Energy Landscape of a Small Peptide Revealed by Dihedral Angle Principal Component Analysis

Yuguang Mu,* Phuong H. Nguyen, and Gerhard Stock*
Institute of Physical and Theoretical Chemistry, J. W. Goethe University, Frankfurt, Germany

ABSTRACT **A 100 ns molecular dynamics simulation of penta-alanine in explicit water is performed to study the reversible folding and unfolding of the peptide. Employing a standard principal component analysis (PCA) using Cartesian coordinates, the resulting free-energy landscape is found to have a single minimum, thus suggesting a simple, relatively smooth free-energy landscape. Introducing a novel PCA based on a transformation of the peptide dihedral angles, it is found, however, that there are numerous free energy minima of comparable energy ($\lesssim$ 1 kcal/mol), which correspond to well-defined structures with characteristic hydrogen-bonding patterns. That is, the true free-energy landscape is actually quite rugged and its smooth appearance in the Cartesian PCA represents an artifact of the mixing of internal and overall motion. Well-separated minima corresponding to specific conformational structures are also found in the unfolded part of the free energy landscape, revealing that the unfolded state of penta-alanine is structured rather than random. Performing a connectivity analysis, it is shown that neighboring states are connected by low barriers of similar height and that each state typically makes transitions to three or four neighbor states. Several principal pathways for helix nucleation are identified and discussed in some detail. Proteins 2005;58:45–52.**

Key words: folding; helix nucleation; pathways; internal motion

INTRODUCTION

The concept of describing biomolecular processes in terms of the molecule's energy landscape has promoted much of the recent progress in understanding the process of protein folding.[1–4] Classical molecular dynamics (MD) simulations have proven to be a valuable tool to account for the energetics and kinetics of folding and unfolding peptides and small proteins.[5–8] However, since it is neither possible nor desirable to represent the free energy as a function of all 3N-6 coordinates of a biomolecule, one needs to invoke a strategy to identify the most important conformational degrees of freedom of a simulation. To this end, the principal component analysis (PCA) method, also called quasiharmonic analysis or essential dynamics method, has been found useful.[9–13] The approach is based on the covariance matrix

$$\sigma_{ij} = \langle (x_i - \langle x_i \rangle)\,(x_j - \langle x_j \rangle) \rangle, \qquad (1)$$

where $x_1, \ldots x_{3N}$ are the mass-weighted Cartesian coordinates of the N-particle system and $\langle \ldots \rangle$ denotes the average over all sampled conformations. Hence, the covariance matrix provides information on the correlated fluctuations of pairs of atoms. The eigenvectors and eigenvalues of σ yield the modes of collective motion and their amplitudes. It has been shown that a large part of the system's fluctuations can be described in terms of only a few PCA eigenvectors.[9–15]

To study the conformational dynamics of a molecule by a PCA, one first needs to discriminate the (interesting) internal motion from the (trivial) overall motion, that is, translation and overall rotation. The translation can readily be separated by requiring that the molecule's center of mass is fixed. The elimination of overall rotation, however, is only straightforward for a rigid body or in the case of small fluctuations around a reference structure. In this case, the separation can be achieved by superimposing the structures of trajectory with the reference structure such that the root-mean-square deviations (RMSD) of atomic fluctuations is minimized.[16] While this procedure correctly removes the overall rotation in the limit of small fluctuations,[17] it breaks down in the case of large amplitude motion, where it is not possible to unambiguously define a single reference structure.[18] In particular, this is true for a folding peptide, in which case the outcome of a PCA has been shown to depend significantly on the reference structure chosen.[19]

As a remedy, a number of methods have been proposed which all have their virtues and problems and often yield complementary information of the conformational dynamics. For example, Schieborr and Rüterjans suggested an approach that minimizes the internal motion, and therefore yields a lower limit of the correlated motion.[20] Prompers and Brüschweiler proposed an isotropically distributed ensemble analysis,[19] which has proven valuable in describing NMR experiments on flexible systems.[21] However, the orientational average makes the illustration of the corre-

