PROTEINS: Structure, Function, and Genetics 34:269–280 (1999)

RESEARCH ARTICLES

Folding–Unfolding Thermodynamics of a β-Heptapeptide From Equilibrium Simulations

Xavier Daura, Wilfred F. van Gunsteren,* and Alan E. Mark
Laboratorium für Physikalische Chemie, ETH Zentrum, Zürich, Switzerland

ABSTRACT **The thermodynamics of folding and unfolding of a β-heptapeptide in methanol solution has been studied at four different temperatures, 298 K, 340 K, 350 K, and 360 K, by molecular dynamics simulation. At each of these temperatures, the 50-ns simulations were sufficient to generate an equilibrium distribution between a relatively small number of conformations (~10^2), showing that, even above the melting temperature (~340 K), the peptide does not randomly sample conformational space. The free energy of folding and the free energy difference between pairs of conformations have been calculated from their relative populations. The experimentally determined folded conformation at 298 K, a left-handed 3_1-helix, is at each of the four temperatures the predominant conformation, with its probability and average lifetime decreasing with increasing temperature. The most common intermediates of folding and unfolding are also the same at the four temperatures. Paths and rates of interconversion between different conformations have been determined. It has been found that folding can occur through multiple pathways, not necessarily downhill in free energy, although the final step involves a reduced number of intermediates. Proteins 1999;34:269–280.** © 1999 Wiley-Liss, Inc.

Key words: molecular dynamics; computer simulation; GROMOS; β-peptides; peptide folding; folding pathways; folding free energy; unfolded conformations

INTRODUCTION

The three-dimensional structure adopted by a peptide is dependent on its environment, the temperature, the pressure, and the solvent or molecular surroundings. Under any given set of conditions the structure of a peptide will not be unique, but instead there will be an equilibrium distribution between different structures. If a particular conformation predominates under a given set of conditions the peptide is commonly referred to as being folded. This folded conformation can (in favourable circumstances) be determined by fitting observed experimental data to one or a small set of structures. The experimental data used in such structure determination procedures corresponds by necessity to an ensemble and time average over different structures. If the experimental data cannot be adequately fitted to a single or a small number of structures the peptide is referred to as unfolded or unstructured. This does not, however, mean that the peptide randomly samples conformational space or that the peptide does not adopt a small number of preferred conformations. It simply indicates that the spatial and time resolution of the experimental technique used to determine the peptide structure, for example, NMR spectroscopy, does not permit one to distinguish between different structures coexisting in the sample during the time of the measurement. In order to investigate the conformational dynamics of an isolated peptide we must turn to theoretical approaches. Molecular dynamics (MD) simulation techniques currently offer the perhaps only means of studying at atomic resolution the dynamic equilibrium between the different conformations accessible to a peptide under realistic conditions. The peptide may be placed in a given environment and the time evolution of the system simulated until an equilibrium distribution between different conformations is established. Once an equilibrium distribution has been established, statistical mechanics can be used to analyse the thermodynamics of the system. For example, from the probability of finding the peptide in a given conformation it is possible to estimate the free energy of folding, or more generally, the difference in free energy between any two conformations. It is also possible to examine preferred paths between any two conformations, and to detect common folding or unfolding intermediates.

Such analysis is not dependent on the whole of conformational space accessible to the peptide being sampled in the simulation, which is, in practice, not possible. Only the ratio between the populations of different conformations is required. The statistical significance of the calculated free

*Correspondence to: Wilfred F. van Gunsteren, Laboratorium für Physikalische Chemie, ETH-Zürich, ETH-Zentrum, CH-8092 Zürich, Switzerland. E-mail:wfvgn@igc.phys.chem.ethz.ch
Received 10 August 1998; Accepted 19 October 1998

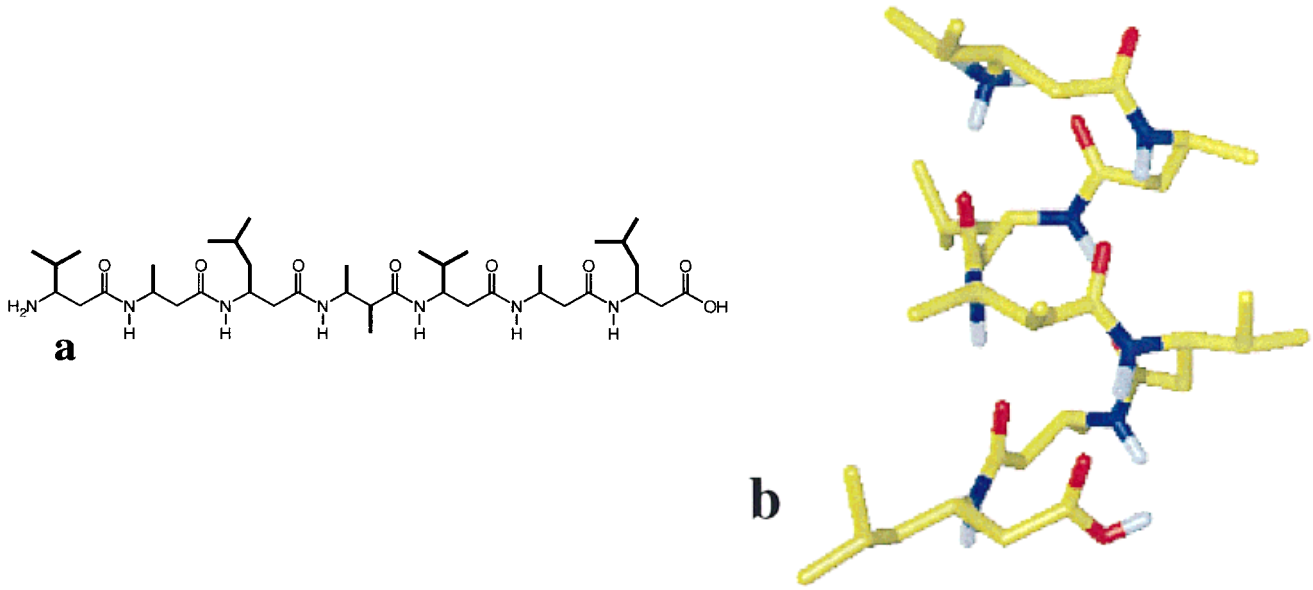

Fig. 1. **a:** Structural formula of the β-heptapeptide studied (H-β-HVal-β-HAla-β-HLeu-(*S,S*)-β-HAla(αMe)-β-HVal-β-HAla-β-HLeu-OH). Note, in the simulations both end groups were protonated in line with experimental data. **b:** Model conformation derived from NMR data at 298 K. (From Reference 2, with permission.)

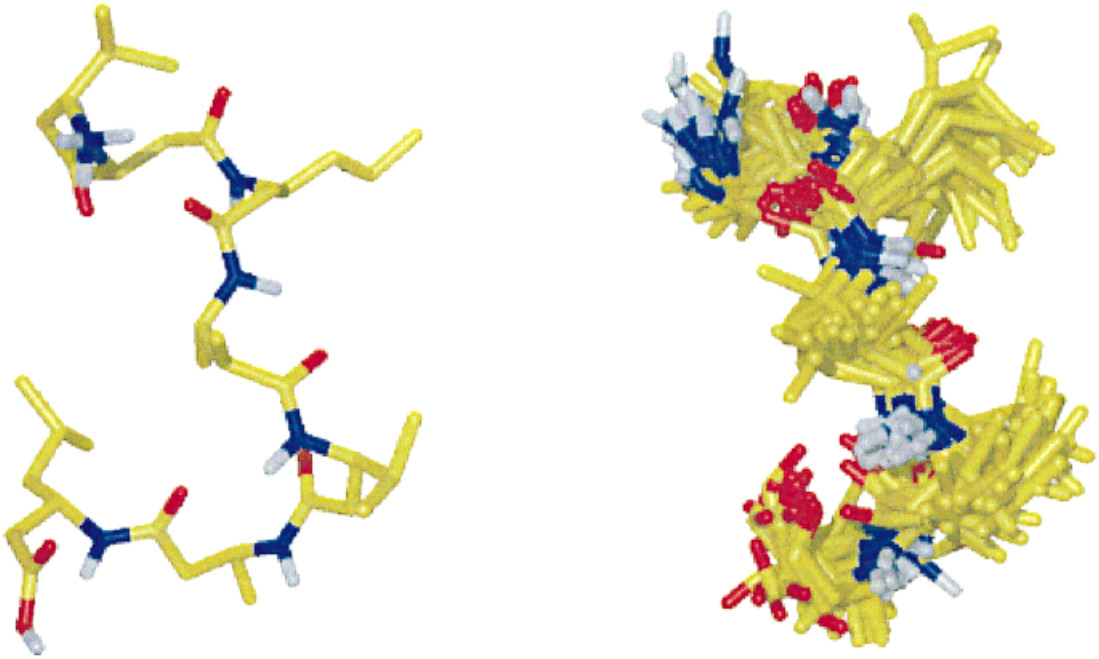

Fig. 2. Central structure of cluster 36 at 360 K, and superposition (residues 2–6) of the structures belonging to this cluster. Of the 20 structures belonging to this cluster, 13 have a backbone atom-positional RMSD (residues 2–6) from the central structure of 0.09 nm, and the maximum RMSD between any two structures is 0.16 nm. It is thus an example of a cluster with a high spread of member structures.

