from the solvation was kept constant within each of these large time steps. In our previous study of helix folding,[33] the results with the large time step were compared with those with 0.002 ps time step for the surface area calculation. The structures observed and the folding time were similar during the simulations. The larger time step (0.02 ps) for the solvent effect calculation speeds up the simulation by about 6-fold. Multiple time steps have been widely used in MD simulations to increase computational efficiency[58] and more elaborate methods have been developed.[59] Here, we simply increased the time step to 0.02 ps for the solvent effect calculation and numerical instabilities were not encountered.

The content of β-sheet and β-hairpin conformation, or the β-ratio, in the peptide conformations was analyzed by the DSSP program.[60] Peptide coordinates were recorded every 0.01 ns during the simulation, and the secondary structure analysis was carried out on each recorded conformation along the simulation trajectory. The overall β-ratio was calculated by counting the total number of residues participating in the β-turns and the β-strands divided by the number of residues in all recorded conformations. Residues in the loop region between two β-strands that did not form β-turns were treated as not having a β structure. All figures in this article were created by the molecular modeling software InsightII and the graphic software SigmaPlot. All simulations and analysis were carried out on an SGI O2 workstation.

Results and Discussion

Folding Simulation of Ac-TWIQNGSTKWYQNGSTKIYT-NH_2. De Alba et al.[25] showed that the peptide TWIQNGSTKWYQNGSTKIYT (GS hereafter) forms a significant population of monomeric three-strand antiparallel β-sheet in aqueous solution. The 20-residue peptide was designed to be able to form two type II′ β-turns at the two GS locations, with the other parts forming three β-strands. Turn residues GS were selected according to their high statistical probability of being at positions $i + 1$ and $i + 2$ of a type II′ β-turn. Residues with high β-sheet propensity, such as I, Y, W, and T, were used for the β-strand sequences. Two lysine residues, which have low β-sheet propensity, were included to increase peptide solubility and prevent aggregation. Nuclear magnetic resonance (NMR) data obtained at 283 K are consistent with the presence of a single structure form, the three-strand β-sheet, in equilibrium with the coil conformations. The β-sheet population was 13−31% based on $C_\alpha H_i - C_\alpha H_j$ NOE intensity, 30−55% based on the chemical shift of $C_\alpha H$ protons.

(45) McCammon, J. A.; Northrup, S. H.; Karplus, M.; Levy, R. M. *Biopolymers* **1980**, *19*, 2033−2045.
(46) Rigby, D.; Roe, R. J. *J. Chem. Phys.* **1987**, *87*, 7285−7292.
(47) Okamoto, Y. *Biopolymers* **1994**, *34*, 529−539.
(48) Hingerty, B. E.; Ritchie, R. H.; Ferrell, T. L.; Turner, J. E. *Biopolymers* **1985**, *24*, 427−439.
(49) Daggett, V.; Kollman, P. A.; Kuntz, I. D. *Biopolymers* **1991**, *31*, 285−304.
(50) Lazarids, T.; Karplus, M. *Proteins* **1999**, *35*, 133−152.
(51) Eisenberg, D.; McLachlan, A. D. *Nature* **1986**, *319*, 199−203.
(52) Schiffer, C. A.; Caldwell, J. W.; Stroud, R. M.; Kollman, P. A. *Protein Sci.* **1992**, *1*, 396−400.
(53) Wesson, L.; Eisenberg, D. *Protein Sci.* **1992**, *1*, 227−235.
(54) Schiffer, C. A.; Caldwell, J. W.; Kollman, P. A.; Stoud, R. M. *Mol. Simulation* **1993**, *10*, 121−149.
(55) Radzicka, A.; Wolfenden, R. *Biochemistry* **1988**, *27*, 1664−1670.
(56) Ryckaert, J. P.; Ciccott, G.; Berendsen, H. J. C. *J. Comput. Chem.* **1976**, *23*, 327−341.
(57) Kundrot, C. E.; Ponder, J. W.; Richards, F. M. *J. Comput. Chem.* **1991**, *12*, 402−409.
(58) Allen, M. P.; Tildesley, D. J. *Computer simulation of liquids*; Clarendon Press: Oxford, 1987.
(59) Tuckerman, M.; Berne, B. J. *J. Chem. Phys.* **1992**, *97*, 1990−2001.
(60) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577−2637.

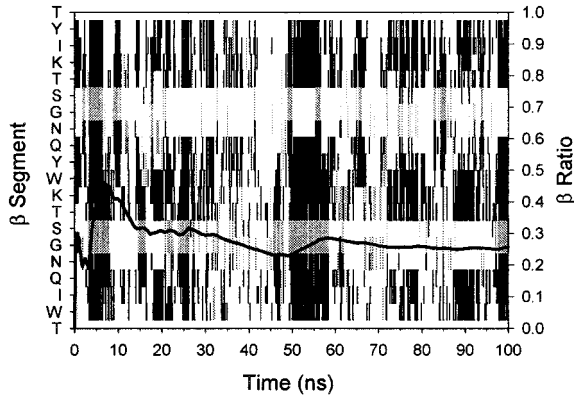

Figure 1. Location and time of occurrence of β-sheet (and β-hairpin) structures during the simulation of peptide GS. The solid vertical lines represent the β-strands and the dotted vertical lines the β-turns. The thick curve shows the accumulated β-ratio.

The MD simulation of GS was carried out for 100 ns at the experimental temperature of 283 K. The simulation started with a standard extended conformation, with the backbone ϕ, ψ dihedral angles assigned to 180° for all residues. A short β-hairpin formed rapidly within 0.2 ns at the C-terminal part of the peptide. This β-hairpin conformation has a three-residue turn at the C-terminal NGS location and four interstrand hydrogen bonds. This β-hairpin unfolded after about 1 ns. Another C-terminal β-hairpin, which has a type II′ β-turn at the C-terminal GS location and four interstrand hydrogen bonds, folded around 1.8 ns. This C-terminal β-hairpin unfolded at 2.4 ns. At about 3.2 ns, a short N-terminal β-hairpin with a type II′ β-turn at the N-terminal GS location formed; this β-hairpin quickly developed into an antiparallel three-strand β-sheet 0.2 ns later. This three-strand β-sheet conformation has two type II′ β-turns at the two GS locations. The β-sheet was stable for about 3 ns, until the C-terminal β-strand unfolded at 6.4 ns. The N-terminal β-hairpin remained folded for another 1.5 ns before the peptide unfolded completely at 8 ns. This three-strand β-sheet conformation folded again during 49−50 and 55−56.5 ns. The peptide adopted various conformations with some β-hairpin contents after 8 ns, before it folded into another three-strand β-sheet conformation at about 18 ns. This new β-sheet conformation has a three-residue turn at the N-terminal NGS location and a type I′ β-turn at the C-terminal NG location. It unfolded shortly before 20 ns. This β-sheet conformation folded again during 72−79 and 88−92 ns. In this β-sheet conformation, the C-terminal NG β-turn was not stable and often unfolded into a four-residue turn at the C-terminal QNGS location. Another stable three-strand β-sheet conformation, which was first observed around 25 ns, has a type II′ β-turn at the N-terminal GS location and a three-residue turn at the C-terminal NGS location. This β-sheet conformation folded during 25−26, 51−55, and 97.5−100 ns.