*Correspondence to: Yuguang Mu whose present address is the School of Biological Science, Nanyang Technological University, Nanyang Drive 60, Singapore 637551. E-mail: ygmu@ntu.edu.sg or to Gerhard Stock, Institute of Physical and Theoretical Chemistry, J. W. Goethe University, Marie-Curie, Str., 11, D-60439 Frankfurt, Germany. E-mail: stock@theochem.uni-frankfurt.de

Received 17 May 2004; Accepted 19 July 2004

Published online 1 November 2004 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20310

sponding PCA vectors difficult in Cartesian space. Furthermore, the overall motion is safely eliminated if internal coordinates are used in the PCA such as dihedral angles[22,23] and distances.[24] However, the periodicity of dihedral angles may lead to artefacts and covariances in distance space describe correlated motions of three and four atoms, that are therefore different from the correlation between two atoms described by equation 1.

In this work, we propose a method, referred to as dPCA, that is based on the dihedral angles (ϕ_i,ψ_i) of the peptide backbone. Clearly, the consideration of dihedral angles in a flexible molecule is appealing because other internal coordinates such as bond lengths, bond angles usually do not undergo changes of large amplitudes. However, attention has to be paid to the fact that—unlike Cartesian coordinates—dihedral angles are periodic. As a consequence, the arithmetic mean of dihedral angles can not be easily calculated as in Cartesian coordinates, e.g., the mean of two dihedral angles [(3π)/4] and −[(3π)/4] should be π, while simple arithmetic average gives 0. To avoid this and related problems, we perform a transformation from the space of dihedral angles to a *metric* coordinate space (i.e., a vector space with a well-defined distance between any two points) built up by the trigonometric functions of the dihedral angles. The resulting representation of the dihedral angles is unique and therefore allows us to readily calculate the mean and the covariance matrix.

To investigate the potential of the dPCA method, we perform a detailed MD study on penta-alanine Ac-Ala_5-NHMe (A_5) in explicit water solvent. A_5 has been adopted by several authors as a model system to study the helix nucleation process in proteins.[25,26] Due to the short length of the peptide, the MD trajectory of A_5 is found to perform numerous folding and unfolding events within the simulation time of 100 ns. Employing a standard PCA using Cartesian coordinates, the free energy surface is found to have a single pronounced minimum. In agreement with previous works,[25,26] this finding suggests a simple, relatively smooth free-energy landscape. Employing the dPCA using dihedral angles, however, we find that there are actually numerous free-energy minima, which correspond to well-defined structures with characteristic hydrogen-bonding patterns. That is, the true free energy landscape is actually quite rugged and its smooth appearance in the Cartesian PCA represents an artifact of the mixing of internal and overall motion. Employing the free-energy landscape provided by the dPCA, we construct a map of connectivity between the various free energy minima, which allows us to give a detailed description of the folding and unfolding dynamics of the peptide.

METHODS

MD Simulations

The GROMACS program suite[27,28] and the GROMOS force field 45A3[29] were used in the simulations of A_5. The peptide was solvated in a rectangular box of SPC water,[30] keeping a minimum distance of 10 Å between the solute and each face of the box. The final system contained 2385 atoms within a box dimension of 32 Å. The equation of motion was integrated by using a leapfrog algorithm with a time step of 2 fs. Covalent bond lengths were constrained by the procedure SHAKE[31] with a relative geometric tolerance of 0.0001. A twin-range cutoff of 0.8/1.4 nm was used for the nonbonded interactions, and a reaction-field correction with permittivity $\epsilon_{RF} = 54$ was employed.[32] The nonbonded interaction pair-list was updated every 10 fs. The solute and solvent were separately weakly coupled to external temperature baths at 300 K.[33] The temperature-coupling constant was 0.1 ps (0.01 during the first 10 ps). The total system was also weakly coupled to an external pressure bath at 1 atm using a coupling constant of 0.5 ps (0.05 during the first 10 ps).