energy of folding, for example, is thus primarily determined by the number of folding–unfolding events occurring during the simulation. The simulation time required to calculate accurately the free energy difference between any two conformations of the peptide depends on the rates of interconversion between each of these two conformations and other conformations and not on the degree to which the conformational space accessible to the peptide has been sampled. Nevertheless, the nature of the ensemble of unfolded conformations, whether this encompasses a large or a small part of the theoretical conformational space of the peptide, does determine whether, in the MD simulations, an equilibrium can be established between different unfolded conformations and the reliability

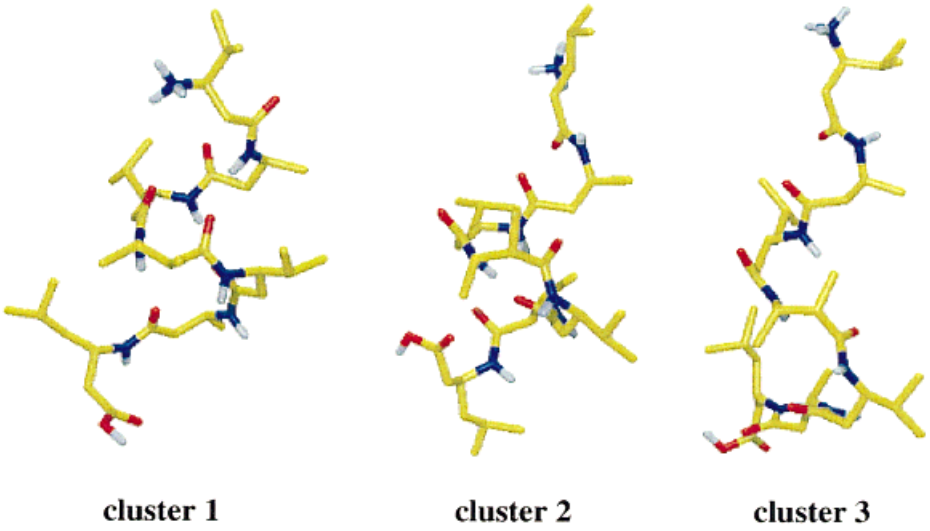

Fig. 3. Central structures of the three most populated clusters of structures at 298 K.

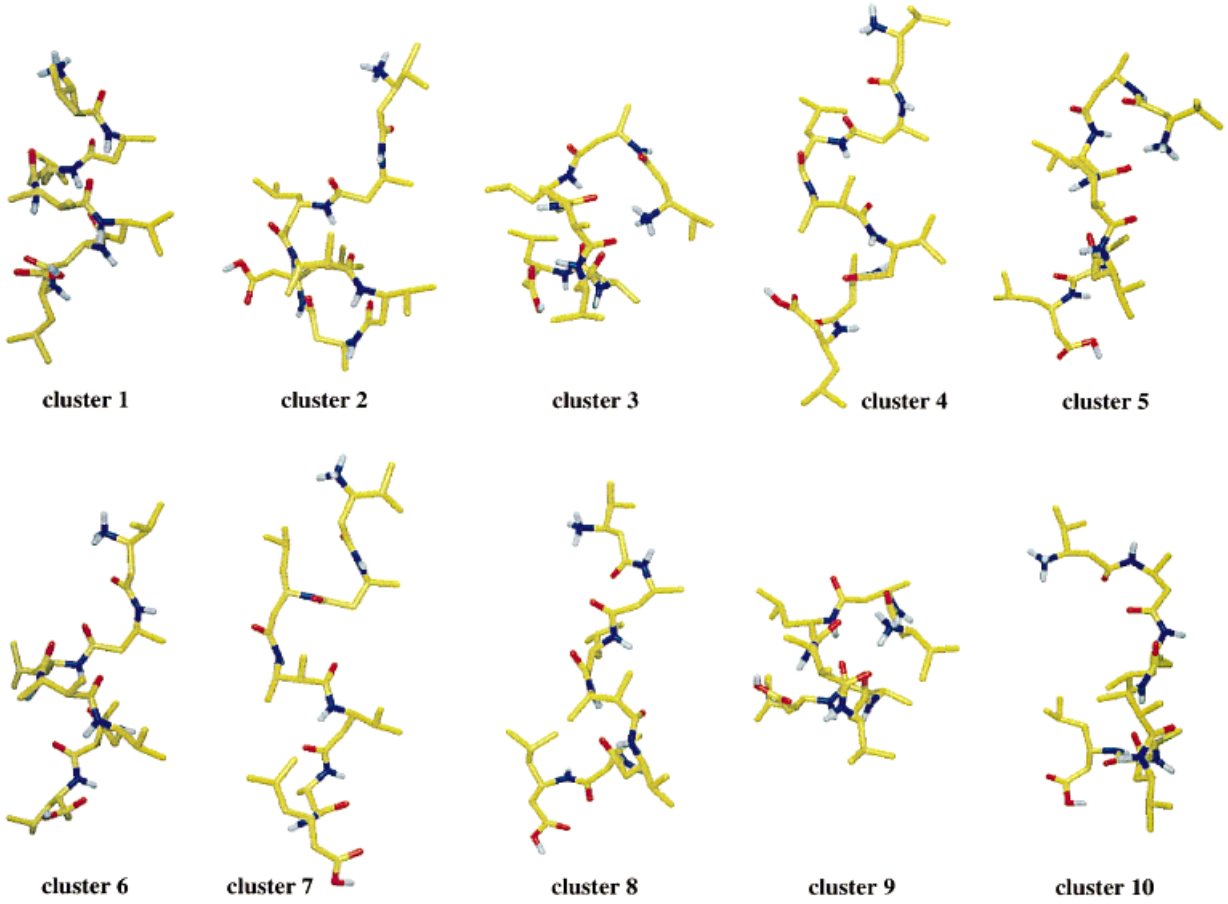

Fig. 4. Central structures of the 10 most populated clusters of structures at 340 K.

to which free energy differences can be determined. As long as the ensemble of unfolded conformations covers a relatively small proportion of the conformational space accessible to the peptide, that is, the peptide repeatedly visits a relatively small number of conformations on a time scale short compared with the simulation time scale, the necessary statistics on the interconversion between different conformations will be obtained.

In this report we present MD simulation studies on the reversible folding of a β-heptapeptide (Fig. 1a) in methanol