The overall percentage of the β structure (β-ratio) in this 100-ns simulation is 26.5%, as calculated by the DSSP algorithm. The β-ratio agrees quite well with the experimental estimations of the β-sheet population for this peptide.[25] Figure 1 shows the location of the three-strand β-sheet and the β-hairpin conformations in every 0.1 ns during the simulation, as well as the accumulated β-ratio. The accumulated β-ratio is drawn as a thick curve in the figure. The β-hairpin and β-sheet conformations are represented by vertical lines, with the solid lines for the β-strands and the dotted lines for β-turns. The accumulated β-ratio at time t is the average percentage of the β structure over the period of time from 0 to t. In the beginning of the simulation, because the period of time for averaging is shorter, a structural change has a more apparent effect on the shape of the accumulated β-ratio curve than in the later stage of the simulation. Hence, a sharp peak is often observed, corresponding to a transient β structure. The folding process that occurred between 3.1 and 3.4 ns is illustrated in Figure 2. At 3.1 ns, the peptide adopted a coil conformation. At 3.2 ns, a type II′ β-turn formed at the N-terminal GS location. A complete N-terminal β-hairpin formed at 3.25 ns, and the C-terminal part of the peptide adopted a U-shaped conformation. With more interstrand hydrogen bonds formed in the C-terminal, the peptide folded into a complete three-strand, antiparallel β-sheet conformation with two type II′ β-turns at the two GS locations at 3.39 ns.

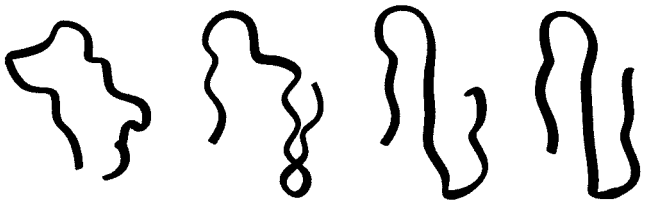

Figure 2. Conformations illustrating a β-sheet folding event. The conformations of peptide GS at 3.10, 3.20, 3.25, and 3.39 ns are shown from left to right. The backbone of the peptide is shown as ribbons; side-chain atoms are not shown. The N-terminus of the peptide is located at the lower left of each structure.

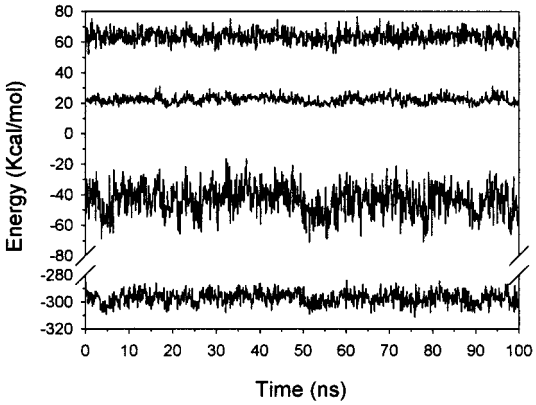

Figure 3. Energy changes during the simulation of peptide GS. The top curve shows the VDW energy, the second curve from the top the solvation energy, the third curve from the top the total potential energy, and the bottom curve the electrostatic energy.

The important energies were recorded every 0.01 ns during the 100-ns simulation, as shown in Figure 3. The top curve in this figure shows the VDW energy, the second the solvation energy, the third the total potential energy, and the bottom the electrostatic energy. Upon the folding of β-sheet structures, all energy terms decreased. The average potential energy for two β-sheet regions (3.5−6.0 and 50.5−53.0 ns; 500 structures) is 14.5 kcal/mol lower than the average potential energy of three coil regions (38.9−39.9, 42.0−45.0, and 69.0−70.0 ns; 500 structures). The average solvation energy, electrostatic energy, and VDW energy for the β-sheet regions is 3.6, 7.9, and 1.2 kcal/mol lower than that of the coil regions, respectively. The contributions to the potential energy difference from covalent bonding energies are smaller in value, totaling about 1.8 kcal/mol. Compared with coil conformation, the lower electrostatic energy of β-sheet comes from the large number of hydrogen bonds formed by backbone atoms. The lower solvation energy and the lower VDW interaction energy resulted from the better side-chain packing in the β-sheet.

During the simulation, four typical three-strand β-sheet conformations along with two typical β-hairpin conformations were observed (Figure 4). Of the four β-sheets, the first one

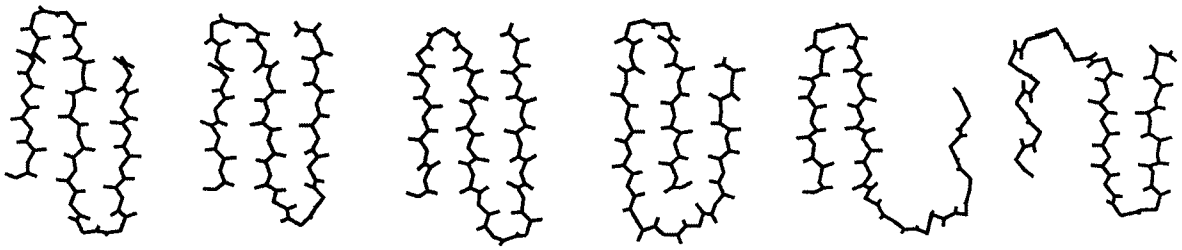

Figure 4. Different types of β-sheet and β-hairpin conformations observed during the folding simulation of peptide GS. From left to right, a three-strand β-sheet with two type II′ β-turn at the two GS locations (5 ns), a three-strand β-sheet with a type II′ β-turn at the N-terminal GS location and a three-residue turn at the C-terminal NGS location (25 ns), a three-strand β-sheet with a three-residue turn at the N-terminal NGS location and a type II′ β-turn at the C-terminal GS location (88 ns), a three-strand β-sheet with the N-terminal β-strand as the central β-strand (57 ns), an N-terminal β-hairpin (47 ns), and a C-terminal β-hairpin (10 ns). Side-chain atoms are not shown. The N-terminus is located at the lower part of each structure.

has two type II′ β-turns at the two GS locations with six hydrogen bonds formed between each edge strand and the central strand. The second β-sheet has a type II′ β-turn at the N-terminal GS location and a three-residue turn at the C-terminal NGS location. The N-terminal β-hairpin has six interstrand hydrogen bonds, whereas the C-terminal β-hairpin has five. The third β-sheet has a three-residue turn at the N-terminal NGS location and a type I′ β-turn at the C-terminal NG location with the N-terminal β-hairpin having five interstrand hydrogen bonds and the C-terminal β-hairpin having six. The fourth β-sheet is a mixed parallel−antiparallel β-sheet with the N-terminal β-strand as its central strand. This β-sheet conformation was only briefly observed, at 57.4−57.8 ns in the simulation. With only two β-strands formed, the peptide can fold into various β-hairpin conformations. The first β-hairpin conformation in Figure 4 has a type II′ β-turn at the N-terminal GS location and six interstrand hydrogen bonds. The N-terminal 2/3 of the peptide formed a β-hairpin conformation, and the C-terminal 1/3 adopted a coil conformation. The second β-hairpin conformation in Figure 4 has a type II′ β-turn at the C-terminal GS location and five interstrand hydrogen bonds. The C-terminal 2/3 of the peptide formed a β-hairpin conformation, and the N-terminal 1/3 adopted a coil conformation. Other β-hairpin conformations with different turn and strand configurations were also observed during the simulation. All β-sheet and β-hairpin conformations had a right-handed twist characteristic, which is consistent with experimental structures of the β-sheet in proteins.[61] All of these conformations have lower potential energies compared with the coil structures.