Starting with a fully extended configuration of penta-alanine (state 16 in Fig. 3), the equilibration protocol consisted of 100 steps of steepest-descent minimization applied to the solvent molecules with fixed solute, followed by 5 ps of MD simulation at 300 K of the solvent with fixed A_5, and another 5 ps simulation without position constraining of A_5. The simulation was then continued for 100 ns, whereby the coordinates were saved every 0.1 ps for analysis. As a test of the convergence of the sampling, we compared the free-energy surfaces obtained for 50 and 100 ns simulation times, which were found to be quite similar.

Principal Component Analyses

As explained in the Introduction, we first performed a standard PCA using Cartesian coordinates, cf. equation 1. In an attempt to remove the overall rotation, we chose the completely folded α-helix (state 1 in Fig. 3) as reference structure for the rotational fit. Choosing other structures as reference state (e. g., the initial extended state), we obtained qualitatively similar free energy surfaces (data not shown), although the details of the plots may look quite different.

By diagonalizing the covariance matrix σ we obtain the eigenvectors v_n and the eigenvalues λ_n, which are organized in descending order, i.e., λ_1 represents the largest eigenvalue. In the Cartesian PCA, the motions along the first two eigenvectors v_1, v_2 of σ contain 51% of the fluctuations of the A_5 simulation. Restricting ourselves to these first two eigenvectors, the free energy surface along these vectors is given by

$$\Delta G(v_1,v_2) = -k_B T[\ln P(v_1,v_2) - \ln P_{max}]. \qquad (2)$$

Here $P(v_1,v_2)$ is the probability distribution obtained from a histogram of the MD data and P_{max} denotes the maximum of the distribution, which is subtracted to ensure that $\Delta G = 0$ for the lowest free energy minimum. The representation of the free energy surface as a function of higher eigenvectors was found to not yield significantly more information.

In the dPCA method, we consider the dihedral angles (ϕ_i,ψ_i) of the peptide backbone, where $i = 1, \ldots, N_p$ with N_p being the number of peptide groups. Instead of using the dihedral angles directly in the PCA, we introduce the variables

$$x_{4i-3} = \cos(\phi_i), \quad x_{4i-2} = \sin(\phi_i),$$
$$x_{4i-1} = \cos(\psi_i), \quad x_{4i} = \sin(\psi_i). \qquad (3)$$

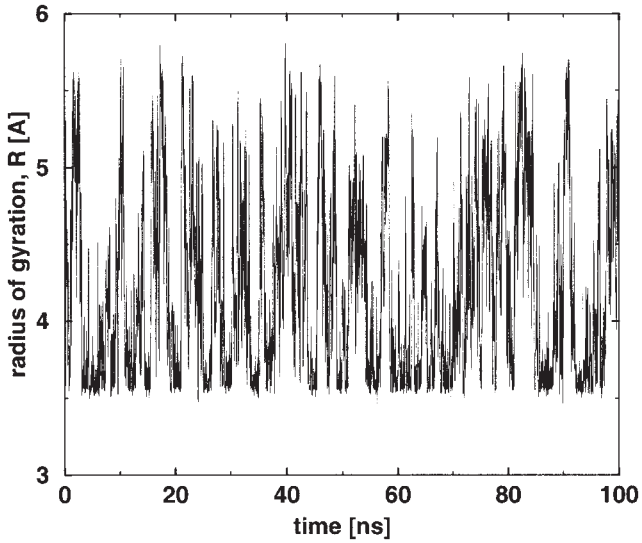

Fig. 1. Time evolution of the radius of gyration $\Re$ obtained from the MD simulation of penta-alanine, showing frequent transitions between extended conformations ($\Re \gtrsim 5$ Å) and α-helical conformations ($\Re \approx 3.6$ Å).

Note that the dimension of the coordinate vectors in the dPCA is 4 N_p, twice as large as the number of backbone dihedral angles. It is easy to show that the thus defined space of dPCA vectors is metric, that is, we can define a unique distance function $\Delta = \Sigma\,[x_i(1) - x_i(2)]^2$ between any two vectors x(1) and x(2). The resulting representation of the dihedral angles is nonperiodic but unique and therefore allows us to unambiguously calculate mean values and other quantities. Interestingly, this is not the case if one considers only cosine terms (or only sine terms) in the PCA. In particular, the thus obtained free energy surfaces do not resolve the rugged structure of the energy landscape found in the full dPCA of equation 3.