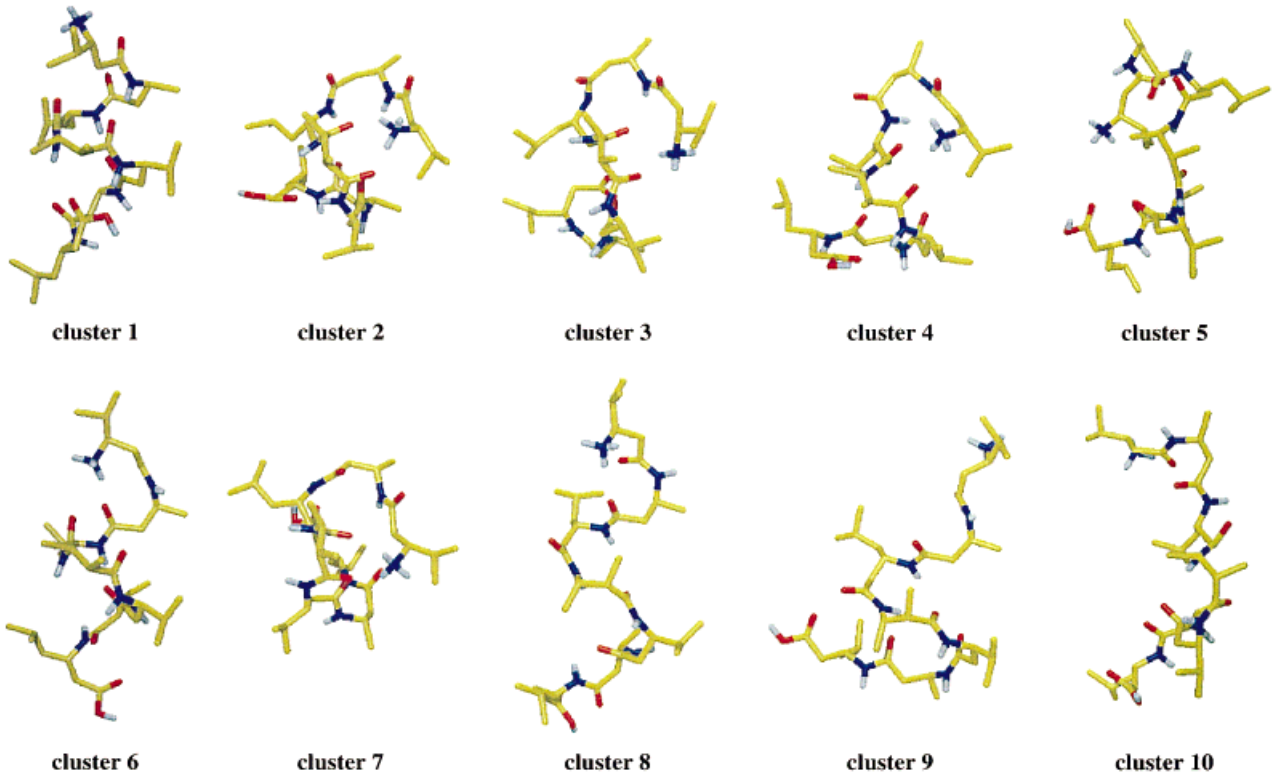

Fig. 5. Central structures of the 10 most populated clusters of structures at 350 K.

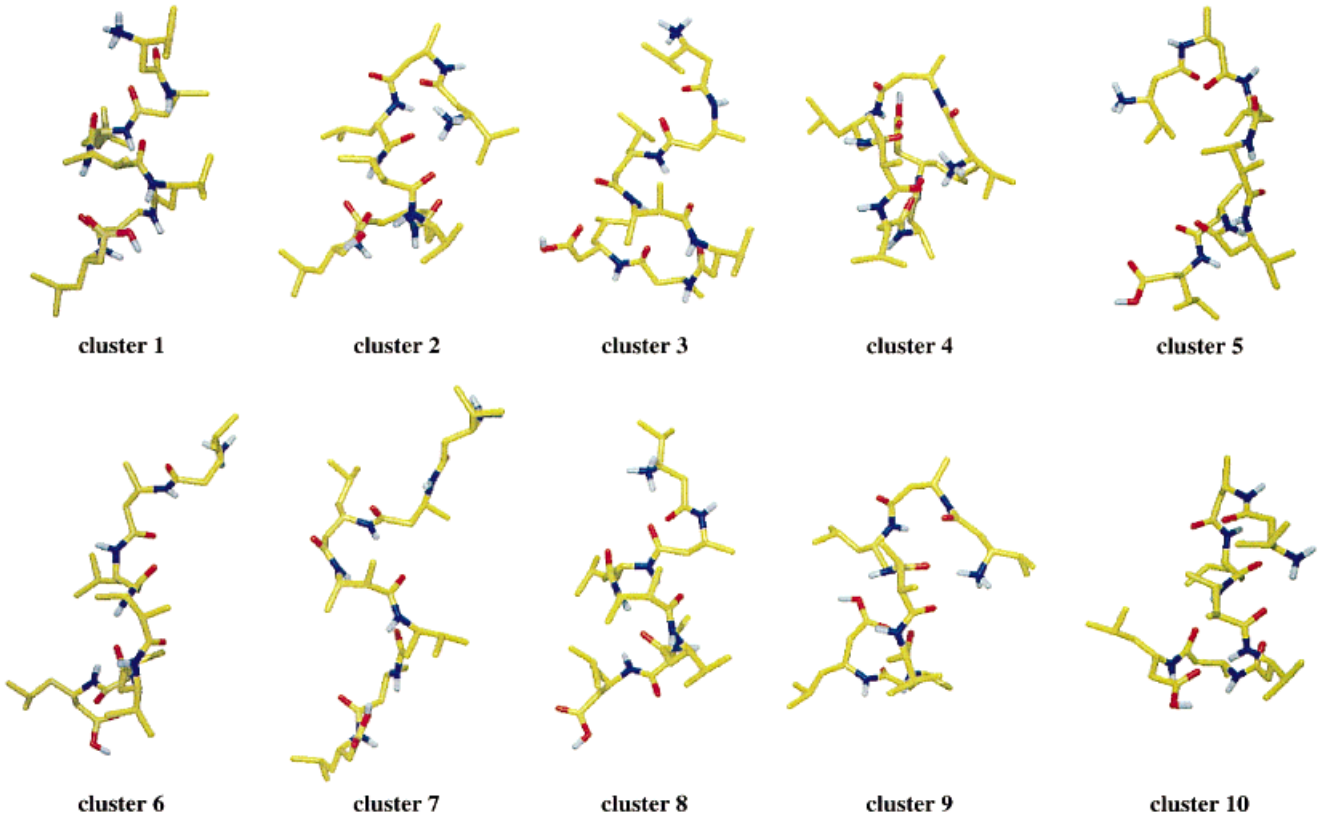

Fig. 6. Central structures of the 10 most populated clusters of structures at 360 K.

TABLE I. Clusters With a Minimum of 20 Member Structures at 298 K, 340 K, 350 K, and 360 K†

	298 K			340 K			350 K			360 K		
Cluster	RMSD (nm)	Members	ΔG (kJ mol^{-1})	RMSD (nm)	Members	ΔG (kJ mol^{-1})	RMSD (nm)	Members	ΔG (kJ mol^{-1})	RMSD (nm)	Members	ΔG (kJ mol^{-1})
1	0.05	4855	0.0	0.06	2503	0.0	0.05	1937	0.0	0.06	1242	0.0
2	0.14	80	10.2	0.13	298	6.0	0.21	1063	1.7	0.19	581	2.3
3	0.20	38	12.0	0.22	249	6.5	0.23	259	5.9	0.14	332	3.9
4				0.17	197	7.2	0.17	182	6.9	0.24	255	4.7
5				0.25	156	7.8	0.33	131	7.8	0.33	187	5.7
6				0.15	122	8.5	0.15	110	8.3	0.26	127	6.8
7				0.25	102	9.0	0.26	100	8.6	0.19	123	6.9
8				0.20	72	10.0	0.17	86	9.1	0.16	106	7.4
9				0.22	66	10.3	0.11	74	9.5	0.24	85	8.0
10				0.30	65	10.3	0.29	61	10.1	0.21	79	8.2
11				0.32	51	11.0	0.31	54	10.4	0.33	78	8.3
12				0.25	51	11.0	0.26	51	10.6	0.18	69	8.6
13				0.25	51	11.0	0.19	45	10.9	0.27	66	8.8
14				0.23	43	11.5	0.22	42	11.1	0.33	59	9.1
15				0.18	38	11.8	0.34	42	11.1	0.34	56	9.3
16				0.22	37	11.9	0.23	42	11.1	0.22	51	9.6
17				0.31	36	12.0	0.30	39	11.4	0.33	49	9.7
18				0.19	33	12.2	0.29	38	11.4	0.30	48	9.7
19				0.18	29	12.6	0.29	35	11.7	0.29	47	9.8
20				0.29	29	12.6	0.19	32	11.9	0.19	46	9.9
21				0.32	29	12.6	0.22	26	12.5	0.26	43	10.1
22				0.30	27	12.8	0.20	24	12.8	0.30	43	10.1
23				0.32	26	12.9	0.22	20	13.3	0.26	42	10.1
24				0.28	26	12.9	0.27	20	13.3	0.20	41	10.2
25				0.31	25	13.0				0.31	35	10.7
26				0.30	24	13.1				0.20	34	10.8
27				0.30	23	13.2				0.34	34	10.8
28				0.23	22	13.4				0.19	34	10.8
29				0.08	22	13.4				0.24	31	11.0
30				0.22	21	13.5				0.09	31	11.0
31										0.26	28	11.3
32										0.31	26	11.6
33										0.29	23	11.9
34										0.18	23	11.9
35										0.36	21	12.2
36										0.31	20	12.4

†The backbone atom-positional RMSD (residues 2–6) between the central structures of the cluster and the NMR model structure, the number of member structures in the cluster, and the free energy difference with cluster 1 are shown.

at four different temperatures.[1] This peptide adopts a left-handed 3_1-helix (Fig. 1b) in methanol solution at room temperature,[2] and MD simulations using the GROMOS 43A1 force field[3] have been previously shown to reproduce much of the available experimental data on this system.[1,4] By simulating on a time scale that is long compared with the average lifetime of any specific conformation of the peptide it has been possible to establish a dynamic equilibrium between different conformational states at a series of temperatures. The structures sampled in the simulations have been clustered and these clusters then used to analyse the thermodynamics of folding and unfolding in the system.

