doi:10.1006/jmbi.2000.4400 available online at http://www.idealibrary.com on IDEAL® *J. Mol. Biol.* (2001) **306**, 837–850

JMB

Native Topology or Specific Interactions: What is More Important for Protein Folding?

Philippe Ferrara and Amedeo Caflisch*

Department of Biochemistry
University of Zürich
Winterthurerstrasse 190
CH-8057, Zürich
Switzerland

Fifty-five molecular dynamics runs of two three-stranded antiparallel β-sheet peptides were performed to investigate the relative importance of amino acid sequence and native topology. The two peptides consist of 20 residues each and have a sequence identity of 15%. One peptide has Gly-Ser (GS) at both turns, while the other has D-Pro-Gly (^{D}PG). The simulations successfully reproduce the NMR solution conformations, irrespective of the starting structure. The large number of folding events sampled along the trajectories at 360 K (total simulation time of about 5 μs) yield a projection of the free-energy landscape onto two significant progress variables. The two peptides have compact denatured states, similar free-energy surfaces, and folding pathways that involve the formation of a β-hairpin followed by consolidation of the unstructured strand. For the GS peptide, there are 33 folding events that start by the formation of the 2-3 β-hairpin and 17 with first the 1-2 β-hairpin. For the ^{D}PG peptide, the statistical predominance is opposite, 16 and 47 folding events start from the 2-3 β-hairpin and the 1-2 β-hairpin, respectively. These simulation results indicate that the overall shape of the free-energy surface is defined primarily by the native-state topology, in agreement with an ever-increasing amount of experimental and theoretical evidence, while the amino acid sequence determines the statistically predominant order of the events.

Keywords: structured peptide folding; three-stranded antiparallel β-sheet; molecular dynamics simulation; implicit solvation model; free-energy surface

*Corresponding author

Introduction

In the post-genomic era, the protein-folding problem is probably the greatest grand-challenge in biology. Although interesting findings about the folding process have been provided by theoretical and experimental studies (Dobson & Karplus, 1999), a more detailed understanding will lead to significant progresses in biology and medicine. Regular elements of secondary structure, α-helices and/or β-sheets, are ubiquitous in gene products. Therefore, an in-depth knowledge of the formation of secondary structure will improve the understanding of the protein-folding reaction (Lacroix *et al.*, 1999; Imperiali & Ottesen, 1999; Crane *et al.*, 2000). Even for a small protein it is currently not yet feasible to simulate the complete process of folding using molecular dynamics (MD) simulations with an all-atom model (Duan & Kollman, 1998). Hence, the common approach taken in the past was to unfold starting from the native structure (Caflisch & Karplus, 1994, 1995). For a β-heptapeptide, the reversible folding has been achieved by MD simulations in explicit methanol (Daura *et al.*, 1998). The thermodynamic properties of two peptides (an α-helix and a β-hairpin) have been determined with the help of an implicit solvation model and adaptive umbrella sampling (Schaefer *et al.*, 1998). Furthermore, the free-energy surface of Betanova, a three-stranded β-sheet peptide, has been constructed starting from conformations obtained during unfolding simulations in explicit water at elevated temperatures (between 350 K and 400 K) (Bursulaya & Brooks, 1999).

Lately, for a designed sequence of 20 amino acid residues, we have performed MD simulations with

Abbreviations used: GS peptide, a three-stranded antiparallel β-sheet with turns at Gly6-Ser7 and Gly14-Ser15; ^{D}PG peptide, as GS peptide but with turns at D-Pro6-Gly7 and D-Pro14-Gly15; MD, molecular dynamics; ECO, effective contact order.

E-mail address of the corresponding author: caflisch@bioc.unizh.ch

0022-2836/01/040837–14 $35.00/0

an implicit model of the solvent, and demonstrated the reversible folding (Ferrara & Caflisch, 2000) to the NMR conformation, a three-stranded antiparallel β-sheet with turns at Gly6-Ser7 and Gly14-Ser15 (GS peptide; De Alba *et al*., 1999). In this work, the folding pathways and free-energy surface of another three-stranded antiparallel β-sheet of 20 amino acid residues with turns at D-Pro6-Gly7 and D-Pro14-Gly15 (^{D}PG peptide) are investigated by MD and compared to the GS peptide. ^{D}PG is a designed amino acid sequence (Ac-Val1-Phe2-Ile3-Thr4-Ser5-D-Pro6-Gly7-Lys8-Thr9-Tyr10-Thr11-Glu12-Val13-D-Pro14-Gly15-Orn16-Lys17-Ile18-Leu19-Gln20-NH_2), where Orn stands for ornithine. Circular dichroism and chemical shift data have provided evidence that ^{D}PG adopts the expected three-stranded antiparallel β-sheet conformation at 24 °C in aqueous solution (Figure 1) (Schenck & Gellman, 1998). Moreover, ^{D}PG was shown to be monomeric in aqueous solution by equilibrium sedimentation. Although the percentage of β-sheet population was not estimated, nuclear Overhauser enhancements indicate that both hairpins are highly populated at 24 °C. Here, an implicit solvation model is used to overcome the problem of the CPU-intensive MD simulations in explicit water. At 360 K, the 15 simulations of ^{D}PG started from random conformations reached the native state at least once, with an average of 4.4 folding events every 100 ns. This allows us to estimate the free-energy surface as a function of relevant progress variables. The low sequence identity (15 %) between the GS and ^{D}PG peptides serves to emphasize the features of the folding process that depend on the topology of the native state rather than the amino acid sequence.

Recently, Wang and Sung have simulated the folding of Betanova, the GS, and ^{D}PG peptides by MD with implicit solvent at temperatures slightly below 300 K (Wang & Sung, 2000). They sampled a few folding and unfolding events in a 100 ns run for each peptide. Our results are closer to equilibrium conditions because MD simulations at 360 K yield a significant amount (ca 50) of folding and unfolding transitions of three-stranded antiparallel β-sheet peptides. This allows us to estimate the free-energy surface and permits a statistically significant study of the sequence of events and folding pathways.

Results and Discussion

Reversible folding

The present analysis focuses on the ^{D}PG peptide, whose folding behavior is compared to that of the GS peptide. A detailed analysis of the energy surface and folding pathways of the GS peptide has been presented elsewhere (Ferrara & Caflisch, 2000). Since the NMR conformation of ^{D}PG was not available when this study was initiated, a 100 ns simulation at 360 K was started from a completely extended conformation. Five folding transitions to an antiparallel three-stranded β-sheet were observed and a conformation was picked randomly after the first folding event. It was the starting point of two 100 ns simulations, one at 300 K (Figure 2(a) and (d)) and the other at 330 K (Figure 2(b) and (e)). The C^{α} root-mean-square deviation (RMSD) between the average structure over the whole 100 ns simulation at 300 K and the average NMR conformation is 1.33 Å (1.15 Å for the segment 2-19). The trajectory at 300 K was used to calculate average inter-proton distance violations, d_{viol}, which are defined as $\langle r(t)^{-6}\rangle^{-1/6} - r_{exp}$, where $r(t)$ is the inter-proton dis-

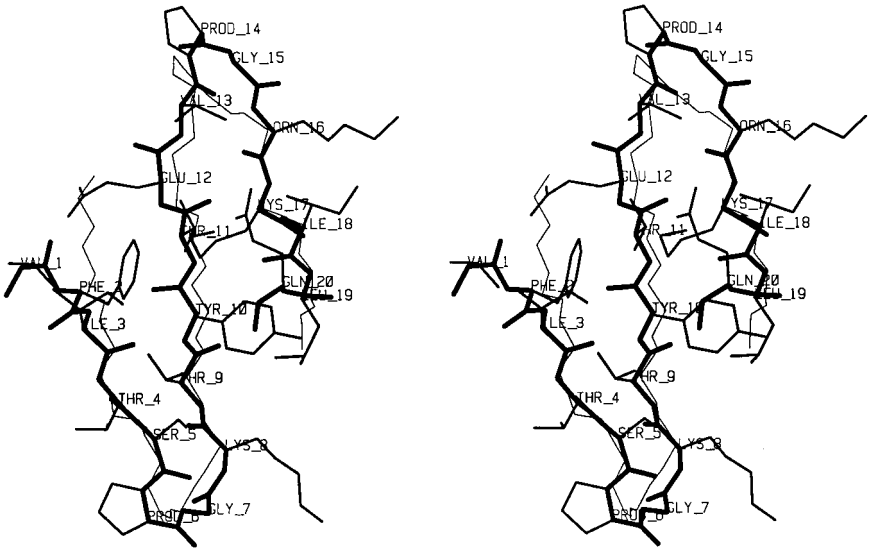

Figure 1. Relaxed eyes stereo picture of the first NMR model of the ^{D}PG peptide (backbone and side-chains in thick and medium lines, respectively) and the average structure over the whole 100 ns simulation at 300 K (thin backbone lines). The C^{α} RMSD between the two structures is 1.7 Å (1.1 Å for the segment 2-19). PROD stands for D-Pro.