It is noted that the total fluctuations covered by the first dPCA vectors converge somewhat slower than in the case of the Cartesian PCA. In the simulation of A_5, for example, the first two and three eigenvectors contain 37% and 48% of the overall fluctuations of the peptide, respectively. Roughly speaking, we therefore need three eigenvectors (instead of two as in the case of the Cartesian PCA) to represent the main features of the energy landscape of A_5.

RESULTS AND DISCUSSION

Folding Dynamics

In the 100 ns MD simulation, A_5 does not maintain a stable α helical conformation but exhibits frequent folding and unfolding events. This is clearly manifested in the time trace of the radius of gyration which is shown in Figure 1. The radius of gyration is roughly 3.6 Å in the α-helical conformations and above 5.0 Å in the extended conformers. The decay time constant of the autocorrelation function of the radius of gyration is found to be 0.5 ns. This reflects the fast folding rate of A_5, which is consistent with previous MD simulations.[25,26] The time scale of 1 ns for the overall folding process should not be confused with the time it takes to perform an individual conformational transition during the folding. As shown below in Figure 4 the typical time scale of such a transition involving the making or breaking of structural hydrogen bonds is about 100 ps.

Free Energy Landscape

Let us first discuss the results of the standard PCA using Cartesian coordinates. Figure 2(a) shows the free energy surface as a function of the first two eigenvectors for this case. The free energy landscape exhibits a single prominent minimum which is found to correspond to the α-helical structure. The results are in direct agreement with the simulations of Hummer et al.[25] using the AMBER94 force field.[34] Furthermore, the data qualitatively agree with the simulations of Margulis et al.,[26] who in addition to the α-helical structure identified a weaker minimum corresponding to the extended state. It is noted, however, that the details of their simulation were somewhat different to ours, i.e., they performed relatively short constant-energy (instead of constant-temperature) simulations employing the OPLS-AA force field[35] and used only the backbone heavy atoms in the PCA analysis.

Generally speaking, one should be aware of the fact that highly flexible small peptides occupying several conformational states of similar energy represent a major challenge for a molecular mechanics force field description. Recent comparison studies have shown that commonly employed force fields may predict qualitatively different populations of energetically close-lying conformers.[36–40] In the present study, the GROMOS force field 45A3[29] was used, which was found to be in good agreement with experimental infrared and NMR spectra on trialanine.[37] This finding has been reconfirmed in a recent joint NMR/MD study[41] on uncapped penta-alanine Ala_5 (that -unlike capped A_5- is readily soluable in water), which to some extent may also be considered as an experimental validation of the computational results presented in this work.

While the standard PCA using Cartesian coordinates suggests a simple and relatively smooth free energy landscape, the results of the dPCA using dihedral angles look quite different. For this case, Figure 2 shows the free energy as a function of (b) the first two eigenvectors v_1 and v_2 and (c) the first and the third eigenvectors v_1 and v_3, respectively. Obviously, the dPCA can clearly resolve numerous minima of the free-energy surface. By using a distance criterion in the space of the first three dPCA eigenvectors, 16 conformational states corresponding to the free-energy minima are identified, which cover nearly 90% of the configurations sampled in the simulation. Representative structures of these states are shown in Figure 3. The structures cover a broad range of the conformational space of the peptide, from the all-extended conformer, structure 16, to the all-α-helical conformer, structure 1. The corresponding average dihedral angles, relative populations, free energies, and average lifetimes of the conformational states are listed in Table I. The data reveal that all states are relevant in the sense that they have similar average life times of ≈ 20–30 ps and show