METHODS

Simulations

Four 50-ns MD simulations were performed using the GROMOS96 package of programs[3] and the GROMOS96 43A1 force field.[3] The dynamics of the β-heptapeptide (Fig. 1) in methanol were studied at a series of temperatures, 298 K, 340 K, 350 K, and 360 K, at 1 atm pressure and with periodic boundary conditions. The temperature and pressure were maintained by weak coupling to an external bath. The initial structure of the peptide for the simulations at 298 K, 340 K, and 350 K was the 3_1-helical fold shown in Figure 1b. The system contained the β-heptapeptide and 962 methanol molecules in a rectangular box.[1] For the simulation at 360 K the peptide was initially fully extended (all backbone dihedral angles set to 180°), and the system contained the β-heptapeptide and 1,778 methanol molecules in a truncated octahedron.[1] A twin-range cutoff of 0.8 nm/1.4 nm was used for all nonbonded interactions. The shortest distance peptide-wall was initially 1.4 nm. In all simulations the (periodic) box was sufficiently large that a totally extended conformation of the β-heptapeptide would not span its shortest axis. Full

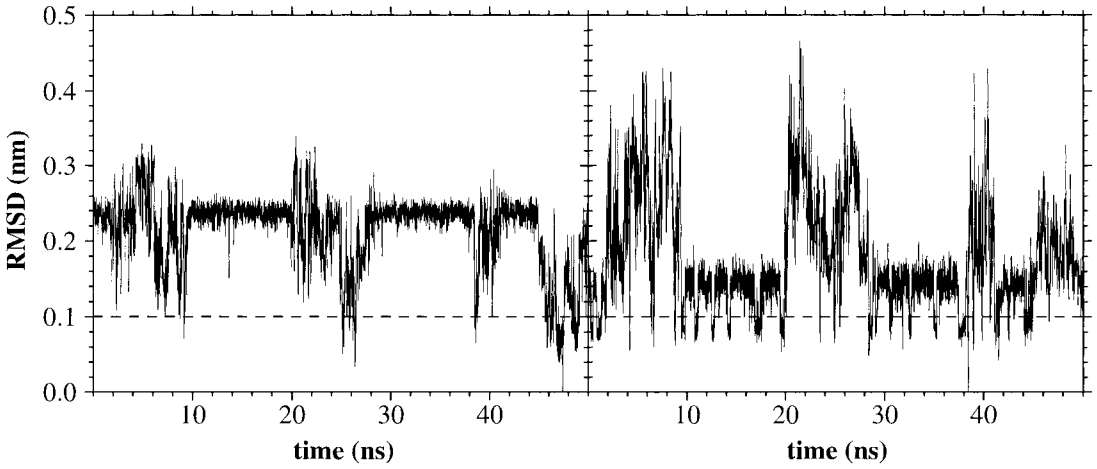

Fig. 7. Backbone atom-positional RMSD (residues 2–6) from the central structure of cluster 5 (left-hand plot) and cluster 6 (right-hand plot) at 340 K as a function of time. Structures with an RMSD below the dashed line potentially belong to the same cluster.

details of the setup of the simulations have been described previously.[1]

Cluster Analysis

Structures of the peptide were extracted from the trajectories at 0.01 ns intervals for analysis (a total of 5,000 structures per simulation). Structures separated in time by less than 0.01 ns are, in general, correlated. The clustering was performed in cartesian space. For each pair of structures a least-squares translational and rotational fit was performed using the backbone (N, C_β, C_α, C) atoms of residues 2 to 6, and the atom-positional root-mean-square difference (RMSD) for this set of atoms was calculated. Using as criterion of similarity of two structures an RMSD $\leq$ 0.1 nm for the backbone atoms of residues 2–6, the number of neighbours (i.e., structures satisfying the similarity criterion) for each of the structures in the initial pool of 5,000 was determined. The structure with the highest number of neighbours was then taken as the centre of the first cluster of structures. All the structures belonging to this cluster were thereafter removed from the pool. For each of the remaining structures the number of neighbours was again computed. The structure with the most neighbours became the centre of the second cluster of structures. Structures belonging to this second cluster were then also removed from the pool. This process was iterated until all structures were assigned to a cluster. This type of clustering favours the most populated cluster, that is, the folded conformation, and ensures a minimum difference in atom-positional RMSD between centres of clusters equal to the similarity criterion. It also results in many clusters with only one member. These are not necessarily conformations that have been sampled only once in the simulation. In general they have similar structures (RMSD $\leq$ 0.1 nm) too, but lie just outside other clusters. This algorithm is similar in effect to that used by Karpen et al.[5] in that the clusters generated are subject to the constraint that no member of the cluster is more than a specified distance from the cluster's central member. An illustration of the spread of structures within a cluster is given in Figure 2.

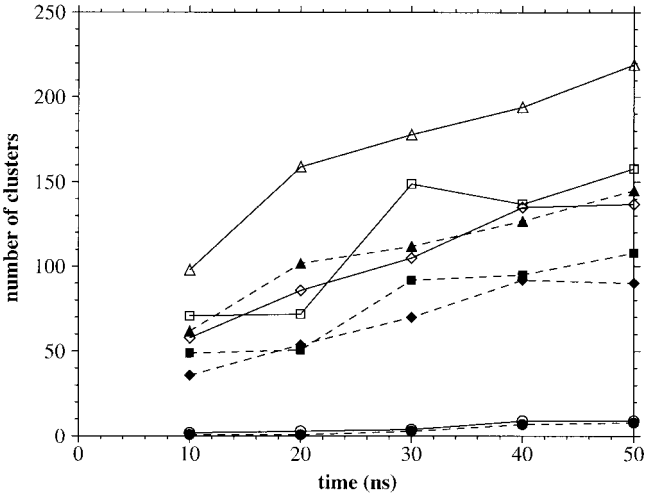

Fig. 8. Number of clusters as a function of time. Solid lines: total number of clusters; dashed lines: number of clusters with more than one member structure; circles: number of clusters at 298 K; squares: number of clusters at 340 K; diamonds: number of clusters at 350 K; triangles: number of clusters at 360 K.

Free Energy

Given a system in thermodynamic equilibrium, the change in free energy on going from a state A of the system to a state B (e.g., from unfolded to folded) can be calculated as

$$\Delta G_{A\rightarrow B} = -k_B T \ln \frac{p_B}{p_A}$$

where k_B is the Boltzmann constant, T is the temperature, and p_A and p_B are the relative probabilities of finding the system in state A and state B, respectively. A and B in this study refer to clusters or groups of clusters, and p_A and p_B

TABLE II. Comparison of Clusters With a Minimum of 20 Member Structures at 298 K, 340 K, 350 K, and 360 K†

Temp1 (K)	Clust1	Temp2 (K)	Clust2	RMSD (nm)	Temp1 (K)	Clust1	Temp2 (K)	Clust2	RMSD (nm)
298	1	340	1	0.04	340	9	360	20	0.06
298	1	350	1	0.04	340	10	350	10	0.07
298	1	360	1	0.04	340	13	350	3	0.07
298	2	340	6	0.05	340	13	360	9	0.05
298	2	350	6	0.05	340	13	360	31	0.07
298	2	360	8	0.06	340	15	360	34	0.07
298	3	340	8	0.06	340	29	360	30	0.05
298	3	350	8	0.07	350	1	360	1	0.03
340	1	350	1	0.02	350	2	360	4	0.07
340	1	360	1	0.03	350	2	360	20	0.07
340	2	350	9	0.06	350	3	360	9	0.04
340	2	360	3	0.04	350	4	360	2	0.05
340	3	350	2	0.05	350	4	360	10	0.07
340	3	360	4	0.05	350	5	360	5	0.07
340	4	350	8	0.05	350	6	360	8	0.07
340	4	360	7	0.05	350	7	360	4	0.06
340	5	360	6	0.05	350	9	360	3	0.06
340	6	350	6	0.03	350	10	360	33	0.07
340	6	360	8	0.06	350	11	360	11	0.06
340	9	350	2	0.07	350	19	360	31	0.07
340	9	350	16	0.05					

†The temperature of the simulation (temp1 and temp2 for the two clusters that are compared), the cluster number (clust1 and clust2 for the two clusters that are compared), and the backbone atom-positional RMSD (residues 2–6) between the central member structures of clust1 and clust2 are shown for those pairs of clusters with RMSD <0.08 nm.