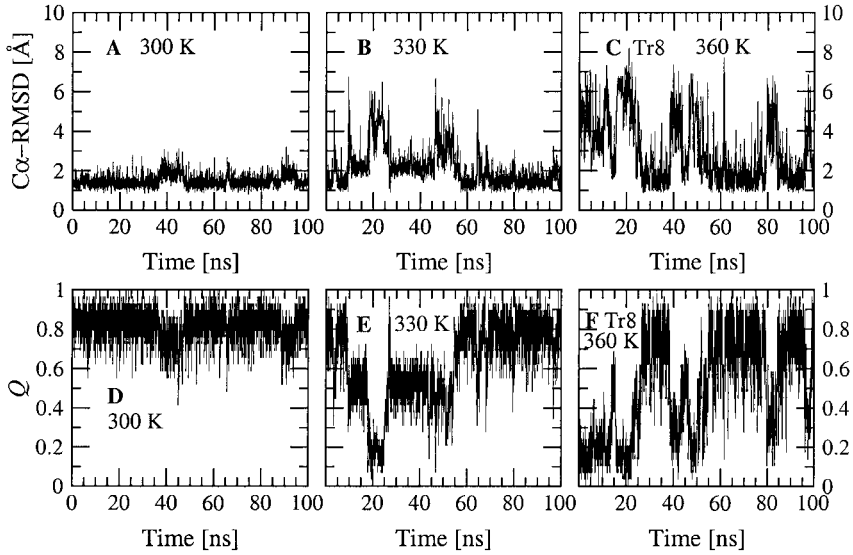

Figure 2. Time-dependence of (a)-(c) the C^{α} RMSD from the mean NMR structure and (d)-(f) the fraction Q of native contacts in the ^{D}PG peptide. (a) and (d) The 300 K simulation started from the folded conformation. (b) and (e) The 330 K simulation started from the folded conformation. (c) and (f) The 360 K simulation started from random conformation 8.

tance at simulation time t, r_{exp} is the NOE upper distance limit, and $\langle\ \rangle$ represents a time average. The 44 NOE restraints are satisfied ($d_{viol} < 0.0$ Å for 36 distances and $d_{viol} < 1.0$ Å for four distances), except for four distances between Q^{ε} Phe2-Thr9 $Q^{\gamma 2}$, $Q^{\delta 1}$ Ile3-Tyr10 Q^{ε}, Q^{ε} Tyr10-Lys17 Q^{β} and HN Val13-Gly15 Q^{α}, with d_{viol} values of 1.7 Å, 2.3 Å, 1.4 Å and 1.5 Å, respectively. Turns of type I′ and II′ were observed in the simulation at 300 K. Type I′ turns were present twice as often as type II′ at residues 6-7, whereas both types of turns were found to be populated nearly equally at residues 14-15. Although the ^{D}PG peptide has been designed to adopt type II′ β-turns (Schenck & Gellman, 1998), with the available experimental data (NMR chemical shifts) it is not possible to discriminate between type I′ and II′. In fact, in the 20 NMR models, 30 and ten dihedral angles (ϕ,ψ) in ^{D}Pro are close to the ideal values (Hutchinson & Thornton, 1994) of type I′ and II′, respectively. On the contrary, most of the dihedral angles (ϕ,ψ) of the two Gly involved in the turns are more than 30° away from the ideal values of either type I′ or II′. The fact that ^{D}PG can accommodate both types of turn may explain, for entropic reasons, the strong propensity of the ^{D}PG segment for β-hairpin formation (Stanger & Gellman, 1998). Yet, the possibility cannot be ruled out that different turns originate from the approximations inherent to the force-field and solvation model.

The fraction of native contacts Q is a progress variable used often to monitor folding (Dobson *et al.*, 1998). Tables 1 and 2 contain a list of the native contacts for the ^{D}PG and GS peptide, respectively. At 300 K, there is no conformation with $Q < 0.3$ (Figure 2(d)), while at 330 K the ^{D}PG peptide is completely unfolded between 19 and 27 ns (Figure 2(e)). Furthermore, at the higher temperature it is partially unfolded between 3 and 4 ns, 11 and 18 ns, and 28 and 55 ns. The structures sampled during these time-intervals are essentially β-hairpins, where most of the contacts between strands 1 and 2 are formed, whereas most of the interactions between the second and third strands are broken. The peptide stays folded most of the time in the second half of the simulation, although it undergoes a few transient unfolding events.

To assess the importance of solvation, a 100 ns simulation at 360 K was performed from the extended conformation of the ^{D}PG peptide with a distance-dependent dielectric function ($\varepsilon(r) = 2r$) but without the solvent-accessible surface energy contribution. The peptide underwent two folding and unfolding transitions (not shown). In a similar test run, the GS peptide did not reach the folded state within 200 ns (Ferrara & Caflisch, 2000). This indicates that direct solvation, i.e. the effect of the solvent on the solute-solvent interactions, plays a more important role in the folding of the GS than the ^{D}PG peptide.

Folding pathways

Sampling a statistically significant number of folding events at 300 K and 330 K is too time-consuming, despite the use of an implicit solvation model. Therefore, most of the simulations from random conformations were performed at 360 K. For the GS peptide, 40 simulations were started from random structures. The 13 simulations that did not reach the folded state (defined by a criterion of $Q > 0.85$) in 50 ns, were restarted and run for an additional 50 ns or 100 ns. After 150 ns, only two out of 40 simulations did not reach a Q value

Table 1. List of native contacts in the ^{D}PG peptide

First residue	Second residue	Strands	Location with respect to turn
A. *Hydrogen bonds*			
Val1 H	O Glu12	1-2	Distal
Val1 O	H Glu12	1-2	Distal
Ile3 H	O Tyr10	1-2	Distal
Ile3 O	H Tyr10	1-2	
Ser5 H	O Lys8	1-2	Proximal
Ser5 O	H Lys8	1-2	Proximal
Thr11 H	O Ile18	2-3	Distal
Thr11 O	H Ile18	2-3	
Val13 H	O Orn16	2-3	Proximal
Val13 O	H Orn16	2-3	Proximal
B. *Side-chain contacts*			
Val1	Ile3	1-1	
Val1	Glu12	1-2	Distal
Phe2	Thr9	1-2	Distal
Phe2	Thr11	1-2	Distal
Ile3	Ser5	1-1	
Ile3	Glu12	1-2	Distal
Thr4	Thr9	1-2	Proximal
Ser5	Tyr10	1-2	
Lys8	Tyr10	2-2	
Tyr10	Lys17	2-3	Distal
Tyr10	Leu19	2-3	Distal
Thr11	Val13	2-2	
Thr11	Ile18	2-3	Distal
Thr11	Gln20	2-3	Distal
Glu12	^{D}Pro14	2-2	
Glu12	Lys17	2-3	Proximal
Val13	Orn16	2-3	Proximal
Val13	Ile18	2-3	
Orn16	Ile18	3-3	

A hydrogen bond is defined as native contact if the O···H distance is smaller than 2.6 Å for more than 50 % of the conformations saved during the 100 ns of the 300 K simulation started from the folded state. An interaction between side-chains of residues not adjacent in sequence is defined as native contact if the average distance between geometrical centers is smaller than 6.7 Å.

Table 2. List of native contacts in the GS peptide

First residue	Second residue	Strands	Location with respect to turn
A. *Hydrogen bonds*			
Thr1 O	H Gln12	1-2	Distal
Ile3 H	O Trp10	1-2	Distal
Ile3 O	H Trp10	1-2	
Asn5 H	O Thr8	1-2	Proximal
Asn5 O	H Thr8	1-2	Proximal
Lys9 O	H Thr20	2-3	Distal
Tyr11 H	O Ile18	2-3	Distal
Tyr11 O	H Ile18	2-3	
Asn13 H	O Thr16	2-3	Proximal
Asn13 O	H Thr16	2-3	Proximal
B. *Side-chain contacts*			
Thr1	Gln12	1-2	Distal
Trp2	Tyr11	1-2	Distal
Ile3	Gln12	1-2	Distal
Ile3	Trp10	1-2	Distal
Gln4	Lys9	1-2	Proximal
Asn5	Thr8	1-2	Proximal
Thr8	Trp10	2-2	
Trp10	Lys17	2-3	Distal
Trp10	Tyr19	2-3	Distal
Tyr11	Asn13	2-2	
Tyr11	Ile18	2-3	Distal
Gln12	Lys17	2-3	Proximal
Asn13	Thr16	2-3	Proximal
Asn13	Ile18	2-3	
Thr16	Ile18	3-3	
Ile18	Thr20	3-3	

See the legend to Table 1 for the criteria used to determine the native contacts. Contacts were defined using the first 100 ns of the 200 ns simulation at 300 K started from the folded state.

of at least 0.85. For the ^{D}PG peptide, 15 100 ns trajectories were generated, starting from random conformations. At least one folding event was observed in each simulation. Hence, with the present force-field and solvation model, the two peptides fold irrespective of the initial conformation. Time-series of some of these simulations are shown in Figures 2(c) and (f) and 3 for the ^{D}PG peptide, and in Figure 4 for the GS peptide. As in the 330 K run from the native conformation, intermediates with Q values of about 0.4-0.6 are often observed at 360 K. These are mainly β-hairpins, in which most of the interactions between two strands are formed. For the ^{D}PG peptide, β-hairpins with most of the native contacts formed between strands 1 and 2 were found much more often than β-hairpin 2-3. For example, the latter is present between 14 and 15 ns in trajectory 8 (Tr8) (Figure 2(c) and (f)), whereas the former is observed between 44 and 46 ns in Tr8, and in the 20-25 ns and 44-53 ns intervals in Tr6 (Figure 3). Such intermediates are present in the simulations of the GS peptide.