Different turn configurations have been found for the three-strand β-sheet conformations. Type II′ β-turns at the two GS locations were frequently observed in the three-strand β-sheet structures. Of these two GS turns, the N-terminal β-turn is more stable compared with the C-terminal GS turn. Three-residue turns at the two NGS locations were also observed in the β-sheet conformations. Even though GS were designed to form the two-residue β-turn, the NG segment also has a high statistical probability of forming a type I′ β-turn.[20] A type I′ β-turn at the C-terminal NG formed in one of the typical β-sheet conformations, but it usually unfolded into a four-residue turn. A type I′ β-turn at the N-terminal NG was not observed in the β-sheet structures, though it was observed in some of the β-hairpin structures. Different configurations of the turn regions in the β-sheet structures were a result of the competing effects of forming a type II′ β-turn at GS and a type I′ β-turn at NG. The type I′ β-turn suits the β-hairpin conformation better than the type II′ β-turn due to their larger right-handed twist. The type II′ β-turn is more planar, allowing better backbone interstrand hydrogen bonding. Type II′ β-turns at the GS locations were more stable than type I′ NG β-turns because for small β-sheet structures, a stable hydrogen bond network is needed. Although NMR data are compatible with the presence of a single three-strand β-sheet form, linear peptides in solution commonly exist as conformational ensembles of fast interconverting structures. This is the reason that multiple β-sheet structures were observed in our simulation.

Table 1. Compatibility of Different β-Sheet Conformations of Peptide GS with Experimentally Observable Interstrand NOEs

residue I	residue J	NOE intensity[a]	β-sheet 1[b]	2	3	4
CαH W2	CαH Y11	m-s	Y	Y	Y	Y
Cε3H W2	CαH Y11	w	Y	Y	Y	Y
CαH Q4	CαH K9	s	Y	Y	Y	Y
CαH W10	CαH Y19	m-s	Y	Y	N	N
CαH Q12	CαH K17	m-s	Y	Y	N	N
CαH T1	Cε3H Y11	w	Y	Y	Y	Y
Cββ′H W2	Cε3H W10	vw	N	N	Y	N
Cε3H W2	Cβ′H N13	w	N	Y	N	N
Cγ3H/Cδ3H I3	Cε′3H W10	w	Y	Y	N	Y
Cγ3H/Cδ3H I3	Cδ1H W10	m-w	Y	Y	N	Y
CδH Y11	Cγ′H I18	vw	Y	N	Y	N
CδH Y11	CγH I18	w	Y	N	Y	N
CδH Y11	Cγ3H I18	m	Y	N	Y	N

[a] The intensities of the NOEs are classified as follows: s, strong; m, medium; m-s, intermediate between strong and medium; w, weak; m-w, intermediate between weak and medium; vw, very weak. [b] Y indicates that the distance between the atom pair is small enough in the structure for NOE to be observable. N indicates that the distance between the atom pair is too large in the structure for NOE to be observable.

In the three-strand β-sheet conformations, the side-chains were separated into two groups along a plane formed by the backbone atoms. In the first β-sheet in Figure 4, the hydrophobic side-chain atoms of I3 of the first β-strand, T8 and W10 of the second β-strand, and Y19 of the third β-strand packed together to form a hydrophobic cluster along one side of the backbone plane. On the other side, the hydrophobic side-chain atoms of W2 of the first β-strand, K9 and Y11 of the second β-strand, and I18 of the third strand packed together to form the hydrophobic cluster. The pairing and side-chain packing pattern in this β-sheet conformation agree with all the NMR findings except for two weak NOEs indicating side-chain interactions in W2···W10, and W2···N13 pairs, as listed in Table 1. The side-chains of W2 and W10 sit on different sides of the backbone plane in this conformation, making it impossible for any side-chain interactions between these two residues. Even though their side-chains are aligned on the same side of the backbone plane, the distance between W2 and N13 is too large to have any strong side-chain interaction between them.

(61) Chothia, C. *J. Mol. Biol.* **1973**, *75*, 295−302.

In the second β-sheet in Figure 4, the hydrophobic side-chain atoms of I3, W10, T16, and I18 form a hydrophobic cluster along one side of the backbone plane. The hydrophobic side-chain atoms of W2, K9, Y11, and Y19 form a hydrophobic cluster on the other side. The pairing and side-chain packing pattern in this β-sheet conformation satisfy some of the NOE constrains, including the signal indicating side-chain interactions between W2 and N13. Here, W2 and N13 are close enough to have strong side-chain interaction between them. But, Y11 and I18 are on different sides of the backbone plane, inconsistent with the observed NOEs between the side-chain atoms of these two residues.

In the third β-sheet in Figure 4, the hydrophobic side-chain atoms of I3, K9, Y11, and I18 form a hydrophobic cluster on one side of the backbone plane. The hydrophobic side-chain atoms of W2, W10, K17, and Y19 form a hydrophobic cluster on the other side. The side-chains of W2 and W10 are very close in this conformation, satisfying the NOE constraint observed between these two residues. But the pairing and side-chain packing pattern does not satisfy most of the other NOE constrains as listed in Table 1. For the β-sheet with the N-terminal strand as the central strand, the hydrophobic side-chain atoms of I3, W10, and Y19 form a hydrophobic cluster along one side of the backbone plane. The hydrophobic side-chain atoms of W2, K9, Y11, and I18 form a hydrophobic cluster on the other side. The pairing and side-chain packing pattern do not satisfy most of the interstrand NOEs. As none of the β-sheet conformations satisfied all the observed NOEs, a likely explanation is that the peptide exists in three different β-sheet conformations in aqueous solution, with the most populated form being the first β-sheet. The second and third β-sheet were present in a much smaller population as indicated by the weak NOEs of W2···W10 and W2···N13 interactions.

Folding Simulation of Ac-RGWSVQNGKYTNNGKT-TEGR-NH_2. Using a four-residue per strand and two-residue β-turns structural template and an iterative hierarchical approach, Kortemme et al.[23] have successfully designed a 20-amino acid, three-stranded β-sheet peptide, RGWSVQNGKYTNNGKT-TEGR (NG hereafter), based on a designed de novo β-hairpin peptide.[14] In the selection of the peptide sequence, experimental information on β-hairpin stability, amino acid β-sheet propensities, and statistical preferences for interstrand residue pairing have been considered. The turn sequences were selected to be optimal for type I′ β-turns.[19] Two arginine residues at both ends of the peptide were introduced to increase its solubility and prevent aggregation. NMR spectroscopy at 273 K provides strong evidence that the 20-residue peptide forms a monomeric, three-stranded, antiparallel β-sheet in aqueous solution, with the NG residues forming the two β-turns.