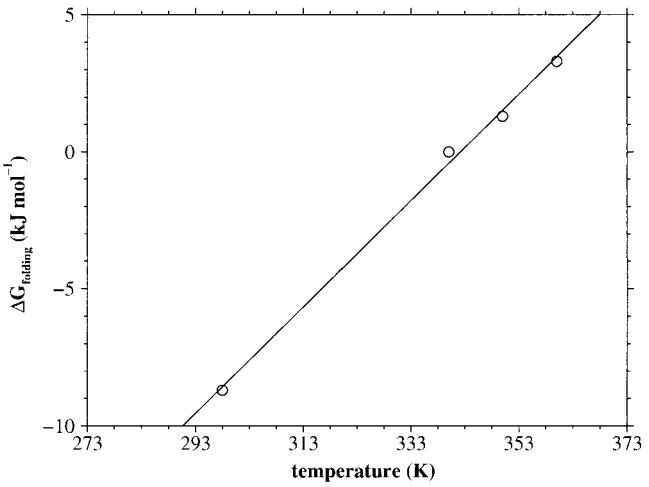

Fig. 9. Free energy of folding as a function of temperature. Circles: simulation data; solid line: linear regression (slope: 0.194 kJ mol^{-1} K^{-1}; intercept: −66.383 kJ mol^{-1}; correlation coefficient: 0.998).

TABLE III. Average Lifetime of Clusters with a Minimum of 50 Member Structures at 298 K, 340 K, 350 K, and 360 K

	Average lifetime (ns)			
Cluster	298 K	340 K	350 K	360 K
1	1.200	0.417	0.208	0.195
2	0.026	0.040	0.098	0.052
3		0.034	0.027	0.030
4		0.021	0.067	0.037
5		0.031	0.036	0.022
6		0.021	0.019	0.020
7		0.020	0.019	0.018
8		0.017	0.028	0.026
9		0.020	0.019	0.019
10		0.021	0.020	0.018
11		0.016	0.015	0.016
12		0.032	0.016	0.014
13		0.022		0.060
14				0.014
15				0.018
16				0.014

can simply be taken as the number of structures in clusters A and B, respectively.

Transition Rates

Given two conformations (clusters of structures) A and B, the transition rate from A to B (unidirectional) is calculated as the average number of times per nanosecond in the simulation that conformation A evolves within a 0.01-ns time interval (the time resolution of the analysis) to conformation B. The transition rate between A and B (bidirectional) is calculated as the average of the transition rates from A to B and from B to A.

RESULTS

Clusters

Clusters with 20 or more member structures (≥0.4% of the ensemble of 5,000 structures) are listed in Table I for each of the four simulations. The central structures of the first 10 clusters (first three at 298 K) at each temperature are shown in Figures 3 to 6. At 298 K a total of nine clusters was found, of which one contained only a single structure. The first cluster at 298 K incorporates approximately 97% of the ensemble and corresponds to the 3_1-helical fold. The central structure of this cluster has a

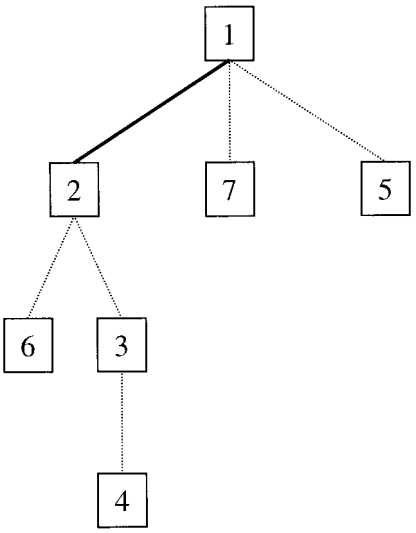

Fig. 10. Paths between clusters of structures at 298 K. Clusters are represented by squares containing the cluster number. Connections between clusters represent paths. Only those paths that have been sampled at least twice in each direction are shown. The approximate average (forth and back) number of times that a path has been sampled (within 50 ns) is indicated by a line code. Dotted line: 2 to 5 times; dot-dashed line: 6 to 10 times; dashed line: 11 to 15 times; thin-solid line: 16 to 20 times; thick-solid line: more than 20 times.

backbone atom-positional RMSD from the NMR model structure of 0.05 nm (residues 2–6). At 340 K a total of 158 clusters was found, of which 50 contained only a single structure. The first cluster incorporates approximately 50% of the ensemble and again corresponds to the 3_1-helical fold. The central structure of this cluster has a backbone atom-positional RMSD from the NMR model structure of 0.06 nm (residues 2–6). Clusters 2–11 incorporate 55% of the ensemble of unfolded structures (clusters 2–158), while for clusters 2–30 (Table I) this percentage increases to 79%. At 350 K a total of 137 clusters was found, of which 47 contained only a single structure. The first cluster, which again has the 3_1-helical fold, incorporates approximately 39% of the ensemble. The central structure of this cluster has a backbone atom-positional RMSD from the NMR model structure of 0.05 nm (residues 2–6). Clusters 2–11 incorporate 69% of the ensemble of unfolded structures (clusters 2–137), while for clusters 2–24 (Table I) this percentage increases to 84%. At 360 K a total of 219 clusters was found, of which 74 contained only a single structure. The first cluster, which again has the 3_1-helical fold, incorporates approximately 25% of the ensemble. The central structure of this cluster has a backbone atom-positional RMSD from the NMR model structure of 0.06 nm (residues 2–6). Clusters 2–11 incorporate 52% of the ensemble of unfolded structures (clusters 2–219), while for clusters 2–36 (Table I) this percentage increases to 79%.

As with the folded (predominant) conformation, which is folding and unfolding several times in each of the simulations, a particular unfolded conformation is in general found repeatedly at different times in the simulation, that is, the structures constituting a cluster are not necessarily consecutive in time. To illustrate this, the backbone atom-positional RMSD (residues 2 –6) from the central structure of two example clusters at 340 K is plotted as a function of simulation time in Figure 7.

A slight difference should be noted between the clustering presented here and that presented previously in Daura et al.[1] The clustering of the structures from the simulation at 360 K in Daura et al.[−1] was based on a least-squares translational and rotational fitting of all trajectory structures to the NMR model structure, as opposed to fitting of individual pairs from the trajectory. This was sufficient for the purpose intended, but resulted in a considerably larger number of clusters (310 instead of 219 at 360 K), especially of clusters consisting of a single structure (132 instead of 74).

Convergence of Number of Clusters

An indication of how much of the conformational space accessible to the peptide at each of the temperatures has been sampled is given in Figure 8, which shows a plot of the number of clusters found as a function of time in the four simulations. At 298 K there are insufficient statistics for a meaningful analysis. At 340 K and 350 K the number of clusters seems to converge. It is expected, however, that the number of clusters at 350 K would be higher than at 340 K, as the system can sample structures which are higher in energy. This is not the case in Figure 8 and so the apparent convergence, at least of the simulation at 350 K, might not be significant. Longer simulation times would be required to conclude that all significant conformations had been sampled. At 360 K the number of clusters is still clearly increasing at 50 ns. Exclusion of clusters with only one member structure (dashed lines in Fig. 8) does not alter these trends.