To gain insights into the folding process, it is important to define appropriate progress variables. Note that a progress variable is only a structural measure of how folded the peptide is. It may not include any information on the kinetics of folding and unfolding. For this, one would need to define a reaction coordinate. Hence, most of the results of this study are concerned with the thermodynamics. As in a previous study of the GS peptide (Ferrara & Caflisch, 2000), $Q_{1\text{-}2}$ is defined as the fraction of native contacts between the first and second strands, while $Q_{2\text{-}3}$ represents the fraction of native contacts between strands 2 and 3 (Tables 1 and 2). $Q_{1\text{-}2}$ and $Q_{2\text{-}3}$ are useful for a clear description of the pathway and were subdivided into the contacts close to the turn ($Q_{1\text{-}2}^{\text{prox}}$ and $Q_{2\text{-}3}^{\text{prox}}$) and the contacts far away from the turn ($Q_{1\text{-}2}^{\text{dist}}$ and $Q_{2\text{-}3}^{\text{dist}}$). There are four folding and three unfolding events in Tr6 of ^{D}PG (Figure 3). The projection into the $Q_{1\text{-}2}Q_{2\text{-}3}$-plane indicates that in the last three folding transitions there is first the formation of most of the contacts between strands 1 and 2, followed by the association of the C-terminal strand onto the preformed 1-2 β-hairpin. On the contrary, the formation of most of the contacts between strands 2 and 3 precedes the appearance of the 1-2 interstrand contacts in the first folding event. Of the three unfolding transitions, the first and the last are essentially the reverse of the last three folding events, whereas in the second the loss of the contacts between strands 1 and 2 precedes the rupture

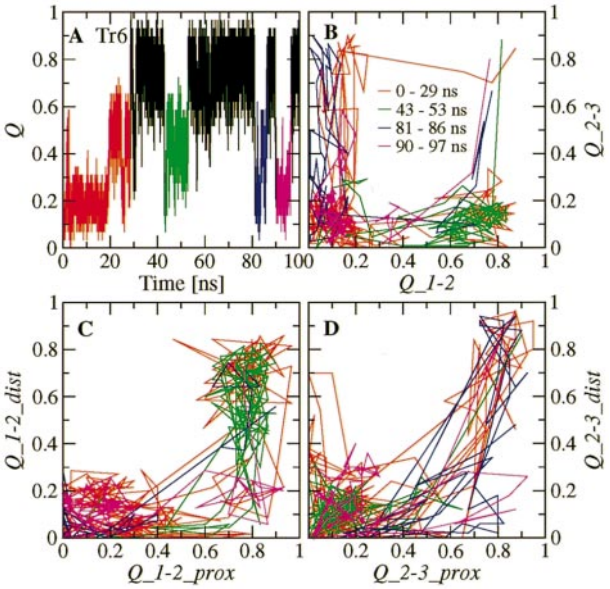

Figure 3. (a) Time-dependence of the fraction of native contacts Q for Tr6 of the ^{D}PG peptide at 360 K. (b) Projection into the $Q_{1\text{-}2}Q_{2\text{-}3}$-plane; (c) the $Q_{1\text{-}2}^{prox}Q_{1\text{-}2}^{dist}$-plane; and (d) the $Q_{2\text{-}3}^{prox}Q_{2\text{-}3}^{dist}$-plane for the same trajectory. The four folding events are shown with the red, gree, blue, and magenta curves. In (b)-(d), the three unfolding events are not shown for clarity reasons and each point corresponds to the average over 100 ps (ten conformations).

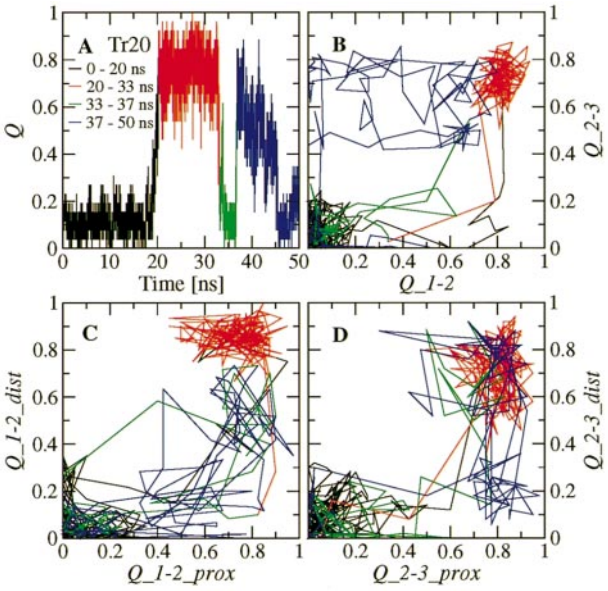

Figure 4. Same as in Figure 2 for Tr20 of the GS peptide. In (b)-(d), both folding and unfolding events are shown.

of 2-3 contacts (not shown). The projection of Tr6 into the $Q_{1\text{-}2}^{prox}Q_{1\text{-}2}^{dist}$ and $Q_{2\text{-}3}^{prox}Q_{2\text{-}3}^{dist}$ shows that contacts propagate from the turn to the terminus of the hairpin in a ''zip-up'' mechanism. A similar zip-up mechanism was found for the GS peptide. Representative trajectories of GS have been presented elsewhere (Ferrara & Caflisch, 2000). The two folding events in Tr20 occur along the less-predominant pathway for the GS peptide (Figure 4). This trajectory was selected to show that it can be misleading to draw conclusions on relatively rare events from a single MD simulation.

Although the exact order of events might depend on the amino acid sequence (see below), the early formation of the turn contacts is in accord with recent implicit solvent MD simulations of Betanova, and the GS and ^{D}PG peptides (Wang & Sung, 2000), with studies of Betanova in explicit solvent (Bursulaya & Brooks, 1999) and the β-hairpin fragment (residues 10-28) of tendamistat (Bonvin & van Gunsteren, 2000), as well as a statistical mechanical model developed to explain laser temperature-jump experiments of the folding of the β-hairpin fragment (41-56) of protein G (Muñoz *et al.*, 1997). On the other hand, recent computational studies on the same fragment of protein G have shown different behavior (Pande & Rokhsar, 1999; Dinner *et al.*, 1999).

Energy surface

For both peptides, the average effective energy (intramolecular plus solvation) as a function of the $Q_{1\text{-}2}$ and $Q_{2\text{-}3}$ progress variables has an almost downhill profile with a single minimum corresponding to the fully folded conformation (Figure 5, left plots). The coloring shows that the effective energy plot of GS is more symmetric than that of ^{D}PG. In the latter, the region corresponding to the 1-2 β-hairpin ($0.6 < Q_{1\text{-}2} < 0.9$ and $0 < Q_{2\text{-}3} < 0.4$) is slightly more favorable than the 2-3 β-hairpin ($0 < Q_{1\text{-}2} < 0.4$ and $0.6 < Q_{2\text{-}3} < 0.9$). The same asymmetry is present in the free-energy surface of ^{D}PG (Figure 5, top right). For both peptides, two thermodynamically defined transition state regions are located at TS1 ($Q_{1\text{-}2} \approx 0.3$, $0.5 < Q_{2\text{-}3} < 0.9$) and TS2 ($0.6 < Q_{1\text{-}2} < 0.8$, $Q_{2\text{-}3} \approx 0.4$). The barriers arise from the loss of conformational entropy associated with fixing about two-thirds of the chain into a β-hairpin. For ^{D}PG the lower barrier in TS2 (free-energy difference between TS2 and denatured state of about 1.3 kcal/mol) corresponds to the statistically predominant folding pathway, which is observed 47 times, whereas only 16 folding events go through the TS1 transition state (barrier of about 2.4 kcal/mol). ^{D}PG unfolded 36 times passing over the TS2 and 17 times over the TS1 transition state. In the GS peptide, the slightly lower barrier in TS1 (free-energy difference between TS1 and denatured state of about 3 kcal/mol) corresponds to the main folding pathway, which occurs 33 times, whereas 17 folding events go through the TS2 transition state (barrier of about 3.5 kcal/mol). For unfolding, 24 trajectories