The simulation of NG was carried out for 100 ns at the experimental temperature of 273 K. The simulation started with a standard extended conformation. A β-hairpin formed at the N-terminal part of the peptide at about 0.1 ns. This β-hairpin has a three-residue turn at the N-terminal NGK location and four interstrand hydrogen bonds. This β-hairpin unfolded after about 0.5 ns. Another N-terminal β-hairpin, which has a type II′ β-turn at the N-terminal GK location and four interstrand hydrogen bonds, and a C-terminal β-hairpin, which has a three-residue turn at the C-terminal NGK location and four interstrand hydrogen bonds, folded occasionally during the first 5-ns simulation. The C-terminal β-hairpin with three-residue turn folded again at 5 ns and developed into a three-strand β-sheet conformation at around 5.8 ns. The three-strand β-sheet has two three-residue turns at the two NGK locations, and it unfolded around 6 ns. The peptide adopted various conformations with some β-hairpin contents before it folded into a very stable β-hairpin conformation at about 44 ns. This β-hairpin has a β-turn at the KY location with up to eight interstrand hydrogen bonds. It unfolded shortly before 53 ns. The three-strand β-sheet was observed briefly again between 66 and 67 ns. The most stable three-strand β-sheet region in this 100-ns simulation occurred between 77 and 79 ns as the peptide folded for about 2 ns.

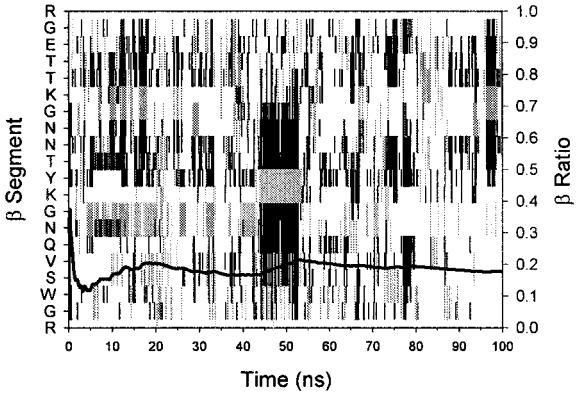

Figure 5. Location and time of occurrence of β-sheet (and β-hairpin) structures during the simulation of peptide NG. The solid vertical lines represent the β-strands and the dotted vertical lines the β-turns. The thick curve shows the accumulated β-ratio.

The overall β-ratio in this 100-ns simulation is 17.8%, as calculated by the DSSP algorithm. Figure 5 shows the location of the β-hairpin and β-sheet conformations every 0.1 ns during the simulation, as well as the accumulated β-ratio. The β-ratio is mostly contributed by the β-hairpin content in different conformations. Three-strand β-sheet conformations were observed during the simulation, but none of them were very stable. Unlike the three-strand β-sheet structure determined by the NMR experiment, which has two type I′ β-turns at the two NG locations, the most stable three-strand β-sheet observed in our simulation has two three-residue turns at the two NGK locations.

Upon the folding of β-sheet and β-hairpin structures, all energy terms decreased compared with those of the coil conformations. The average potential energy for one stable β-sheet region (77.0−78.0 ns; 100 structures) is 8.7 kcal/mol lower than the average potential energy of three coil regions (27.0−28.2, 33.8−36.0, and 84.6−86.2 ns; 500 structures). The average solvation energy, electrostatic energy, and VDW energy for the β-sheet regions is 1.1, 4.9, and 2.1 kcal/mol lower than that of the coil regions, respectively. The contributions to the potential energy difference from covalent bonding energies are smaller in value, totaling about 0.7 kcal/mol. The β-sheet conformations in this simulation are much less stable compared with those of peptide GS. The smaller value in the potential energy difference between β-sheet and coil conformations could be an indication of the instability of the β-sheet structures for this peptide.

During the simulation, three types of the three-strand β-sheet were observed (Figure 6). Of the three typical β-sheet conformations, the first one has two three-residue turns at the two NGK locations with five hydrogen bonds formed between each edge strand and the central strand. The second β-sheet has a β-turn at the QN location and a six-residue loop at the NNGKTT location. The N-terminal β-hairpin has five interstrand hydrogen bonds, and the C-terminal four. Unlike the previous two β-sheet conformations, the third β-sheet uses the C-terminal β-strand as the central strand. The small side-chains of the C-terminal

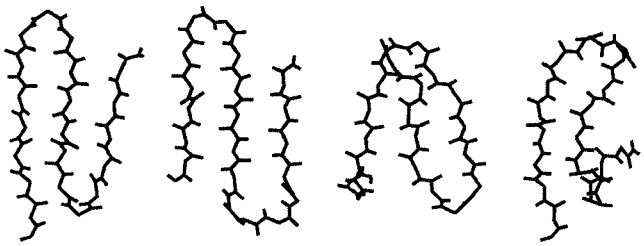

Figure 6. Different types of β-sheet and β-hairpin conformations observed during the folding simulation of peptide NG. From left to right, a three-strand β-sheet with two three-residue turns at the two NGK locations (77 ns), a three-strand β-sheet with a β-turn at the QN location and a six-residue loop at the NNGKTT location (66 ns), a three-strand β-sheet with the C-terminal β-strand as the central strand (37 ns), and a β-hairpin with a β-turn at the KY location (45 ns). Side-chain atoms are not shown. The N-terminus is located at the lower left of each structure.

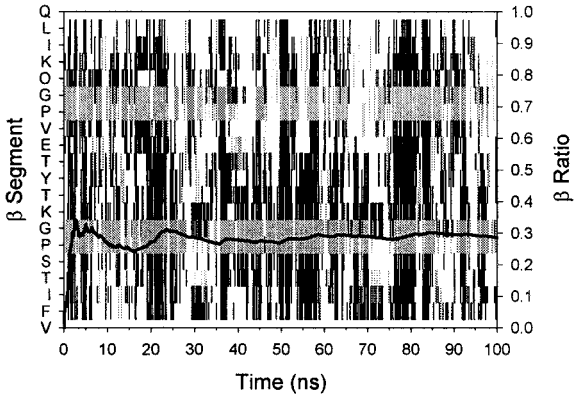

Figure 7. Location and time of occurrence of β-sheet (and β-hairpin) structures during the simulation of peptide PG. The solid vertical lines represent the β-strands and the dotted vertical lines the β-turns. The thick curve shows the accumulated β-ratio.

β-strand fit well between the other two β-strands. With only two β-strands formed, the peptide can also fold into different β-hairpin conformations. The β-hairpin formed by the N-terminal 2/3 of the peptide usually has a three-residue turn at the N-terminal NGK location, whereas the β-hairpin formed by the C-terminal 2/3 usually has a type II′ β-turn at the C-terminal GK location or a three-residue turn at the C-terminal NGK location. C-Terminal β-hairpin conformations were more frequently observed than N-terminal ones during the simulation. Figure 6 also shows the structure of a stable long β-hairpin observed during the simulation, which has a KY β-turn and eight interstrand hydrogen bonds. This long β-hairpin structure appeared to be a result of the side-chain interactions. During the folding process, the side-chain of N7 formed hydrogen bonds with C═O of K9 and NH of T11, and the side-chain of T11 formed a hydrogen bond with NH of G9. These hydrogen bonds helped the formation of a turn at the KY location. The side-chains of N7 and T11 formed a hydrogen bond between them in the β-hairpin conformation, which stabilizes the β-hairpin structure. All β-sheet and β-hairpin conformations had a right-handed twist characteristic and lower potential energies compared with the coil structures.