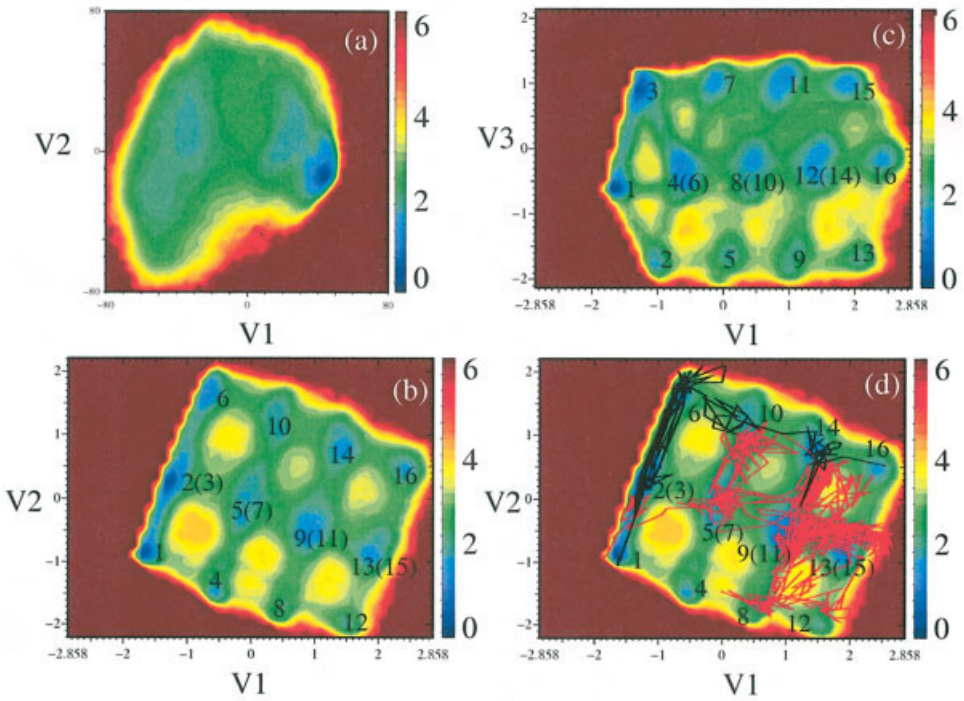

Fig. 2. Free energy landscape (in kcal/mol) of penta-alanine as obtained from various PCAs of the 100 ns MD simulation. Compared are the results of (**a**) the standard PCA using Cartesian coordinates and (**b**–**d**) the dPCA using dihedral angles. In (**a**), (**b**), and (**d**) the energy is plotted in the plane of the first two PCA eigenvectors v_1 and v_2, in (**c**) the first and third PCA eigenvectors v_1 and v_3 are used. The red and black lines in (**d**) reflect two representative folding pathways of the peptide.

comparable population probabilities corresponding to a free energy differences of $\lesssim 1$ kcal/mol.

To characterize the conformational states of A_5, we may consider the hydrogen (H) bonding of the system, in particular the formation of (i,i+4) α-helical bonds, of (i,i+3) 3_{10}-helical bonds, and of (i,i+2) turns. Employing the H-bonding criteria derived by Kabsch and Sander,[42] we have calculated the relative occurrences of the various H-bonding types and listed them in Table I. Clearly, the helical appearance of structures 1, 3, 6, 4, and 2 seen in Figure 3 is caused by three, two, and one (i,i+4) H-bonds, respectively. On the other hand, the formation of (i,i+3) and (i,i+2) structures is less prominent and typically only one of these H-bonds is observed at a time.

Having identified its minima, the next step to characterize the free energy landscape is to study the possible transitions between these minima. To this end, we have constructed a map of connectivity shown in Figure 4, in which each conformational state is represented by a circle. If two circles are connected by a line, direct conformation transitions between the two states exists (in both directions) and the number of these transitions during the 100 ns trajectory are written next to the line. Hereby only transition are considered, which were performed more than 50 times within the simulation time of 100 ns. The map of connectivity in Figure 4 reveals that most conformational states are connected to either three or four next neighbors. The exceptions are the two less prominent structures 12 and 13, which make transitions to only two next neighbors. To relate the connectivity to the free energy surface, it is instructive to calculate the (dimensionless) distances between the various conformational states in the dPCA space (data not shown). Interestingly, we obtain an almost perfect correlation between the state-to-state distances and the probabilities of the corresponding conformational transitions, that is, virtually all connected minima are less than 2.0 apart. It is noted that a distance of 2.0 in dPCA space corresponds, for example, to a 180° difference of a single dihedral angle. The well-resolved structures and distances of the dPCA free energy surfaces therefore reflect the sensitivity of the present method in monitoring large amplitude changes in dihedral angle space.