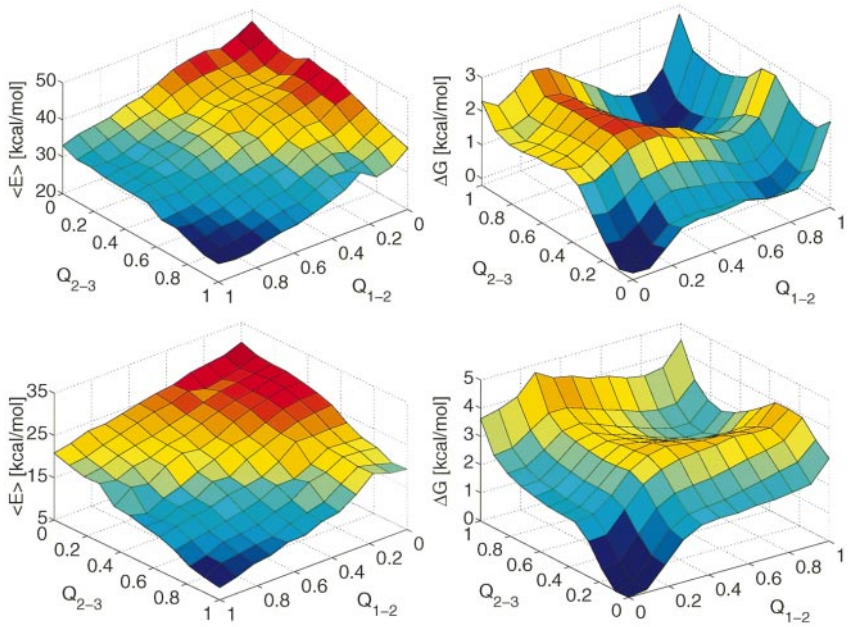

Figure 5. Top: Peptide ^{D}PG, average effective energy ($\langle E \rangle$, left) and free-energy surface (ΔG, right) at 360 K as a function of the fraction of native contacts between residues in strands 1 and 2 ($Q_{1\text{-}2}$), and between residues in strands 2 and 3 ($Q_{2\text{-}3}$). A total of 1.5×10^{-5} conformations sampled during the 15 simulations at 360 K were used. To make it clearer, the plot of ΔG is rotated with respect to that of $\langle E \rangle$ by 180° around an axis going through the center of the horizontal plane. $\langle E \rangle$ was evaluated by averaging the effective energy values of the conformations within a bin without minimizing them. ΔG was computed as $-k_B\ T\ \ln(N_{n,m}/N_{0,0})$, where $N_{n,m}$ denotes the number of conformations with n (m) contacts formed between strands 1 and 2 (2 and 3). The minimum and maximum values of $N_{n,m}$ are 87 and 5120, respectively. The error in ΔG is estimated by separating the 15 simulations started from random conformations into two sets of seven and eight simulations. The average and maximal error of $\langle E \rangle$ are 0.7 and 3.9 kcal/mol (bin with $n = 1$ and $m = 11$), respectively. The average and maximal error of ΔG are 0.3 and 0.9 kcal/mol ($n = 1$, $m = 10$), respectively. Bottom: Peptide GS, the minimum and maximum values of $N_{n,m}$ are 42 and 31,987, respectively. The error in ΔG is estimated by separating the 40 simulations started from random conformations into two sets of 20 simulations each. The average and maximal error of $\langle E \rangle$ are 0.8 and 3.7 kcal/mol (bin with $n = 8$ and $m = 4$), respectively. The average and maximal error of ΔG are 0.2 and 0.6 kcal/mol ($n = 11$, $m = 11$), respectively.

follow the route over the TS1 and 17 over the TS2 transition state.

The relatively high barrier for conformations with few contacts equally distributed in the two partially formed hairpins (i.e. $Q_{1\text{-}2} \approx 0.4 \approx Q_{2\text{-}3}$) is of entropical origin, since the average effective energy is smooth. The entropic penalty of the partial freezing of both hairpins is higher than that involved in complete formation of one β-hairpin because the unstructured strand is almost completely flexible, whereas even a few contacts, if equally distributed between the two hairpins, significantly restrict the available conformational space (see below).

On average, 2.3 side-chain contacts and 1.7 hydrogen bonds are formed between strands 1 and 2 in the conformations with $Q_{1\text{-}2} = 0.3$ and $Q_{2\text{-}3} = 0.7$. Similarly, 3.0 side-chain contacts and 1.0 hydrogen bonds are formed between strands 2 and 3 in the conformations with $Q_{1\text{-}2} = 0.75$ and $Q_{2\text{-}3} = 0.4$. For the GS peptide, it was found that around 2.8 side-chain contacts and 1.2 hydrogen bonds are formed in TS1 and TS2 (Ferrara & Caflisch, 2000). These results indicate that native side-chain interactions are more important than hydrogen bonds to drive the folding of ^{D}PG and GS, as found previously in simulations of Betanova (Bursulaya & Brooks, 1999) and a β-hairpin (Pande & Rokhsar, 1999; Dinner *et al.*, 1999). The most frequent contacts in the transition states of ^{D}PG are close to the turn. They are Thr4-Thr9 (78 %) in TS1 and Val13-Orn16 (87 %) in TS2.

The free-energy surfaces as a function of $Q_{1\text{-}2}^{dist}$ and $Q_{1\text{-}2}^{prox}$, or $Q_{2\text{-}3}^{dist}$ and $Q_{2\text{-}3}^{prox}$ are shown in Figure 6. On this projection, the barriers are less pronounced, which indicates a greater variability in the pathways. Yet, both hairpins in both peptides show a similar free-energy surface. On average, folding is initiated by the formation of two to three proximal contacts in one of the two native turns and zero or one distal contact.

Denaturated state ensemble

To investigate the unfolded state at 360 K, a cluster analysis was performed on the conformations with $Q < 0.3$. For the ^{D}PG peptide, one structure every eighth snapshot saved along the simulation intervals with $Q < 0.3$ was used for clustering, whereas one every 40th snapshot was used for the GS peptide. In this way, for each peptide, nearly 5200 unfolded conformations were used for clustering. For the ^{D}PG peptide, a total of 3650 clusters (455 of which with more than one member) were found. The largest cluster incorporates 1.6 % of the analyzed structures with $Q < 0.3$ and the ten largest clusters 9.2 %. The center of the six most populated clusters are non-native three-stranded β-sheets (Figure 7). In the clusters 1 to 6, there are nine, nine, seven, five, seven and eight interstrands hydrogen bonds, respectively. Non-

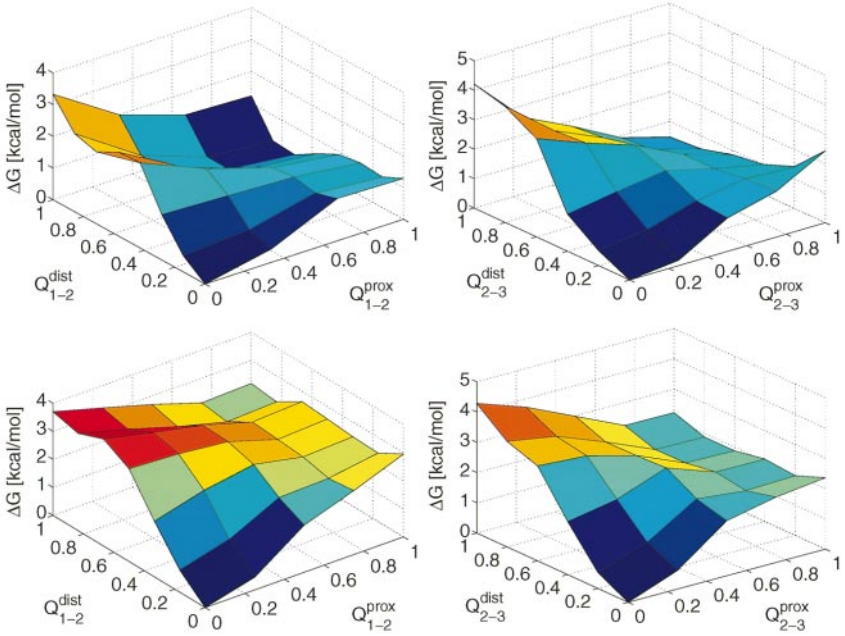

Figure 6. Top: Peptide ^{D}PG. Left, free energy surface (ΔG) at 360 K as a function of the fraction of native contacts between residues in strands 1 and 2 close to the turn (Q^{prox}_{1-2}) and far away from the turn (Q^{dist}_{1-2}). Right, free-energy surface as a function of the fraction of native contacts between residues in strands 2 and 3 close to the turn (Q^{prox}_{2-3}) and far away from the turn (Q^{dist}_{2-3}). Bottom: Same as in the top panels for the GS peptide.

native two-residue or three-residue β-turns are often observed, e.g. Val13-^{D}Pro14 in clusters 1, 4 and 6. The central conformation of cluster 2 has the C-terminal β-strand as the central β-strand.

For the GS peptide, 4441 clusters (334 of which with more than one member) were obtained. The largest cluster incorporates 0.8 % of the analyzed conformations with $Q < 0.3$ and the ten largest clusters incorporate 3.1 %. The central conformation of the six most populated clusters are shown in Figure 8. As for ^{D}PG, there are often non-native backbone hydrogen bonds. The representative of cluster 1 is characterized by an out-of-register three-stranded β-sheet with the C-terminal β-strand as the central β-strand, a three-residue-loop (Gly14-Ser15-Thr16) and five interstrand hydrogen bonds (Trp2 O-Thr20 H, Trp2 H-Ile18 O, Tyr11 O-Tyr19 H, Lys17 O-Asn13 H and Tyr19 O-Trp10 H). A structure comparable to the center of cluster 1 was observed in simulations of the GS peptide with a force-field different from that used here, except that the N-terminal β-strand was the central β-strand (Wang & Sung, 2000). A large amount of helical structure is present in the center of cluster 2 with seven backbone hydrogen bonds (Thr1 O-Gln4 H, Thr1 O-Asn5 H, Trp2 O-Ser7 H, Gly6 O-Lys9 H, Gly6 O-Trp10 H, Ser7 O-Tyr11 H and Trp10 O-Gly14 H). The N terminus is packed against the C terminus through interactions that involve mainly Trp2, Trp10, Thr16 and Tyr19. Little secondary structure is present in the central conformations of clusters 3 and 6, apart from a hydrogen-bonded turn with two hydrogen bonds in cluster 3 and six in cluster 6. In cluster 3, the β-turn is stabilized by the side-chain of Asn5, which makes hydrogen bonds with Tyr11, Gln12 and Ser15. In cluster 6, Trp10 occupies the ''core'' and is surrounded by Trp2, Lys9 and Tyr19. The representatives of clusters 4 and 5 are non-native three-stranded β-sheets. In cluster 4, there are seven interstrand hydrogen bonds with a one (Gly6) and a two-residue loop (Asn13-Gly14). Eight interstrand hydrogen bonds are present in cluster 5. In this conformation, the native β-turn is formed at the N-terminal β-strand, whereas there is an out-of-register β-turn at the C terminus. The larger number of clusters for GS is due to the better sampling and perhaps to the lower stability of GS at 360 K. For both peptides, there is a large number of clusters, which indicates that at 360 K the ensemble of denatured states is broad without any predominant conformation.