Folding Simulation of Ac-VFITS^{D}PGKTYTEV^{D}PGOKILQ-NH$_2$. Peptide VFITS^{D}PGKTYTEV^{D}PGOKILQ (PG hereafter) was found[24] to adopt a three-strand antiparallel β-sheet conformation in aqueous solution. Site-specific conformational data from NMR spectroscopy provide strong evidence that high populations of the three-strand β-sheet are present in aqueous solution at 297 K. PG also displays a β-sheet signature in circular dichroism measurements. This peptide incorporated three unnatural amino acids (two D-prolines and one ornithine). The constraint of the pyrrolidine ring of ^{D}P restricts its backbone ϕ torsion angle to a value close to the ideal ϕ_{i+1} value of $+60°$ in a type I′ or type II′ β-turn, promoting the formation of two β-turns at the two ^{D}PG locations. It is intended that the formation of these two β-turns will further promote the formation of two β-hairpins with one strand in common. Ornithine has a side-chain similar to that of lysine, but one CH_2 unit shorter.

The simulation of PG was carried out for 100 ns at the experimental temperature of 297 K. The simulation started with an extended conformation generated by assigning the backbone ϕ, ψ dihedral angles to 180° for all residues except the two D-prolines, whose ϕ and ψ dihedral angles were set to 60° and 180°, respectively. Within 0.2 ns of the simulation, various β-turns formed at several locations along the peptide sequence, including the two ^{D}PG locations, as the peptide system searched for low-energy conformations. The β-turns formed at ^{D}PG were always type II′. A β-turn located at the C-terminal ^{D}PG sequence developed into a short β-hairpin with three hydrogen bonds at around 0.3 ns. This first short β-hairpin was not stable and unfolded very quickly. A three-strand β-sheet conformation first formed at 0.9 ns, with two β-turns at the two ^{D}PG locations and less than three hydrogen bonds between the central β-strand and the edge β-strands. The N-terminal β-hairpin of this three-strand β-sheet unfolded at about 1 ns, while the C-terminal β-hairpin remain folded, and the peptide refolded into a three-strand β-sheet structure at 1.4 ns. The new β-sheet conformation has more interstrand hydrogen bonds and is more stable, which unfolded at about 2.0 ns. The peptide underwent several folding and unfolding processes in the first 20-ns simulation, during which three-strand β-sheets with different number of interstrand hydrogen bonds and various β-hairpin conformations with different loop configurations have been observed. Shortly after 20 ns, the peptide refolded into the three-strand β-sheet again; this time the β-sheet conformation was very stable, lasting about 3 ns. The most stable β-sheet region observed during the simulation occurred between 76 and 81 ns, and the peptide stayed in this conformation for about 5 ns.

The overall β-ratio in this 100-ns simulation is 28.5%, as calculated by the DSSP program. Figure 7 shows the location of the β-hairpin and β-sheet conformations every 0.1 ns during the simulation, as well as the accumulated β-ratio. This peptide has the most stable turn regions, especially the N-terminal β-turn, and the highest β-ratio among the three β-sheet-forming peptides. The average potential energy for three β-sheet regions (21.2−23.0, 78.2−80.2, and 83.0−84.2 ns; 500 structures) is 13.5 kcal/mol lower than the average potential energy of three coil regions (14.5−16.3, 27.1−28.3, and 98.0−100.0 ns; 500 structures). The average solvation energy, electrostatic energy, and VDW energy for the β-sheet regions is 1.9, 6.3, and 2.5 kcal/mol lower than that of the coil regions, respectively. The contribution to the potential energy difference from covalent bonding energies is 2.8 kcal/mol.

During the simulation, three typical three-strand β-sheet conformations were observed (Figure 8). Of the three typical β-sheet conformations, the first one has two type II′ β-turns at the two ^{D}PG locations, with five hydrogen bonds formed between each edge strand and the central strand. This β-sheet is the most frequently observed β-sheet conformation during the 100-ns simulation. In this structure, the hydrophobic side-chain atoms of I3 of the first β-strand, K8 and Y10 of the second β-strand, and K17 and L19 of the third β-strand packed together to form a hydrophobic cluster along one side of the backbone

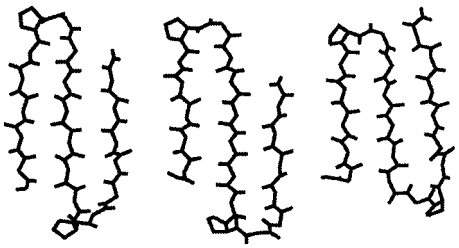

Figure 8. Different types of β-sheet conformations observed during the folding simulation of peptide PG. From left to right, a three-strand β-sheet with two type II′ β-turns at the two ^{D}PG locations (21 ns), a three-strand β-sheet with a type II′ β-turn at the N-terminal ^{D}PG location and a three-residue turn at the ^{D}PGO location (37 ns), and a three-strand β-sheet with a type II′ β-turn at the N-terminal ^{D}PG location and a four-residue turn at the EV^{D}PG location (5 ns). Side-chain atoms except those of the ^{D}P's are not shown. The N-terminal ^{D}P is located at the upper right of each structure.

plane. On the other side, the hydrophobic side-chain atoms of F2 of the first β-strand and T9 and T11 of the second β-strand packed together, while T11 and V13 of the second β-strand packed with O16 and I18 of the third strand, to form two hydrophobic clusters. The side-chain packing pattern in this β-sheet conformation agrees with all the observed side-chain NOEs, i.e. between residues Y10···L19, Y10···K17, and F2···T11.

The other two three-strand β-sheets have identical N-terminal β-hairpins as the first β-sheet, but with different C-terminal conformations. The second β-sheet has a three-residue C-terminal loop with a type II′ β-turn at ^{D}PG and a bulge at O16. The C-terminal β-hairpin, which has only four interstrand hydrogen bonds, is shorter than the N-terminal β-hairpin. The third β-sheet has a four-residue C-terminal loop at the EV^{D}PG location. The C-terminal β-hairpin has the same length as the N-terminal β-hairpin, with five interstrand hydrogen bonds. The second and third β-sheets are much less stable than the first β-sheet, and their side-chain packing patterns are different from those of the NMR findings. In the β-sheet conformations, the C-terminal β-hairpin is less stable than the N-terminal β-hairpin, as shown by the single N-terminal β-hairpin configuration and multiple C-terminal β-hairpin configurations. The reason for this difference is not clear. Stable β-hairpin conformations are also observed, with most of them having a β-turn at the N-terminal or the C-terminal ^{D}PG locations. All β-sheet and β-hairpin conformations had a right-handed twist characteristic and lower potential energies compared with the coil structures.

The Effect of the Turn Residues and the Side-Chain Interactions. The β-sheet folding is usually initiated by the formation of turns. In the folding process, the peptide backbone always forms bends at the sequences with high β-turn propensity, such as GS, NG, GK, and ^{D}PG in the three β-sheet-forming peptides. The bends at these locations are usually preserved even in the coil structures, where the backbone reverses its direction. Among these peptides, the three-strand β-sheet formed by peptide PG is the most stable one, followed by GS and NG. In peptide PG, the backbone usually forms two bends at the two ^{D}PG locations, and the conformation of the bends is usually that of a type II′ β-turn, even in the coil conformations. The dominant preference of forming β-turn at the ^{D}PG location greatly restricted the conformational freedom of the peptide backbone. The GK segment in the peptide also has a tendency of forming β-turn, and is observed in some less populated β-sheet structures.