Comparison of Clusters at Different Temperatures

Table II shows a comparison of clusters from the different simulations. Only clusters listed in Table I were included in the comparison. Pairs of clusters with a backbone atom-positional RMSD < 0.08 nm (residues 2–6) between the respective central structures are listed in Table II together with their RMSD values. Clearly, the central structure of cluster 1 is very similar at each of the four temperatures (backbone RMSD ≤ 0.04 nm, residues 2–6). It is also clear that in many other cases there is a direct one-to-one correspondence between the clusters at different temperatures. For RMSD ≤ 0.06 nm there is a unique correspondence between the clusters. The first cases of multiple overlap appear at an RMSD of 0.07 nm, for example, cluster 2 at 350 K with clusters 3 and 9 at 340 K.

Free Energy of Folding

The free energy of folding can be calculated from the populations of folded and unfolded structures in the simulations using Equation (1). At 340 K the free energy of folding, that is, from clusters 2 to 158 to cluster 1, is $\Delta G_{\text{folding}}(340\text{ K}) = 0$ kJ mol^{-1}. This implies that the melting temperature of the peptide in the force field is around 340 K. At 350 K the free energy of folding, from clusters 2 to 137 to cluster 1, is $\Delta G_{\text{folding}}(350\text{ K}) = 1.3$ kJ mol^{-1}. At 360 K the free energy of

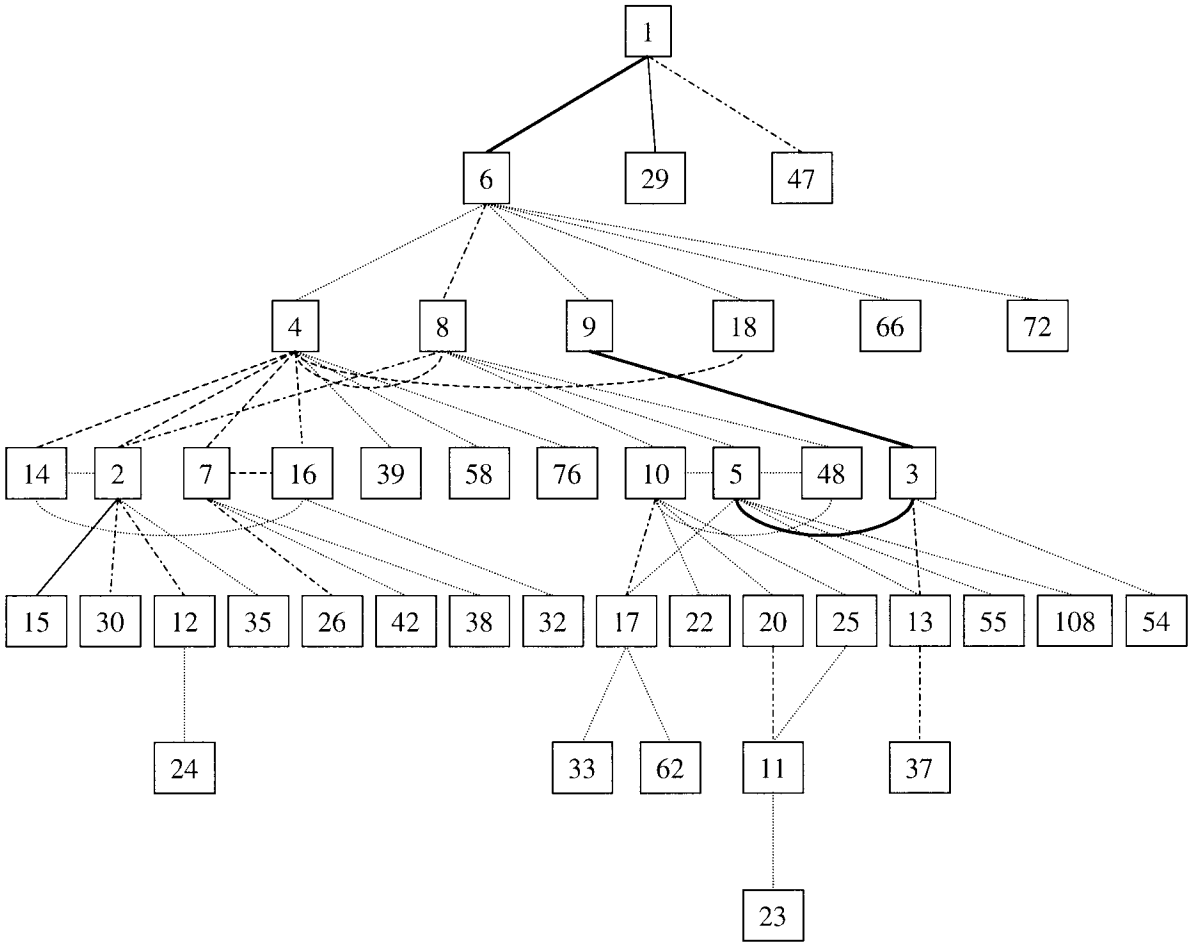

Fig. 11. Paths between clusters of structures at 340 K. See Figure 10 for details.

folding, from clusters 2 to 219 to cluster 1, is $\Delta G_{folding}(360\ K) = 3.3\ kJ\ mol^{-1}$. At 298 K the free energy of folding, from clusters 2 to 9 to cluster 1, is $\Delta G_{folding}(298\ K) = -8.7\ kJ\ mol^{-1}$. At this temperature the calculation of the free energy of folding obviously suffers from the particularly poor statistics of sampling folding–unfolding transitions. The free energy of folding is plotted as a function of temperature in Figure 9, and it is clear that $\Delta G_{folding}$ is approximately linear over the range of temperatures investigated.

Note that Daura et al.[1] calculated the free energy of folding using for the folded conformation a cluster centred at the NMR model structure. Here the cluster number 1 from the simulation is taken for the folded conformation. Unfortunately, there is no unique means to cluster a trajectory. The two approaches are equally reasonable and give differences in the free energies of folding of less than $0.5\ kJ\ mol^{-1}$. In the present study, however, we have tried where possible to avoid biasing the analysis by the incorporation of knowledge of the experimentally derived conformation.

Lifetimes

The average lifetime of clusters with 50 or more member structures is listed in Table III. Note that the lifetimes are dependent on the resolution at which they are calculated—in this case, 0.01 ns. The average lifetime of the folded conformation (cluster 1) decreases as the temperature increases. The average lifetimes of the different unfolded conformations are, at any of the four temperatures, below 0.1 ns, clearly shorter than the average lifetime of the folded conformation. They do not necessarily follow the cluster rank order, that is, unfolded conformations that are more populated than others do not necessarily have also longer average lifetimes.

Low Free Energy Pathways

It is possible to estimate the relative free energy for each pair of clusters from their relative populations in the simulations using Equation (1). From this a free energy landscape for the peptide at each of the temperatures could, in principle, be generated. As an example, the free energy difference between cluster 1 and subsequent clusters is given in Table I for each of the simulations.

Low free energy pathways (pathways with high probability) between clusters are represented in Figures 10 to 13 for each of the four temperatures. These are not necessarily complete pathways actually sampled during the simulation. The connections between clusters in Figures 10 to 13 represent paths that have been sampled at least twice in