At 360 K, the radius of gyration of the denatured state ensemble of ^{D}PG is 8.07(±0.63) Å, whereas for the folded state the average value over the simulation at 300 K is 7.49(±0.15) Å. This shows that at 360 K the unfolded state ensemble consists of rather compact structures with a relatively broad distribution. For the GS peptide, the average radius of gyration is 8.00(±0.50) Å for the denatured conformations at 360 K and 7.66(±0.13) Å for the structures obtained at 300 K. The unfolded structures of the β-hairpin fragment (41-56) of protein G were found to consist of compact conformations at 400 K (Pande & Rokhsar, 1999). On the other hand, more expanded unfolded structures were observed for Betanova (Bursulaya & Brooks, 1999).

For ^{D}PG, nearly 20 % of the unfolded conformations have their backbone dihedral angles in the α-helical region of the Ramachandran map, i.e. $(-180^\circ, -120^\circ) < (\phi,\psi) < (0^\circ, 0^\circ)$. The remaining conformations belong mainly to the extended region $((-180^\circ, 30^\circ) < (\phi,\psi) < (0^\circ, 180^\circ))$. Turns were classified as type I′, II′, I or II if the dihedral angles of two consecutive residues belong to the

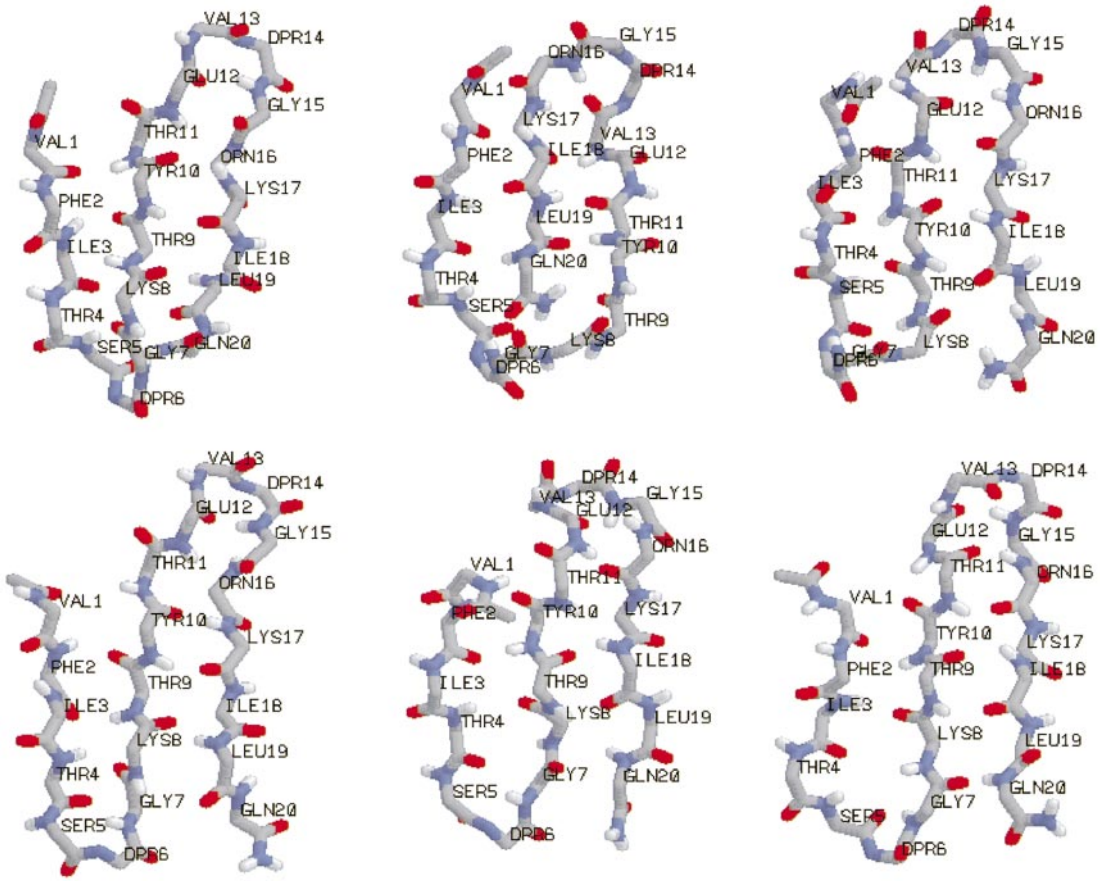

Figure 7. Backbone representation of the center of the six most populated clusters (from left to right and top to bottom) of the ensemble of unfolded states of the ^{D}PG peptide. They incorporate 1.6 %, 1.6 %, 1.1 %, 1.0 %, 0.9 % and 0.8 % of all the unfolded structures. DPR stands for D-Pro.

region defined by the standard values ±30° (Hutchinson & Thornton, 1994). ^{D}Pro6-Gly7 are involved in either a type I′ or II′ turn in nearly 4 % of the unfolded states. For ^{D}Pro14-Gly15, these values are 3 % and 1.7 % in type I′ and II′, respectively. The backbone dihedrals of Gly7 and Gly15 are spread all over the glycine-allowed region of the Ramachandran map. This indicates that the intrinsic propensity of ^{D}Pro-Gly to form the native β-turn is low. Ser5-^{D}Pro6 and Val13-^{D}Pro14 form a type II turn in 19 % and 29 % of the denatured structures, respectively. This is probably a consequence of the restricted accessible space of the φ dihedral angle of ^{D}Pro and the fact that the (φ,ψ) dihedrals of the first residue in a type II turn belong to the extended region. Non negligible percentages of type I turns are observed: 8.5 % for Ile18-Leu19 and 6.2 % for Leu19-Gln20. Non-native type I′ and II′ β-turns are not significantly populated, apart from Gly15-Orn16, which forms a type II′ turn in 1.2 % of the unfolded conformations.

For the GS peptide, around 30 % of the backbone dihedrals belong to the α-helical region. As for ^{D}PG, the intrinsic turn propensity of Gly-Ser is low: nearly 2 % of the unfolded conformations have a type II′ turn at either Gly6-Ser7 or Gly14-Ser15. A similar percentage of type I turn is found at the same location. There is a significant presence of type I turns, particularly in the middle of the sequence. Trp10-Tyr11, Tyr11-Gln12, Lys9-Trp10 and Thr8-Lys9 are the pairs of residues that are the most often involved in a type I turn, with values of 11.3 %, 10.6 %, 10.2 % and 9.7 %, respectively. The largest percentage of type II turn is located at Asn13-Gly14 (2 %). Type I′ and non-native type II′ turns are negligible (<1 %).

Kinetics

At 360 K, the folding time of the ^{D}PG peptide varies among different trajectories from 0.5 ns to 47.6 ns with an average of 10.6 ns, while the mean unfolding time is 12.4 ns. The experimental folding rate is not known. From the free-energy surface, it was assumed that 360 K is close to the melting temperature, at which the folding and unfolding rates should cross. Hence, thermodynamic and kinetic data are consistent. For the GS peptide, the folding time varies among different trajectories from 1.5 ns to 134.8 ns with an average of 31.8 ns, while the mean unfolding time is 7.2 ns. An upper limit for the *in vitro* folding time of 16-45 μs at 5 °C and 4-14 μs at 10 °C was estimated for the GS peptide (De Alba *et al.*, 1999). The lower folding time

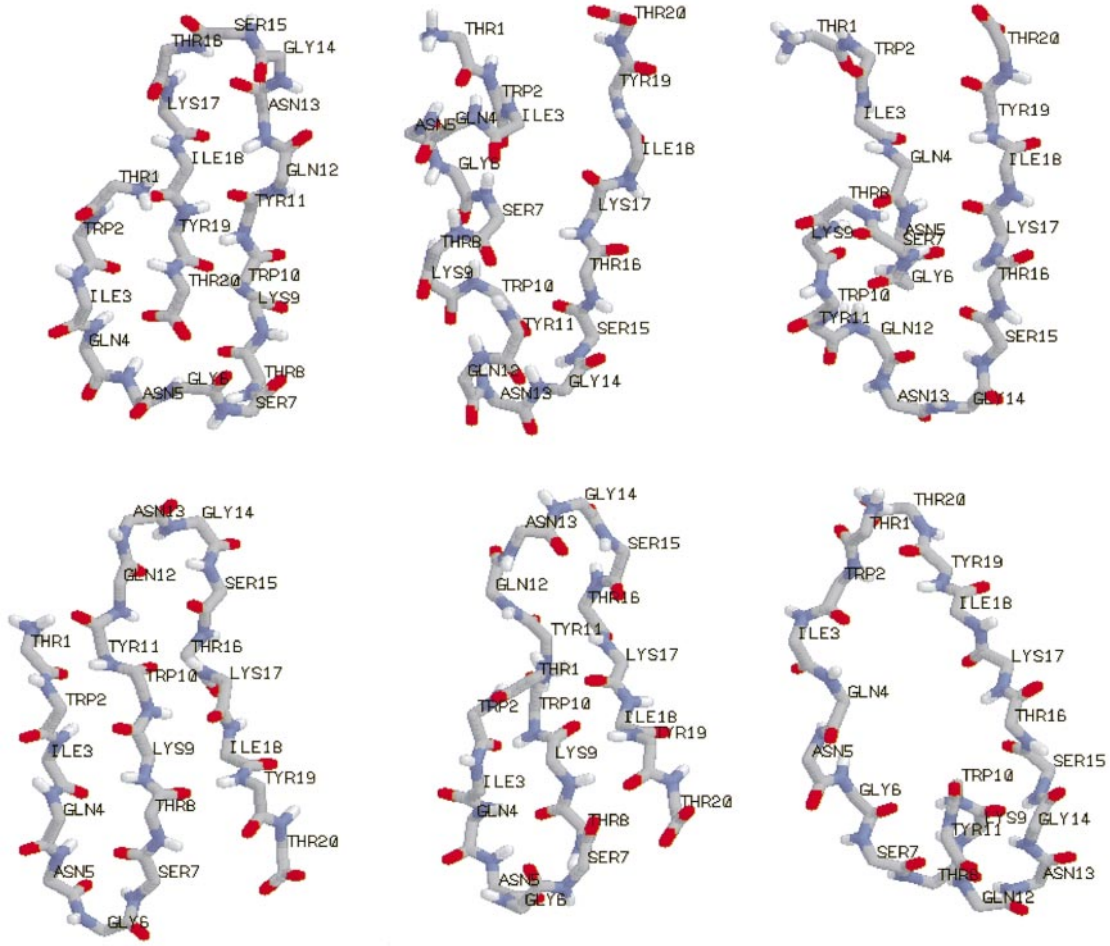

Figure 8. Backbone representation of the center of the six most populated clusters (from left to right and top to bottom) of the ensemble of unfolded states of the GS peptide. They incorporate 0.8 %, 0.4 %, 0.3 %, 0.3 %, 0.3 % and 0.3 % of all the unfolded structures.