With peptide GS, the peptide backbone usually forms two bends at the two GS locations. Unlike ^{D}PG, the conformation of bends formed by GS is not always close to that of a type II′ β-turn in the conformations, indicating more conformational flexibility for this sequence. As the NG segment also has a high probability to form a β-turn, bend at the NGS location is frequently observed, resulting in more flexible turns. As a result, the β-sheets formed by this peptide have more turn configurations and are less stable compared with those formed by peptide PG.

With peptide NG, bending of the backbone usually occurred around the NG segments. The backbone conformation at this location is even more flexible. As the GK segment in the peptide sequence also has a high possibility in forming type II′ β-turns, and its propensity of β-turn formation is as strong as that of the NG segment,[17] the conformation at the two bending positions is very flexible. With the competing effects between NG and GK, the peptide does not form a unique β-turn, leading to unstable side-chain packing. Thus, it is difficult for peptide NG to fold into β-sheet, and the structure is much less stable once folded.

A good turn sequence is not sufficient for the folding of a β-sheet structure, as most coil conformations also form β-turns at these locations; favorable side-chain interaction is needed for the stability of the β-sheet structure. A good side-chain packing pattern is required for a stable β-sheet, while lacking this often results in briefly observed structures. Different turn configuration often leads to different registration in backbone hydrogen bonds and a different side-chain packing pattern. Side-chain packing is not as rigid as the turn location, because the side-chains often easily adjust their conformations to find good packing. The observation of hydrophobic clusters in β-sheet conformations with different turn configuration is a good indication of the flexibility in side-chain interactions.

The Effect of van der Waals Attractions. As mentioned in the Methods section, simulations with a longer truncation distance at $r = 1.25r^*$ for the VDW interaction were tested. These simulations were started with the extended conformations and were carried out for 100 ns at the same temperatures as in the simulations with $r = r^*$. For peptide GS, the first three-strand β-sheet conformation formed shortly after 5 ns, which has a three-residue turn at the N-terminal NGS location and a β-turn at the C-terminal NG location. The three-strand β-sheet was not very stable and unfolded 1.5 ns later. This β-sheet was observed again at 38−40 ns. A new three-strand β-sheet with two type I′ β-turns at the two NG locations folded at 11 ns. It remained folded for about 1.5 ns. The peptide adopted mainly coil conformations in the following 10 ns and refolded into this three-strand β-sheet at 23 ns. The newly formed β-sheet was very stable, remaining folded until 36 ns. But the conformation at the two turn regions is unstable, especially the C-terminal β-turn, often unfolded into a four-residue loop at the QNGS sequence. Another three-strand β-sheet with a type II′ β-turn at the N-terminal GS location and a three-residue turn at the C-terminal NGS location folded briefly between 41 and 42 ns. At about 63 ns, the peptide refolded into yet another three-strand β-sheet conformation, with two three-residue turns at the N-terminal NGS and C-terminal QNG locations. The peptide remains folded in this three-strand β-sheet for about 14 ns. The overall β-ratio in this 100-ns simulation is 33.5%, higher than that with $r = r^*$ (26.5%).

For the peptide NG, a C-terminal β-hairpin formed at 16.8 ns, which has a three-residue turn at the C-terminal NGK location. This β-hairpin quickly developed into a three-strand β-sheet, with the formation of a third β-strand occurring at about 17.4 ns. This three-strand β-sheet has two three-residue turns

at the two NGK locations, with the N-terminal NG adopting a type I′ β-turn. The three-strand β-sheet is very stable. The N-terminal β-strand unfolded at about 31.2 ns and the C-terminal β-hairpin stayed folded for about 4 ns more until the peptide completely unfolded. A short three-strand β-sheet folded again at about 39 ns, lasting for 5 ns. After a long coil region between 44 and 63 ns, the peptide refolded into a three-strand β-sheet conformation between 63.8 and 66.3 ns. Again, the N-terminal β-hairpin unfolded first in the unfolding process. The peptide refolded into a three-strand β-sheet at about 72 ns and unfolded 4.5 ns later. The overall β-ratio in this 100-ns simulation is 24.2%, higher than that with $r = r^*$ (17.8%).

In the simulation of the peptide ^{D}PG, a short β-hairpin conformation with a ^{D}PG β-turn formed at 6.5 ns at the N-terminal of the peptide. The length of this β-hairpin increased gradually during the simulation. With the formation of a third strand at about 8.7 ns, the peptide folded into a three-strand β-sheet with two type II′ β-turns at the two ^{D}PG locations. This β-sheet is stable for the rest of the 100-ns simulation, except the occasional unfolding of one of the edge β-strands. The overall β-ratio in this 100-ns simulation is 67.5%, much higher than that with $r = r^*$ (28.5%).

The qualitative results of these folding simulations are consistent with those of the previous simulations. Quantitatively, with the VDW truncation at $r = 1.25r^*$, the β-ratios are higher, particularly for peptide PG. For peptides GS, NG, and PG, the longer VDW truncation increased the β-ratio by 26%, 36%, and 137%, respectively. With the VDW truncation at $r = r^*$, the average SASA of the β-sheet is about 70% of that of the corresponding extended conformation, whereas the average SASA of the coil conformations is close to 80%. Longer VDW truncation increases the compactness of the structures, as the average SASA of the β-sheet reduces to about 65% of the SASA of the extended conformation, and the average SASA of the coil conformations reduces to about 70%. In general, the β-sheet is more compact than the coil conformation. The attractive VDW interaction makes structures compact and stabilizes compact structures more than other structures. The attractive VDW interaction shifted the equilibrium toward the β-sheet structures, thus increasing the β-ratio. The ^{D}PG residues form tight β-turns, which usually result in a very compact antiparallel β-sheet structure. Therefore, the longer VDW truncation increases the β-ratio of the PG peptide much more than that of the GS and NG peptides. The longer VDW truncation also increases the heights of energy barriers between energy minima. Compared with the simulations with $r = r^*$, the three-strand β-sheet folds later in these simulations, but tends to be more stable for all three peptides.

Different Initial Conformations. Random coil initial conformations were tested for all three peptides we studied. To generate random coil conformations, one method assigns random numbers to the ϕ, φ dihedral angles. However, this method often results in close contact among atoms and additional structural adjustment is needed. Other studies have used high-temperature simulation to generate random coil conformations.[39] To avoid unrealistic distortions of the molecular structure, such as the *trans*-CONH near planer geometry, we used a moderately high temperature (500 K). In addition, the nonbonding interactions (the electrostatic interaction and the solvent effect) were set to zero so that the structural changes were dominated by random thermal motions. The covalent bonding constraints and the repulsive VDW interactions keep the structure from falling apart. For each sequence, two conformations at 1 and 2 ns in the simulation were taken as the initial conformations for the following two β-sheet folding simulations.