To summarize, it has been found that the free energy surface of A_5 obtained from the Cartesian PCA exhibits a smooth appearance and a simple single-minimum funnel-like shape. Considering the results obtained from the dPCA, we have shown, however, that these features are an artifact of the mixing of internal and overall motion. In fact, the dPCA shows that the true free energy landscape of A_5 is quite rugged, that is, it contains numerous minima of comparable probabilities and life times.

Folding Pathways

Let us now employ the connectivity map to identify possible pathways of folding and unfolding of A_5. Since the *complete* pathway from the all-extended state 16 to the all-α-helical state 1 may be quite complicated (see below),

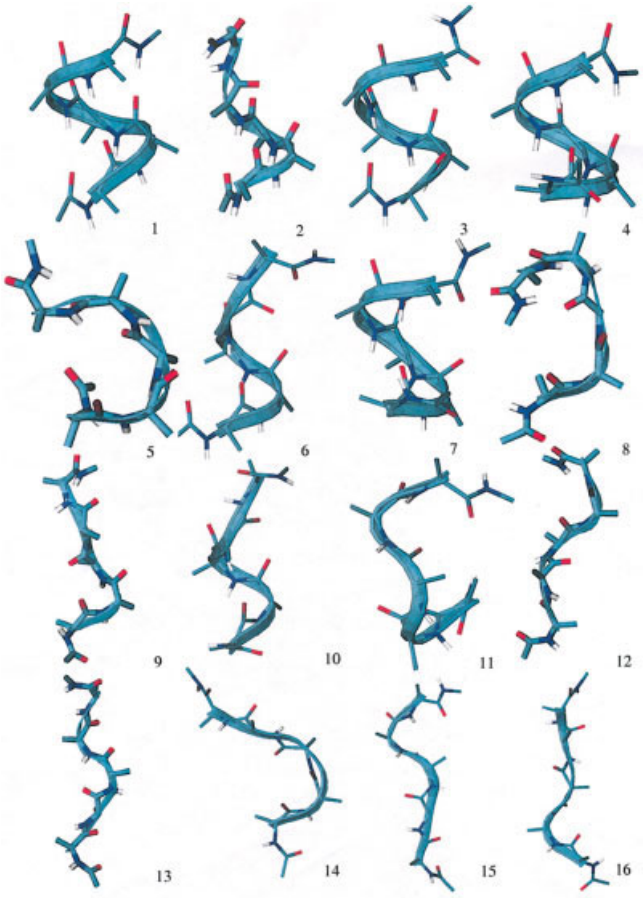

Fig. 3. Representative structures of the conformational states corresponding to the most prominent minima of the free energy landscape. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

it is helpful to first focus on the initial steps of folding and unfolding. Beginning with the unfolding of state 1 with three (i,i+4) α-helical H-bonds, one may observe transitions to three neighbor states: Preferably to state 3, showing two α-helical H-bonds at the N-terminus side, but also to states 4 and 2, which exhibit two and one α-helical H-bond at the C-terminus side, respectively.

According to Figure 4, the next step of the unfolding process may lead the system to states 6, 7, 8, and 5. The first three states have a considerable amount of conformations with a single α-helical H-bond, which is located at the N-terminal (state 6), the peptide center (state 7), and the C-terminal (state 8), respectively. Adopting helix-coil theory,[43,44] in which the first (i,i+4) H-bond that stabilizes a α-helix is regarded as the "nucleus" of the formation of the complete α helix, states 6, 7, 8 can be considered as the nucleus of the folding process. State 5, on the other hand, does not exhibit an α-helical H-bond but rather adopts a global hairpin structure, whose bending position is centered on Ala2 (see Figure 3). As is shown below, this state also acts as a "transition state" between folded and unfolded structures and may therefore be as well considered as a nucleus of folding (although not in the strict sense of helix-coil theory). The four nucleus states can be considered as a separation of the configurational space of A_5 in a folded part (including the four at least partially folded states 1, 3, 4, and 2) and an unfolded part (including the remaining eight states).