in the simulations is due, in part, to the higher temperature, but also to the implicit solvent model, which does not take into account the friction of the water molecules at the surface of the solute. Recently, the conformational transition between the *A* and *B*-DNA forms was observed to be about 20 times faster when using the generalized Born implicit solvation model than it is when using explicit water molecules (Tsui & Case, 2000).

The fraction of unfolded conformations at time t, $P_u(t)$, can be derived from the percentage of trajectories that have not reached the folded state ($Q > 0.85$) at time t. An analogous definition is valid for the fraction of folded conformations, $P_f(t)$, using $Q < 0.15$ as a criterion for the unfolded state. Exponential curves ($P_{u,f}(t) = \exp(-t/\tau_{f,u}^{fit})$) yield characteristic times of decay of $\tau_f^{fit} = 11.4$ ns ($\chi^2 = 0.17$) and $\tau_u^{fit} = 13.1$ ns ($\chi^2 = 0.07$) for ^{D}PG, and $\tau_f^{fit} = 30.3$ ns ($\chi^2 = 0.03$) and $\tau_u^{fit} = 7.6$ ns ($\chi^2 = 0.17$) for GS (Figure 9). No improvement in the χ^2 value was observed when fitting a double exponential. Therefore, the decrease of the unfolded and folded population is essentially monoexponential. This has been observed in two α-helical peptides and a β-hairpin (Ferrara *et al.*, 2000b; Hiltpold *et al.*, 2000).

Role of topology and specific interactions

A variety of computational analysis and experimental data indicate that the pathways and rates of protein folding are determined largely by the native state topology rather than specific interatomic interactions (reviewed by Baker, 2000). Shakhnovich and co-workers studied different 36-mer sequences with the same native fold on a three-dimensional lattice (Abkevich *et al.*, 1994). They determined that eight out of 40 native contacts have to be formed at the transition state (specific nucleus). Moreover, the folding nucleus is a spatially localized substructure that is the same for the 30 sequences analyzed. This clearly points out the crucial role of the native-state topology on the transition state. Based on this result, a strategy was elaborated to identify nucleus residues in proteins with very different sequences but identical folds (superfamilies) (Michnick & Shakhnovich, 1998). Seven potential nucleus residues in ubiquitin

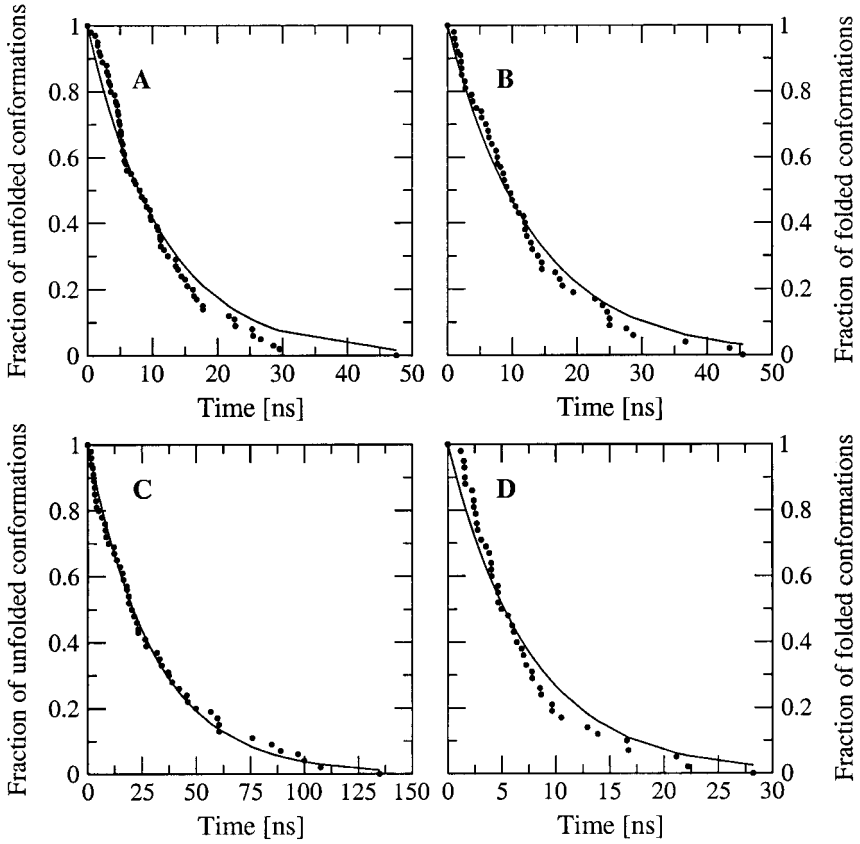

Figure 9. Fraction of (a) unfolded and (b) folded conformations as a function of the simulation time for the ^{D}PG peptide (see the section Kinetics for a full explanation). Fraction of (c) unfolded and (d) folded conformations as a function of the simulation time for the GS peptide. The simulation results are represented by dots, whereas the line represents a monoexponential fit.

superfamily members were identified (Michnick & Shakhnovich, 1998). This led to the concept of "conservatism of conservatism", which was introduced to identify nucleus residues conserved in a superfamily (Mirny *et al.*, 1998; Mirny & Shakhnovich, 1999). A high correlation between conservatism of conservatism and solvent accessibility was found for five different folds; this shows that the most important selection pressure is thermodynamic stabilization of the native fold (Mirny & Shakhnovich, 1999). Residues at positions with a high conservatism of conservatism value but a low correlation with solvent accessibility were under other selective pressure (fast folding and/or function). Protein engineering experiments indicate that for three out of five folds the theoretical approach correctly identifies the folding nucleus.

Of further interest is the high anticorrelation between the folding rate of small proteins and the relative contact order. The latter is a measure of the complexity of the native-state topology and reflects the relative importance of local and non-local contacts (Plaxco *et al.*, 1998). A number of proteins with mainly local contacts have been shown to fold faster than proteins with mainly non-local contacts. Recently, the transition states of the src and α-spectrin SH3 domains, two proteins with the same fold but a sequence identity of only 35%, were found to be similar, although a few differences were pointed out (Martinez & Serrano, 1999; Grantcharova *et al.*, 2000). Lattice simulations (Fiebig & Dill, 1993; Dill *et al.*, 1993), as well as MD simulations (Sheinerman & Brooks, 1998; Ferrara *et al.*, 2000a), have highlighted the role of contact order (average sequence separation) and topology. It was also shown that three different force-fields lead to comparable sequences of unfolding events of chymotrypsin inhibitor 2, which indicates that detailed interatomic interactions do not play a significant role on the sequence of events (Ferrara *et al.*, 2000a).

Formation of non-local contacts costs much more entropy than formation of contacts between residues close in sequence; this is one explanation of the role of topology in protein folding (Plaxco *et al.*, 1998). In ^{D}PG and GS, a conformation in which the proximal contacts are formed in both β-hairpins has a lower contact order than a conformation in which one β-hairpin is fully folded and the remaining strand is unstructured. However, the folding mechanism of ^{D}PG and GS involves the formation of the latter and not the former. This can be understood by introducing the concept of effective contact order (ECO) (Fiebig & Dill, 1993; Dill *et al.*,

1993). ECO measures the size of the conformational search needed to form a contact considering other already formed contacts. For a two-dimensional square lattice, it was shown that contacts with a low ECO are more likely to form before those with a high ECO, because the loss of configurational entropy is smaller in the former (Fiebig & Dill, 1993). The hydrophobic zipper hypothesis proposed by Dill and co-workers implies that hydrophobic contacts act as constraints that facilitate the formation of other contacts (Dill *et al.*, 1993). The folding of the ^{D}PG and GS peptides is consistent with the hydrophobic zipper hypothesis and the ECO concept, because the formation of the contacts between hydrophobic side-chains close to one turn facilitates the zipping of one hairpin, which is entropically more favorable than formation of contacts at both turns. In other words, the distal contacts in the β-hairpin become local in a topological sense upon formation of the proximal contacts. More recently, 125-mers on a three-dimensional cubic lattice were found to fold using an hydrophobic zipper-like mechanism (Dinner *et al.*, 1996, 1998).