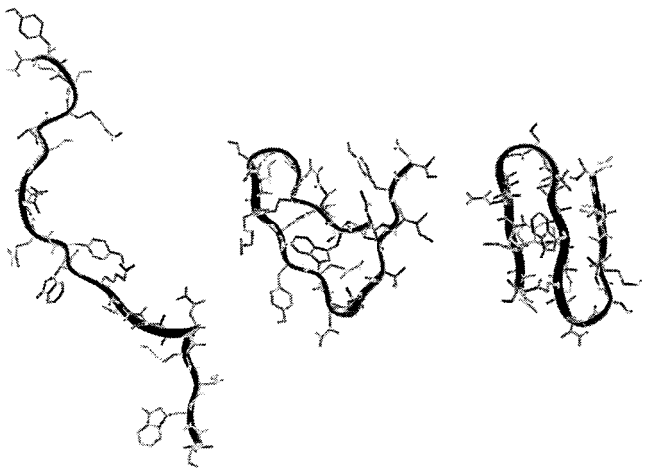

Figure 9. The initial conformation, a compact conformation at 2.6 ns, and a β-sheet conformation at 6.5 ns of the simulation of peptide GS, shown from left to right. The backbone of the peptide is shown as ribbons. The compact conformation has three strands and inter-strand side-chain interactions, but does not have the β-sheet hydrogen bonding pattern. The lower terminus of the peptide is the N-terminus in each structure.

For each peptide, two simulations were carried out at the corresponding experimental temperature with the same energy function as used in the simulation with the extended initial conformation and with VDW truncation at $r = r^*$. For the peptide GS, the simulations were carried out at 283 K. With the first initial conformation (Figure 9), a turn formed at GS in the N-terminal part of the peptide at 0.6 ns. This turn developed into a short β-hairpin conformation with 1 to 3 hydrogen bonds and lasted for about 0.3 ns. At 1.0 ns, a short β-hairpin conformation occurred at the C-terminal part of the peptide, but it was not stable. Frequently, the whole peptide showed a three-strand compact structure with side-chain interactions, similar to the conformation at 2.6 ns shown in Figure 9. From 3.0 ns, a β-hairpin at the C-terminal part was observed with occasional unfolding, but the N-terminal strand had only 1 or 2 backbone hydrogen bonds with the central strand. A three-strand β-sheet conformation formed at 5.7 ns and was stable for about 2 ns. Unfolding and refolding were observed in the remaining part of the simulation. Partial unfolding sometimes occurred, where the structure contains a β-hairpin and an unfolded strand, similar to the last two conformations in Figure 4.

Compact conformations similar to that at 2.6 ns were frequently observed. These conformations have considerable side-chain interactions between each pair of strands, but do not have the backbone hydrogen bonding, except the one or two hydrogen bonds very close to the turn. Similar structures were also observed in other simulations, including those of other peptides, suggesting a general folding mechanism. The side-chain interactions are mainly hydrophobic and are dependent on the nonlocal interaction of the amino acid sequence. It brings the strands together reducing the conformational entropy and shields the backbone polar groups from solvent. The backbone hydrogen bonds often form from the turn to the other end of the strands, locking the structure in the β-sheet conformation. From these observations, the β-sheet folding mechanism may be summarized as follows: coils with turns ⇔ compact structures with side-chain interactions ⇔ β-sheets. Note that each structural conversion is microscopically reversible and the stability of one structure affects that of the previous and

following structures. The turns form very fast and the coil conformations almost always have turns at different positions. The formation of the turn at the proper position is the initiation step of folding. The compact structure with side-chain interactions forms relatively fast, and it includes a large number of similar conformations. Its conversion to a β-sheet includes the reorientation of the side-chains of the compact structure and the partially folded β-sheet, and is much slower than other steps of folding, which could suggest a macroscopic two-state folding kinetics.[67] Similar observations have been reported in previous studies of the β-hairpin folding.[34] It is interesting to note that the compact structure with side-chain interactions resembles some features of the molten globule in protein folding.

With the other initial conformation of the peptide GS, three strand β-sheet conformations were formed at 9.9 ns. For the peptide NG, the two simulations were carried out at 273 K. The β-sheet conformations were formed around 4.0 ns with one initial conformation and around 10.0 ns with the other initial conformation. For the peptide ^{D}PG, the two simulations were carried out at 297 K. The β-sheet conformations were formed around 10.9 ns with one initial conformation and around 14.0 ns with the other initial conformation. During these simulations, the conformations observed for the three peptides are very similar to those shown in Figure 4, Figure 6, and Figure 8, respectively.

To further test the dependence of the simulation results on the initial conformation, a right-handed α-helical initial conformation was used in a simulation of the peptide GS. The initial conformation did not have any close contacts or high covalent bonding tensions. The C-terminus of the helix started unfolding in 0.1 ns and the side-chain interactions increased. At 0.5 ns, the C-terminal part formed an extended strand and the N-terminal part remained helical. At 0.9 ns, the helix unfolded completely, and a three-strand compact structure was observed. At 3.0 ns, a complete three-strand β-sheet formed. With different initial conformations, the exact folding trajectories and folding times were different. However, the qualitative results of β-sheet folding do not depend on the initial conformation.

The Effect of Different Solvation Parameters. To further address the importance of the solvent effects on β-sheet folding, a set of three 50-ns simulations was carried out for peptides GS, NG, and PG with solvation parameters set to zero for all atoms. During these simulations, the peptide conformation was mainly random coil, and no stable β-sheet or β-hairpin conformations formed. These tests showed that the solvent effect is crucial for the folding and stability of β-sheet and β-hairpin structures, a finding that agrees qualitatively with the study of Yang and Honig.[62] Compared with other interactions, such as the electrostatic interaction, the specificity of the solvent effect is primarily in its atom-type dependence. For example, the solvent effect between two hydrophilic atoms is repulsive, regardless of their charges. As shown in our previous studies, the exact distance dependence of the specificity of the solvation effect is secondary compared with the atom-type dependence.[34]

In our simulations, the magnitude of the solvation energy difference between β-sheet conformation and coil conformation is not as large as that in some theoretical studies that ignored the nonlocal interactions (in one-dimensional sequence space) for the coil conformation. Using the two-state assumption, the coil conformation in our calculation is the ensemble of the non-β-structures generated by the simulation. Because of the hydrophobic effect, many coil conformations are compact with side-chain interactions. The nonlocal interactions in these structures significantly decreased the solvation energy of the coil conformations. This is clearly shown in Table 2, which lists the solvation energies for different conformations of the three peptides. The average solvation energies are calculated over the respective structural regions, as mentioned earlier. The solvation energy differences between the extended conformations and the coil conformations are around 10 kcal/mol or more; in contrast, the solvation energy differences between the coil conformations and the β-sheet conformations are less than 4 kcal/mol for all three peptides. With $r = 1.25r^*$, the VDW attraction increases the solvation energy differences between the extended and coil conformations. In both cases, the energy differences between the β-sheet and coil conformations are much smaller than those between the β-sheet and the extended conformations.

Table 2. Solvation Energies for Different Conformations of the Peptides

peptide	c[a]	extend	coil	β-sheet
GS	1.0	38.36	23.70	20.13
	1.25	38.36	22.20	19.53
NG	1.0	27.66	18.13	17.07
	1.25	27.66	16.70	15.19
PG	1.0	43.46	29.44	27.55
	1.25	43.46	26.97	25.41

[a] VDW interaction is truncated at cr^*.

Table 3. Ratios of Hydrophobic Solvent Accessible Surface Area for Different Conformations of the Peptides

peptide	c[a]	extend	coil	β-sheet
GS	1.0	0.58	0.50	0.49
	1.25	0.58	0.51	0.48
NG	1.0	0.49	0.41	0.41
	1.25	0.49	0.43	0.40
PG	1.0	0.68	0.60	0.62
	1.25	0.68	0.62	0.62

[a] VDW interaction is truncated at cr^*.