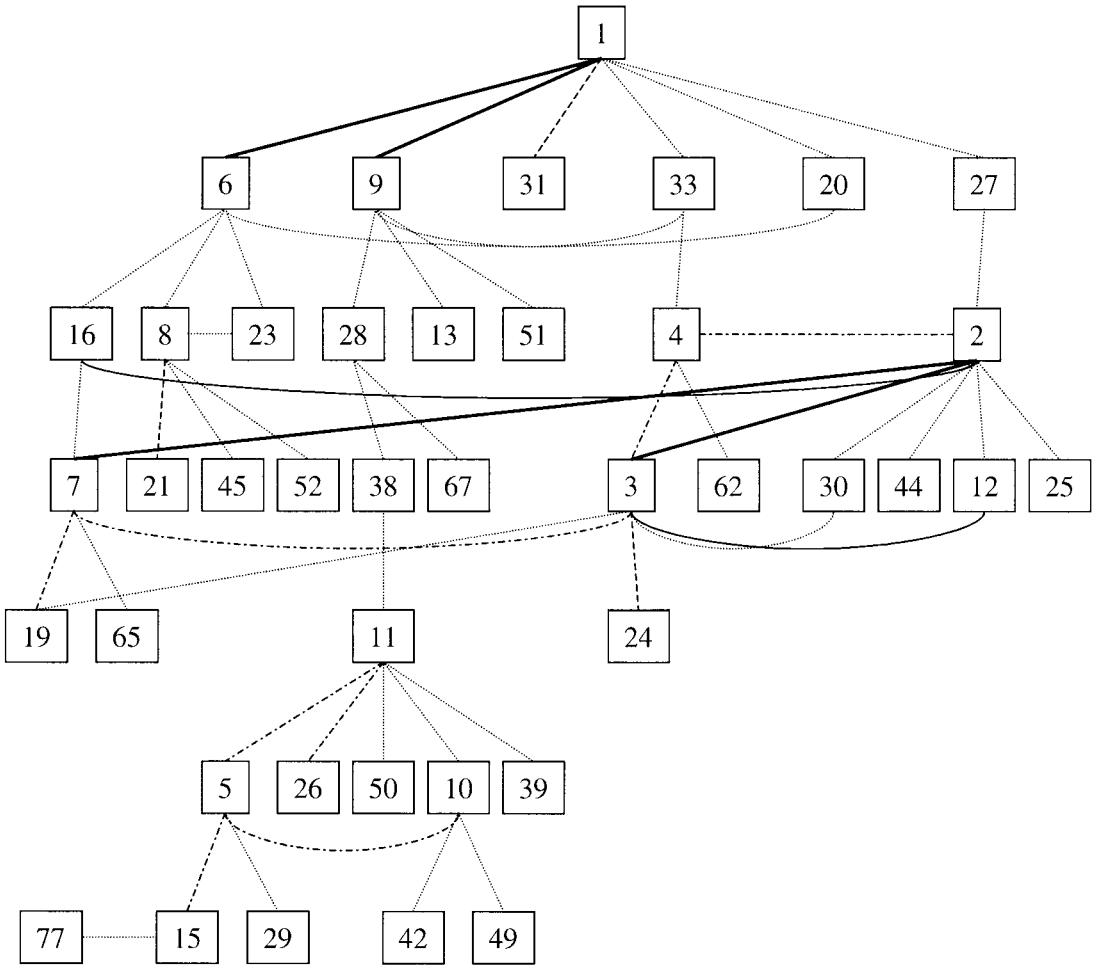

Fig. 12. Paths between clusters of structures at 350 K. See Figure 10 for details.

each direction in the respective simulations. The number of transitions in each direction is in general very similar, as is expected for a system in equilibrium. There are clearly multiple folding pathways from any given cluster representing unfolded structures, or in general, for connecting any two given clusters. Folding pathways are not necessarily downhill in free energy, apart from, obviously, the last step. Examples of this are given in Figure 14. Nevertheless, the number of intermediates of folding and unfolding is found to be small and dependent on temperature. An intermediate of folding and unfolding is here defined as a cluster which appears immediately preceding or immediately following the cluster representing the folded conformation (cluster 1). At 298 K three significant intermediates were found (Fig. 10), clusters 2, 7, and 5, with transition rates (averaged over the two directions) of 0.47 ns^{-1}, 0.09 ns^{-1}, and 0.08 ns^{-1}, respectively. However, clusters 7 and 5 are actually dead ends. At 340 K, with substantially more statistics of folding and unfolding, also only three significant intermediates were found (Fig. 11), clusters 6, 29, and 47, with transition rates of 0.56 ns^{-1}, 0.36 ns^{-1}, and 0.11 ns^{-1}, respectively. Again two of these, clusters 29 and 47 are unproductive dead ends. Cluster 6 is the only productive intermediate of folding and unfolding found at this temperature. At 350 K six significant intermediates were found (Fig. 12), clusters 6, 9, 31, 33, 20, and 27, with transition rates of 0.86 ns^{-1}, 0.43 ns^{-1}, 0.22 ns^{-1}, 0.10 ns^{-1}, 0.06 ns^{-1}, and 0.04 ns^{-1}, respectively. Only cluster 31 represents a dead end. At 360 K seven significant intermediates were found (Fig. 13), clusters 30, 8, 2, 57, 3, 60, and 20, with transition rates of 0.31 ns^{-1}, 0.27 ns^{-1}, 0.24 ns^{-1}, 0.11 ns^{-1}, 0.08 ns^{-1}, 0.05 ns^{-1}, and 0.04 ns^{-1}, respectively, cluster 57 being a dead end. Figures 10 to 13 suggest that as the temperature increases so does the number of intermediates of folding and unfolding and the number of interconnections between them. It is also apparent from these figures (and Table I) that the lowest free energy (most populated, lowest cluster number) unfolded conformations are not necessarily directly connected to the folded conformation. The central structures of clusters 2 at 298 K, 6 at 340 K, 6 at 350 K, and 8 at 360 K, the most common intermediates of folding and unfolding in the respective simulations (second most common at 360 K), are all very similar and represent a common fold (Table II and Figs. 3–6). Moreover, cluster 29 at 340 K is very similar to cluster 30 at 360 K (Table II), and similar also to

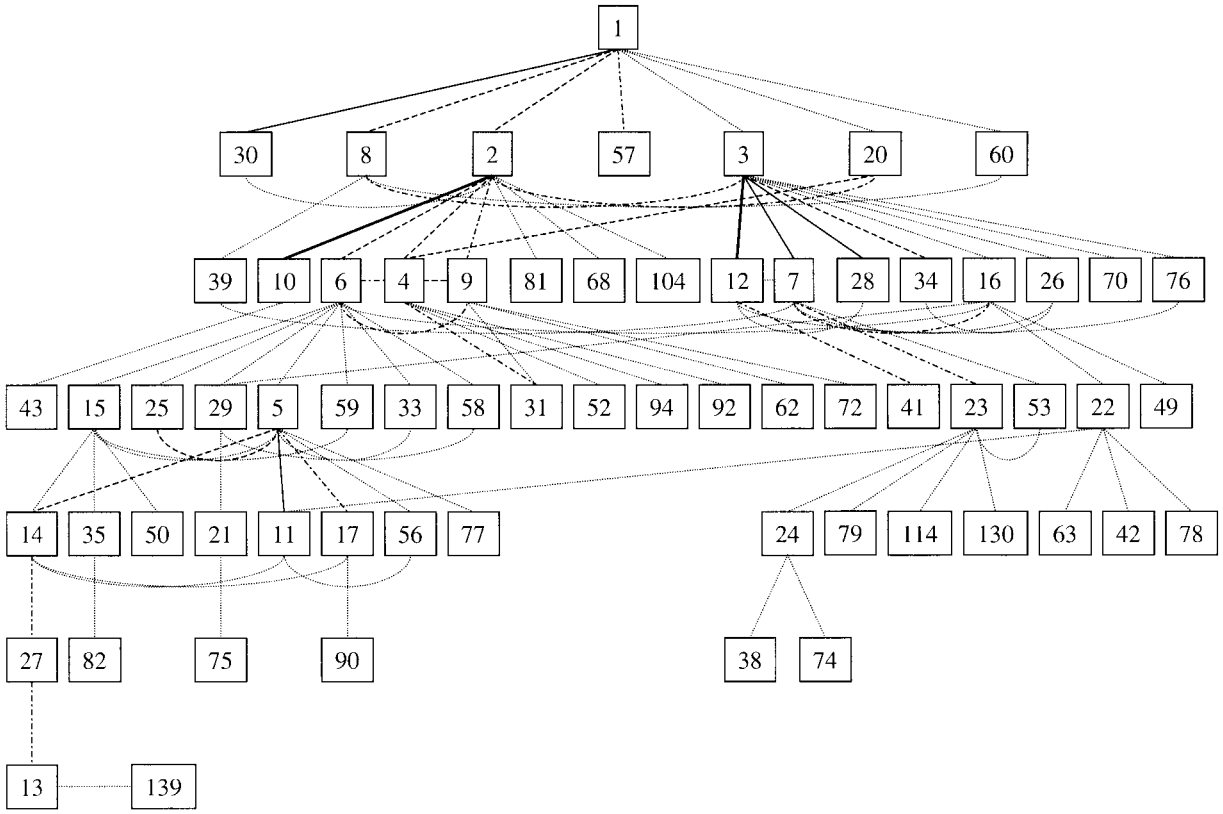

Fig. 13. Paths between clusters of structures at 360 K. See Figure 10 for details.