The free-energy surfaces of the ^{D}PG and GS peptides projected onto the native contacts in the 1-2 and 2-3 β-hairpins are similar and both show two main folding pathways. As the sequence identity is only 15%, this cannot be due to the specific interactions but rather to the topology of the folded state. However, the statistically preferred pathway is not the same in the two peptides, which is a sequence effect and highlights the importance of some specific interatomic interactions. Recently, the transition states of the IgG-binding domains of protein G and protein L were mapped out using the protein engineering technique (McCallister *et al.*, 2000). Both proteins contain an α-helix packed against a four-stranded β-sheet, which consists of two symmetrically disposed β-hairpins. It was found for both proteins that only one of the two β-turns is formed at the transition state. However, in protein L, the β-turn in the second β-hairpin is structured at the transition state, whereas it is the one in the first β-hairpin for protein G. For native topologies with a high degree of similarity, it was concluded that several alternative folding routes may exist for different sequences. The present simulation results suggest that folding happens along multiple pathways with a statistical weight that depends on the sequence. The sequence identity between β-hairpin 1-2 and β-hairpin 2-3 is much higher in the GS (67%) than the ^{D}PG peptide (17%). This may explain the higher symmetry of the free-energy surface of GS projected in the ($Q_{1\text{-}2}$,$Q_{2\text{-}3}$) plane than that of ^{D}PG (Figure 5 right). In the 300 K simulation of the ^{D}PG peptide, the C terminus was significantly more frayed than the N terminus. This may be related to the fact that most of the folding transitions (47 of 63 events) at 360 K start by the formation of the β-hairpin 1-2. For the GS peptide, the difference in flexibility between the two termini is much smaller than for ^{D}PG.

Conclusions

The reversible folding of two designed 20-residue sequences having the same three-stranded antiparallel β-sheet topology was simulated with an implicit model of the solvent based on the accessible surface area. A statistically significant sampling of the conformational space was obtained by means of around 50 folding and unfolding events for each peptide. In the two main folding pathways there is first an almost complete stabilization of one β-hairpin followed by the coalescence of the unstructured strand. Average effective energy and free-energy surface are similar for both peptides, despite the sequence dissimilarity. Since the average effective energy has a downhill profile at 360 K, the free-energy barriers are a consequence of the entropic loss involved in the formation of a β-hairpin that represents two-thirds of the chain. The free-energy surface of the β-sheet peptides is completely different from a recently investigated helical peptide of 31 residues, $Y(MEARA)_6$ (Hiltpold *et al.*, 2000). For the helical peptide, the folding free-energy barrier corresponds to the helix nucleation step, and is much closer to the fully unfolded state than for the β-sheet peptides. This indicates that the native topology determines, to a large extent, the free-energy surface. On the other hand, the ^{D}PG peptide has a statistically predominant folding pathway with a sequence of events that is the inverse of that of the most frequent pathway for the GS peptide. Hence, the amino acid sequence and specific interactions between different side-chains determine the most probable folding route.

Despite a sequence identity of only 15%, the 57 residue IgG-binding domain of protein G and of protein L have the same native topology. Their folded state is symmetric and consists mainly of two β-hairpins connected such that the resulting four-stranded β-sheet is antiparallel, apart from the two central strands, which are parallel (McCallister *et al.*, 2000). The Φ value analysis of protein L and of protein G indicates that for proteins with symmetric native structure more than one folding pathway may be consistent with the native-state topology and the selected route depends on the sequence (McCallister *et al.*, 2000). The present simulation results for two antiparallel three-stranded β-sheet peptides (whose sequence identity is also 15%) go beyond the experimental findings for protein G and for protein L. The MD trajectories show the existence of more than one folding pathway for each peptide sequence. Furthermore, the folding events sampled along the MD runs suggest that the higher the degree of symmetry in the sequence, the more uniformly distributed are the statistical weight of the pathways, i.e. the more symmetric is the free-energy surface.

A few trajectories were generated at 330 K for the GS peptide (Ferrara & Caflisch, 2000). Both the average effective energy and free-energy surface that emerged from these simulations are similar to those at 360 K, despite the fact that the GS peptide is significantly less stable at 360 K than at 330 K. This suggests that for these peptides the folding pathways do not depend strongly on the temperature conditions. Moreover, the simulation results indicate that the main unfolding pathway is the reverse of the main folding route, and the same is true for the secondary pathway. The ensemble of denatured states of both peptides consists of a variety of relatively compact structures, whose average radius of gyration is only slightly larger than in the native state.

Materials and Methods

Models

The ^{D}PG and GS peptides were modeled by explicitly considering all heavy-atoms and the hydrogen atoms bound to nitrogen or oxygen atoms (Brooks *et al*., 1983). The aqueous solvent was approximated by an implicit model based on the solvent-accessible surface (Eisenberg & McLachlan, 1986). In this approximation, the mean solvation term is given by:

$$V_{\mathrm{solv}}(\mathbf{r}) = \sum_{i=1}^{M} \sigma_i A_i(\mathbf{r}) \qquad (1)$$

for a molecule having M heavy-atoms wit Cartesian coordinates $\mathbf{r}=(\mathbf{r}_1,\ldots,\mathbf{r}_M)$. $A_i(\mathbf{r})$ is the solvent-accessible surface area of atom i, computed by an approximate analytical expression (Hasel *et al*., 1988) and using a 1.4 Å probe radius. Furthermore, ionic side-chains were neutralized (Lazaridis & Karplus, 1999) and a linear distance-dependent dielectric function ($\varepsilon(r) = 2r$) was used for the electrostatic interactions. The CHARMM PARAM19 default cutoffs for long-range interactions were used, i.e. a shift function (Brooks *et al*., 1983) was employed with a cutoff at 7.5 Å for both the electrostatic and van der Waals terms. The model contains only two σ parameters; one for carbon and sulfur atoms ($\sigma_{C,S} = 0.012$ kcal/(mol Å^2)), and one for nitrogen and oxygen atoms ($\sigma_{N,O} = -0.060$ kcal/(mol Å^2)) (Ferrara *et al*., 2000a). The implicit solvent model has already been used to simulate the reversible folding of two α-helices and a β-hairpin (Ferrara & Caflisch, 2000; Ferrara *et al*., 2000b; Hiltpold *et al*., 2000).

Simulations

All simulations were carried out with the CHARMM program (Brooks *et al*., 1983). Constant-temperature MD simulations were performed by weak coupling to an external bath with a coupling constant of 5 ps (Berendsen *et al*., 1984). The SHAKE algorithm (Ryckaert *et al*., 1977) was used to fix the length of the covalent bonds involving hydrogen atoms, which allows an integration time-step of 2 fs. The nonbonded interactions were updated every ten dynamics steps and coordinate frames were saved every 10 ps.

The following procedure was used to build the initial random conformations for folding simulation. For the ^{D}PG peptide, 2000 structures were generated by randomizing the dihedral angles of the rotatable bonds, followed by 1000 steps of energy minimization. Structures with one or more native contacts were discarded. The 15 structures with the most favorable energies were retained as starting conformations. Their average C$^{\alpha}$ RMSD from the mean NMR model is 8.5 Å. The same was done for the GS peptide with 5000 structures. The 40 with the lowest energies retained as initial conformations have an average C$^{\alpha}$ RMSD from the mean NMR model of 7.4 Å.

Effective energy and free-energy

For the understanding of protein folding, the important role of effective energy and free-energy surfaces, determined by simulations and experiments, has been reviewed recently (Dinner *et al*., 2000). The effective energy is the sum of the intramolecular energy (CHARMM PARAM19 force-field energy) and the solvation free-energy. The latter is approximated by the solvent-accessible surface term of equation (1) and contains the free-energy contribution of the solvent within the approximations of an implicit model of the water molecules. The effective energy does not include the configurational entropy of the peptide, which consists of conformational and vibrational entropy contributions (Lazaridis & Karplus, 1999). The plots on the left of Figure 5 show the values of the effective energy averaged within a bin defined by discretizing the ($Q_{1\text{-}2}$,$Q_{2\text{-}3}$) space.

For a system in thermodynamic equilibrium, the difference in free-energy in going from a state A to a state B is proportional to the natural logarithm of the quotient of the probability of finding the system in state A divided by the probability of state B. Therefore, an error of a factor of 2 in this quotient corresponds to nearly 0.5 kcal/mol. The free-energy surface is projected onto the aforementioned two-dimensional space of progress variables (plots on the right of Figure 5), as well as on the ($Q_{1\text{-}2}^{\mathrm{prox}}$,$Q_{1\text{-}2}^{\mathrm{dist}}$) and ($Q_{2\text{-}3}^{\mathrm{prox}}$,$Q_{2\text{-}3}^{\mathrm{dist}}$) planes (Figure 6), by using an arbitrarily chosen reference bin as the denominator of the probability quotient.

Cluster analysis

The method for the cluster analysis is based on structural similarity (Daura *et al*., 1999). The C$^{\alpha}$ RMSD is calculated for each pair of structures after optimal superposition. The number of neighbors is then calculated for each conformation using a C$^{\alpha}$ RMSD cutoff of 2.0 Å. The conformation with the highest number of neighbors is considered as the center of the first cluster. All the neighbors of this conformation are removed from the ensemble of conformations. The center of the second cluster is then determined in the same way as for the first cluster, and this procedure is repeated until each structure is assigned to a cluster. To determine the number of hydrogen bonds in the centers of the most populated clusters, a hydrogen bond is considered to be formed if the O$\cdots$H distance is smaller than 2.6 Å.

Folding and unfolding times

The folding time is defined as the time-interval during which the fraction of native contacts (Q) varies from a value smaller than 0.15 to a value larger than 0.85. The beginning and the end of the interval are defined as the

time when the simulation reaches a Q value of less than 0.15 and of more than 0.85, respectively, for the first time. Therefore, this definition of the folding time corresponds to the first passage time. Similarly, the unfolding time is defined as the time-interval during which Q varies from a value larger than 0.85 to a value smaller than 0.15.