Table 3 lists the ratio of hydrophobic SASA against the total SASA for each group of conformations of the three peptides. Among these three peptides, PG is the most hydrophobic peptide, as it has the largest hydrophobic SASA ratio in all three different conformations, and NG is the most hydrophilic peptide. The hydrophobicity of a peptide could be another factor, other than the turn sequences, contributing to the stability of the β-sheet conformation, as the more hydrophobic peptide has a higher β-ratio. This agrees with the finding that hydrophobic residues generally have higher β-sheet propensities. The data also show that the hydrophobic SASA ratios are very close for the coil conformations and β-sheet conformations, both about 8% lower compared with the ratios for extended conformations. Using a longer VDW truncation distance does not significantly change the ratios, in agreement with the fact that little change was observed for the solvation energy difference between the β-sheet and coil conformations.

The Sequence Dependence of Folding. The simulation results obtained using these β-sheet-forming peptides were compared with those from an alanine-based synthetic peptide, Ac-AAQAAAAQAAAAQAAY-NH_2 ($(AAQAA)_3Y$ hereafter). Experimentally, $(AAQAA)_3Y$ folds into a stable helical conformation in aqueous solution, and its helical content is about 50% at 274 K, as measured by CD.[63] The details of the folding simulations of $(AAQAA)_3Y$ have been reported previously.[33]

(62) Yang, A. S.; Honig, B. *J. Mol. Biol.* **1995**, *252*, 366−376.

(63) Scholtz, J. M.; York, E. J.; Stewart, J. M.; Baldwin, R. L. *J. Am. Chem. Soc.* **1991**, *113*, 5102−5104.

(64) Okamoto, Y. *Proteins* **1994**, *19*, 14−23.

The simulations were carried out for 50 ns at 274 K using the same energy function as that used for the GS, NG, and PG peptides. With an extended initial conformation, stable helical conformations formed during the simulations. The overall helical ratio is 48.9% with the VDW truncation at $r = r^*$, and 63.6% with the VDW truncation at $r = 1.25r^*$. These results show that the folding simulations are sequence-dependent and are not a consequence of the bias of the method or the energy function.

Conclusions

Without including solvent effect, the multiple minima problem[40] is a major obstacle for protein folding simulations with the traditional empirical force fields. Either thermal perturbations[64] or high temperatures, such as those used in the simulated annealing method,[65] are needed for peptide-folding simulations. A modification of the force field to include an average solvent effect has made the constant-temperature helix folding simulation possible.[32] With the atom-based solvent effect, β-sheet folding can now be simulated. This model is applicable to qualitative study of the secondary structure folding. Quantitatively, more realistic models are needed. For example, the implicit solvent models often accelerate conformational changes, which is useful for conformational search, but cannot quantitatively estimate the folding time.[37]

From the simulations, it is clear that the first factor determining β-sheet folding is the amino acid sequence. The selection of the turn residues has a profound effect on the folding and stability of the β-sheet structure. The β-sheet folding is usually initiated by the formation of turns. In the folding process, the peptide backbone always forms bends at the sequences with high β-turn propensity. The bends at these locations are usually preserved even in the coil structures readily to form β-turns once favorable side-chain interaction is achieved. Peptide PG has the highest β-ratio and most stable β-sheet structure as a result of the strong preference of forming β-turns at the two ^{D}PG segments. The rigid side-chain of ^{D}P greatly restricted its backbone conformation, leading to dominant type II′ β-turn at the ^{D}PG location. In peptide GS, type II′ β-turns at the two GS locations were frequently observed in the three-strand β-sheet structures. But a three-residue turn at NGS and a type I′ β-turn at NG locations were also found in many β-structures. This is because of the high statistical probability of forming a type I′ β-turn at the NG location as well as the backbone flexibility of glycine residue. Although a type I′ β-turn suits the β-hairpin conformation better than a type II′ β-turn geometrically, the type II′ β-turn allows better backbone interstrand hydrogen bonding, which is essential for small β-sheet structures. Isolated β-sheet and β-hairpins with type I′ β-turns are less stable compared with those with type II′ β-turns, as peptide NG has the lowest β-ratio among the three peptides we investigated. These simulation results also agree with the experimental findings by Stanger and Gellman[66] that replacing the ^{D}PG segment with an NG segment leads to a less stable β-hairpin structure.

Besides the amino acids in the turn, the hydrophobic interactions among the side-chains of the strands played a crucial role. Hydrophobic clusters formed through the packing of hydrophobic side-chain atoms contribute significantly to the stability of all β-sheet structures. A good side-chain packing pattern is required for a stable β-sheet, while lacking good packing often results in briefly observed structures. Because the side-chains readily adjust their conformations, the side-chain packing is not as rigid as the turn location. This interaction may explain why the more hydrophobic peptides often show a higher β-ratio. For all three sequences in this study, no stable β-sheet has been observed in the simulations with solvation parameters set to zero. Therefore, for a proper sequence the solvent effect in the energy functions is a major factor that drives β-sheet folding.[34,35,68]

Kinetically, it is observed during the simulation that the side-chain interactions occur prior to forming backbone hydrogen bonds. Because the side-chains are more flexible than the backbone, they often reach each other before the backbone polar groups do. The favorable side-chain interactions bring the backbones of the strands together and shield the backbone polar groups from the competing hydrogen bonding with solvent. For global folding, some simulation suggested that the initial collapse and the secondary structure formation occurred simultaneously,[39] but the question is currently under debate. For local folding of a β-sheet, our simulations showed that the side-chain interactions occur prior to the backbone hydrogen bonding. The attractive interactions among the side-chains are mainly from the solvent effect, especially the hydrophobic interactions. In this sense, the hydrophobic interaction drives the backbone hydrogen bonding. Once the backbone hydrogen bonds formed, they stabilize the backbone conformation as well as the side-chain interactions. In general, the side-chains, not the backbone, carry the sequence information. The side-chain interaction shows how the amino acid sequence determines folding. The side-chain effect of the turn residues on the turn-forming tendency is well recognized. In our previous helix folding study,[33] some kinetic effects of the interaction between the side-chain and the backbone have been observed. The current β-sheet folding simulations suggest that the mechanism of the β-sheet folding includes the initiation step of forming β-turns, the intermediate step of the side-chain interactions, and the final step of the β-sheet hydrogen bonding. Microscopically, each structural conversion is reversible and the stability of one structure affects that of the previous and following structures. The formation of the turns is usually fast. The conversion from the compact structure with side-chain interactions to the β-sheet is the slowest step. This conversion contains mainly a search for optimal side-chain packing, including the reorientation of the side-chains of the compact and the partially folded structures. For an amino acid sequence that can form a stable β-sheet, the optimal side-chain packing will result in the favorable backbone hydrogen bonding pattern of the β-sheet.

JA992359X

(65) Wilson, S. R.; Cui, W. L. *Biopolymers* **1990**, *29*, 225−235.
(66) Stanger, H. E.; Gellman, S. H. *J. Am. Chem. Soc.* **1998**, *120*, 4236−4237.
(67) Eaton, W. A.; Munoz, V.; Thompson, P. A.; Henry, E. R.; Hofrichter, J. *Acc. Chem. Res.* **1998**, *31*, 745−753.
(68) Wang, H.; Sung, S. S. *Biopolymers* **1999**, *50*, 763−776.