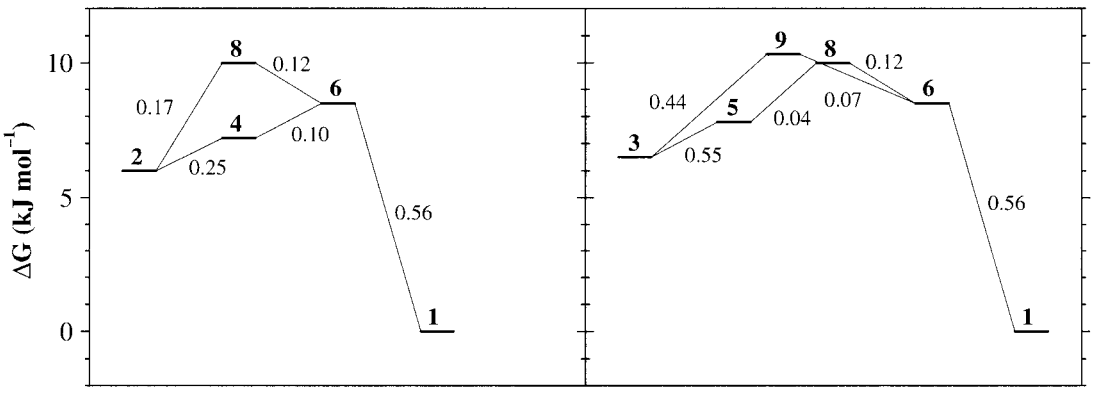

Fig. 14. Example folding pathways from cluster 2 (left-hand plot) and cluster 3 (right-hand plot) at 340 K. The vertical axis indicates the free energy difference with respect to cluster 1. The transition rate (in ns^{-1}) between consecutive clusters is also indicated. Only the two shortest folding pathways (i.e., those with the minimum number of intermediate clusters) are shown. At 340 K (Fig. 11) there are multiple alternative pathways connecting clusters 2 and 1 and clusters 3 and 1 that involve 4 or more intermediates.

cluster 9 at 350 K (RMSD of 0.09 nm between central structures). All these folding intermediates have residual helicity.

DISCUSSION

The results shown in the previous section are based on two important assumptions. First, that there is sufficient statistics in each simulation for the population ratios between the clusters to have converged. Second, that the results are independent of the clustering algorithm. Neither assumption is strictly valid. The ratios between the populations of different conformations would most likely change if the simulations were extended, and the clustering algorithm together with the clustering criterion does affect the results. The quantitative aspects of the work, populations, ratios, and free energy differences between specific clusters must be interpreted cautiously. Nevertheless, the qualitative picture which emerges from the simulations is clear.

In the simulations the peptide does not sample randomly conformational space at any of the four tempera-

tures studied. Instead, there is a rapidly established equilibrium between a relatively small number of conformations, including the conformation that is predominant at room temperature. The relative population and average lifetime of the 3_1-helical folded conformation decreases as the temperature of the system increases, but the folded conformation remains the dominant and most stable one even at 360 K. The melting temperature of the peptide in the force field is around 340 K.[1] Clusters centred around very similar (unfolded) structures were observed in the simulations at different temperatures, although with different weights.

There are multiple folding pathways from any given conformation. These pathways are not necessarily downhill in free energy. Nevertheless, the number of folding and unfolding intermediates (conformations which appear immediately before or immediately after the folded conformation) appears to be small and dependent on temperature. Furthermore, the most common folding and unfolding intermediates are the same at the four temperatures studied. All have residual helicity, which suggests that the central turn may be the point of nucleation of the helix.

There has been a growing tendency in recent years to postulate (free) energy landscapes as a means to explain the mechanism of peptide or protein folding.[6–12] Many of these free energy landscapes have been derived from lattice simulations,[6–9] which allow for the complete enumeration of conformational space if the model is simple enough. Such studies have the advantage of being statistically rigorous, but to what extent ideas and results based on these lattice models can be related to peptide and protein systems is uncertain. Some conclusions drawn from lattice simulations, such as a strong relationship between fast folding and an overall funnel shape of the free energy surface or a large free energy gap between the native and the lowest nonnative structure, have been questioned.[13] There have also been attempts to derive free energy landscapes for protein systems from MD simulations at atomic resolution.[10–12] These are not based on equilibrium simulations of folding and unfolding, as spontaneous folding of proteins is still beyond currently accessible time scales. Instead, a number of unfolded structures is generated and umbrella sampling techniques are used to obtain a free energy profile for folding with respect to a given coordinate such as the radius of gyration or the number of native contacts. This approach is also not without question. For one reason, unfolded structures are commonly generated from high temperature (e.g., 400 K or higher) simulations and used to represent the unfolded structures populated at room temperature.[10–12,14] Highly populated unfolded structures at high temperature are not necessarily representative of unfolded structures at temperatures at which folding is studied (whether with or without denaturants), and it is uncertain if such structures can relax to representative ones on the time scale simulated. A more serious problem, however, involves the projection of the results onto an arbitrary reaction coordinate. As is apparent from the literature there is no unique coordinate by which to describe folding.[6–12] Any form of projection leads by necessity to the superposition of states which are potentially unrelated on a folding pathway. The equilibrium simulations presented here do allow one to directly analyse the mechanism of folding and unfolding. The information is contained in the populations of different states, their average lifetimes and their respective transition rates. Even if there were enough statistics to be certain about the relative populations of every conformation, we doubt if this information could be meaningfully represented as a simple landscape in two or even three dimensions, as is currently so often attempted.

CONCLUSIONS

To date, two main approaches have been used to study peptide and protein folding by simulation. In the first, a physically meaningful model is sacrificed for the sake of rigor in the statistical mechanical analysis. In the second, statistical rigor is sacrificed for the sake of a physically reasonable model. This work shows that there is a third way. We have demonstrated that it is possible to study the process of spontaneous folding and unfolding of a peptide under equilibrium conditions using a physically reasonable, atom-based, model which explicitly includes solvent degrees of freedom. The current state of empirical force fields and accessible computational power is such that, at least for systems such as presented here, it is possible to simulate an equilibrium distribution of folded and unfolded species in solution and use statistical mechanics to directly analyse the thermodynamics of peptide folding and unfolding.

REFERENCES

1. Daura X, Jaun B, Seebach D, van Gunsteren WF, Mark AE. Reversible peptide folding in solution by molecular dynamics simulation. J Mol Biol 1998;280:925–932.
2. Seebach D, Ciceri PE, Overhand M, Jaun B, Rigo D, Oberer L, Hommel U, Amstutz R, Widmer H. Probing the helical secondary structure of short-chain β-peptides. Helv Chim Acta 1996;79:2043–2066.
3. van Gunsteren WF, Billeter SR, Eising AA, Hünenberger PH, Krüger P, Mark AE, Scott WRP, Tironi IG. Biomolecular simulation: the GROMOS96 manual and user guide. Vdf Hochschulverlag AG an der ETH Zürich, Zürich; 1996. p 1–1024.
4. Daura X, van Gunsteren WF, Rigo D, Jaun B, Seebach D. Studying the stability of a helical β-heptapeptide by molecular dynamics simulations. Chem Eur J 1997;3:1410–1417.
5. Karpen ME, Tobias DJ, Brooks CL III. Statistical clustering techniques for the analysis of long molecular dynamics trajectories: analysis of 2.2-ns trajectories of YPGDV. Biochemistry 1993;32:412–420.
6. Shakhnovich EI. Theoretical studies of protein-folding thermodynamics and kinetics. Curr Opin Struct Biol 1997;7:29–40.
7. Onuchic JN, Luthey-Schulten Z, Wolynes PG. Theory of protein folding: the energy landscape perspective. Annu Rev Phys Chem 1997;48:545–600.
8. Chan HS, Dill KA. Protein folding in the landscape perspective: Chevron plots and non-Arrhenius kinetics. Proteins 1998;30:2–33.
9. Dobson CM, Sali A, Karplus M. Protein folding: a perspective from theory and experiment. Angew Chem Int Ed 1998;37:868—893.
10. Guo Z, Brooks CL III, Boczko EM. Exploring the folding free energy surface of a three-helix bundle protein. Proc Natl Acad Sci USA 1997;94:10161–10166.
11. Sheinerman FB, Brooks CL III. Molecular picture of folding of a small α/β protein. Proc Natl Acad Sci USA 1998;95:1562–1567.
12. Sheinerman FB, Brooks CL III. Calculations on folding of segment B1 of streptococcal protein G. J Mol Biol 1998;278:439–456.
13. Crippen GM, Ohkubo YZ. Statistical mechanics of protein folding by exhaustive enumeration. Proteins 1998;32:425–437.
14. Kazmirski SL, Daggett V. Simulations of the structural and dynamical properties of denatured proteins: the "molten coil" state of bovine pancreatic trypsin inhibitor. J Mol Biol 1998;277:487–506.