Note added in proof

After this article was submitted, an NMR study and thermodynamics analysis of another designed 24-residue three-stranded antiparallel β-sheet was published (Griffiths-Jones, S. R. & Searle, M. S. (2000). *J. Am. Chem. Soc.* **122**, 8350-8356). The experimental data indicate the presence of two folding pathways starting by the almost complete formation of either of the two β-hairpins, an agreement with the present simulation study.

Acknowledgments

We thank Professor S. H. Gellman, Dr F. Syud and Dr M. A. Jiménez for providing the NMR conformations. We thank one referee, who brought the concept of effective contact order (Fiebig & Dill, 1993) to our attention. This work was supported by the Swiss National Science Foundation (grant no. 31-53604.98) and the Théodore Ott Foundation.

References

Abkevich, V. I., Gutin, A. M. & Shakhnovich, E. I. (1994). Specific nucleus as the transition state for protein folding: evidence from the lattice model. *Biochemistry,* **33**, 10026-10036.

Baker, D. (2000). A surprising simplicity to protein folding. *Nature,* **405**, 39-42.

Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684-3690.

Bonvin, A. M. J. J. & van Gunsteren, W. F. (2000). β-hairpin stability and folding: molecular dynamics studies of the first β-hairpin of tendamistat. *J. Mol. Biol.* **296**, 255-268.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983). CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4**, 187-217.

Bursulaya, B. D. & Brooks, C. L., III (1999). Folding free-energy surface of a three-stranded β-sheet protein. *J. Am. Chem. Soc.* **121**, 9947-9951.

Caflisch, A. & Karplus, M. (1994). Molecular dynamics simulation of protein denaturation: solvation of the hydrophobic cores and secondary structure of barnase. *Proc. Natl Acad. Sci. USA,* **91**, 1746-1750.

Caflisch, A. & Karplus, M. (1995). Acid and thermal denaturation of barnase investigated by molecular dynamics simulations. *J. Mol. Biol.* **252**, 672-708.

Crane, J. C., Koepf, E. K., Kelly, J. & Gruebele, M. (2000). Mapping the transition state of the WW domain β-sheet. *J. Mol. Biol.* **298**, 283-292.

Daura, X., van Gunsteren, W. F. & Mark, A. E. (1999). Folding-unfolding thermodynamics of a β-heptapeptide from equilibrium simulations. *Proteins: Struct. Funct. Genet.* **34**, 269-280.

De Alba, E., Santoro, J., Rico, M. & Jiménez, M. A. (1999). De novo design of a monomeric three-stranded antiparallel β-sheet. *Protein Sci.* **8**, 854-865.

Dill, K. A., Fiebig, K. M. & Chan, H. S. (1993). Cooperativity in protein-folding kinetics. *Proc. Natl Acad. Sci. USA,* **90**, 1942-1946.

Dinner, A. R., Šali, A. & Karplus, M. (1996). The folding mechanism of larger model proteins: role of native structure. *Proc. Natl Acad. Sci. USA,* **93**, 8356-8361.

Dinner, A. R., So, S. S. & Karplus, M. (1998). Use of quantitative structure property relationships to predict the folding ability of model proteins. *Proteins: Struct. Funct. Genet.* **33**, 177-203.

Dinner, A. R., Lazaridis, T. & Karplus, M. (1999). Understanding β-hairpin formation. *Proc. Natl Acad. Sci. USA,* **96**, 9068-9073.

Dinner, A. R., Šali, A., Smith, L. J., Dobson, C. M. & Karplus, M. (2000). Understanding protein folding via free-energy surfaces from theory and experiment. *Trends Biochem. Sci.* **25**, 331-339.

Dobson, C. M. & Karplus, M. (1999). The fundamentals of protein folding: bringing together theory and experiment. *Curr. Opin. Struct. Biol.* **9**, 92-101.

Dobson, C. M., Šali, A. & Karplus, M. (1998). Protein folding: a perspective from theory and experiment. *Angew. Chem. Int. Ed.* **37**, 869-893.

Duan, Y. & Kollman, P. A. (1998). Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science,* **282**, 740-744.

Eisenberg, D. & McLachlan, A. D. (1986). Solvation energy in protein folding and binding. *Nature,* **319**, 199-203.

Ferrara, P. & Caflisch, A. (2000). Folding simulations of a three-stranded antiparallel β-sheet peptide. *Proc. Natl Acad. Sci. USA,* **97**, 10780-10785.

Ferrara, P., Apostolakis, J. & Caflisch, A. (2000a). Computer simulations of protein folding by targeted molecular dynamics. *Proteins: Struct. Funct. Genet.* **39**, 252-260.

Ferrara, P., Apostolakis, J. & Caflisch, A. (2000b). Thermodynamics and kinetics of folding of two model peptides investigated by molecular dynamics simulations. *J. Phys. Chem. ser. B,* **104**, 5000-5010.

Fiebig, K. M. & Dill, K. A. (1993). Protein core assembly processes. *J. Chem. Phys.* **98**, 3475-3487.

Grantcharova, V. P., Riddle, D. S. & Baker, D. (2000). Long-range order in the src SH3 folding transition state. *Proc. Natl Acad. Sci. USA,* **97**, 7084-7089.

Griffiths-Jones, S. R. & Searle, M. S. (2000). Structure, folding, and energetics of cooperative interactions between the β-strands of a de novo designed three-stranded antiparallel β-sheet peptide. *J. Am. Chem. Soc.* **122**, 8350-8356.

Hasel, W., Hendrickson, T. F. & Still, W. C. (1988). A rapid approximation to the solvent accessible surface areas of atoms. *Tetrahedron Comput. Methodol.* **1**, 103-116.

Hiltpold, A., Ferrara, P., Gsponer, J. & Caflisch, A. (2000). Free energy surface of the helical peptide $Y(MEARA)_6$. *J. Phys. Chem. ser. B,* **104**, 10080-10086.

Hutchinson, E. G. & Thornton, J. M. (1994). A revised set of potentials for β-turn formation in proteins. *Protein Sci.* **3**, 2207-2216.

Imperiali, B. & Ottesen, J. (1999). Uniquely folded miniprotein motifs. *J. Pept. Res.* **54**, 177-184.

Lacroix, E., Kortemme, T., de la Paz, M. & Serrano, L. (1999). The design of linear peptides that fold as monomeric beta-sheet structures. *Curr. Opin. Struct. Biol.* **9**, 487-493.

Lazaridis, T. & Karplus, M. (1999). Effective energy function for proteins in solution. *Proteins: Struct. Funct. Genet.* **35**, 133-152.

Martinez, J. C. & Serrano, L. (1999). The folding transition state between SH3 domains is conformationally restricted and evolutionarily conserved. *Nature Struct. Biol.* **6**, 1010-1016.

McCallister, E. L., Alm, E. & Baker, D. (2000). Critical role of β-hairpin formation in protein G folding. *Nature Struct. Biol.* **7**, 669-673.

Michnick, S. W. & Shakhnovich, E. I. (1998). A strategy for detecting the conservation of folding-nucleus residues in protein superfamilies. *Fold. Des.* **3**, 239-251.

Mirny, L. A. & Shakhnovich, E. I. (1999). Universally conserved positions in protein folds: reading evolutionary signals about stability, folding kinetics and function. *J. Mol. Biol.* **291**, 177-196.

Mirny, L. A., Abkevich, V. I. & Shakhnovich, E. I. (1998). How evolution makes proteins fold quickly. *Proc. Natl Acad. Sci. USA,* **95**, 4976-4981.

Muñoz, V., Thompson, P. A., Hofrichter, J. & Eaton, W. A. (1997). Folding dynamics and mechanism of β-hairpin formation. *Nature,* **390**, 196-199.

Pande, V. S. & Rokhsar, D. S. (1999). Molecular dynamics simulations of unfolding and refolding of a β-hairpin fragment of protein G. *Proc. Natl Acad. Sci. USA,* **96**, 9062-9067.

Plaxco, K., Simons, K. & Baker, D. (1998). Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.* **277**, 985-994.

Ryckaert, J. P., Ciccotti, G. & Berendsen, H. J. C. (1977). Numerical integration of the Cartesian equation of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comp. Phys.* **23**, 327-341.

Schaefer, M., Bartels, C. & Karplus, M. (1998). Solution conformation and thermodynamics of structured peptides: molecular dynamics simulation with an implicit solvation model. *J. Mol. Biol.* **284**, 835-848.

Schenck, H. L. & Gellman, S. H. (1998). Use of a designed triple-stranded antiparallel β-sheet to probe β-sheet cooperativity in aqueous solution. *J. Am. Chem. Soc.* **120**, 4869-4870.

Sheinerman, F. B. & Brooks, C. L., III (1998). Calculations on folding of segment B1 of streptococcal protein G. *J. Mol. Biol.* **278**, 439-456.

Stanger, H. E. & Gellman, S. H. (1998). Rules for antiparallel β-sheet design: D-Pro-Gly is superior to L-Asn-Gly for β-hairpin nucleation. *J. Am. Chem. Soc.* **120**, 4236-4237.

Tsui, V. & Case, D. A. (2000). Molecular dynamics simulations of nucleic acids with a generalized Born solvation model. *J. Am. Chem. Soc.* **122**, 2489-2498.

Wang, H. & Sung, S. (2000). Molecular dynamics simulations of three-strand β-sheet folding. *J. Am. Chem. Soc.* **122**, 1999-2009.

Edited by F. Cohen

(Received 17 August 2000; received in revised form 4 December 2000; accepted 5 December 2000)