While for the first two steps of unfolding there are no other possibilities than described above (besides back-crossing), subsequently many pathways are possible. For this reason, we continue the discussion with the initial steps of the folding pathways. The process begins with the all-extended state 16, which can be best described as a proline II conformation, i.e., ϕ and ψ are centered around $-75°$ and $145°$, respectively. Starting from state 16, we distinguish four principal pathways to reach the above discussed nucleus states 6, 7, 8, and 5:

path A:	16	→	14	→	10	→ 6
path B:	16	→	15	→	11	→ 7
path C:	16	→	15	→	12	→ 8
path D:	16	→	13	→	9	→ 5

Since the states involved in these pathways are effectively interconnected, moreover, there exist a large number of possibilities of back-crossing and connecting paths.

Path A appears to be the simplest folding route. According to Tab.iI, the transitions from state 16 to state 14 and further to state 10 and state 6 involves mainly the change of the ψ-angles of the first three residues. This reflects the stepwise conformational transition from an extended structure of the first three residues in state 16 to their α-helical structure in state 6 and leads to a α-helical H-bond at the N-terminal. Similarly, pathway C exhibits a stepwise development of a α-helical nucleus at the C-terminal folding. Compared to path A, however, in A_5 the C-terminal folding pathway is found to play a minor role, because the free energy barrier between states 8 and 12 is quite high (1.7 kcal/mol higher than in path A).

Path B appears to be more complicated. Going from the all-extended state 16 to state 15, only the ψ_4-angle changes significantly, thus reflecting an α-helix-like structure of the forth residue. The next step to state 11 then again involves mainly the change of the ψ-angles of the first three residues, in particular of ψ_2. Proceeding to state 7, we observe a change of the angles ψ_2 and ψ_3. This yields an α-helical H-bond between residue 1 and 5, that is, the center of Ac-Ala_5-NHMe exhibits α-helical structure. Pathway D, finally, is special as it never attains a complete α-helical H-bond. Although first (in state 13) the N-terminal dihedral angles ϕ_5,ψ_5 and later (in state 9) also the C-terminal angles ϕ_1,ψ_1 and ϕ_2,ψ_2 get close to the α-helical region, the overall structure along this path leading to state 5 is best described by a global hairpin structure.

To summarize, we have found the following folding pathways for helix nucleation, path A which leads to an α-helical H-bond at the N-terminal, path B which leads to an α-helical H-bond at the center of the peptide, path C which has an α-helical H-bond at the C-terminal, and path D which exhibits a global hairpin structure. Due to the limited length of A_5, it is difficult to estimate the relevance

TABLE I: Characterization of the Most Prominent Conformational States of A_5 in Water†

State	P[%]	τ[ps]	ΔG	ϕ_1,ψ_1	ϕ_2,ψ_2	ϕ_3,ψ_3	ϕ_4,ψ_4	ϕ_5,ψ_5	1_α	2_α	3_α	$1_{3_{10}}$	$2_{3_{10}}$	1_t	2_t
1	14	35	0.05	−57,−48	−64,−43	−63,−43	−65,−39	−72,−34	11	27	60	28	7	2	0
2	3	18	0.98	−58,−44	−72,−36	−68,−36	−88,102	−67,−35	70	2	0	29	10	33	3
3	15	25	0	−58,−47	−65,−40	−66,−39	−69,−39	−90,115	25	67	1	26	6	28	0
4	4	20	0.84	−61,59	−58,−18	−68,−10	−63,−39	−76,−36	29	33	0	30	9	24	1
5	3	17	0.90	−65,7	−70,−4	−80,31	−71,110	−68,−37	5	0	0	27	4	41	11
6	7	23	0.44	−60,−43	−70,−37	−70,−38	−93,113	−75,117	68	1	0	24	8	39	7
7	3	14	0.89	−63,75	−56,−28	−68,−14	−69,−42	−91,103	46	2	0	30	12	40	7
8	1	12	1.58	−71,110	−59,32	−68,55	−60,−43	−85,−35	22	0	0	27	7	36	6
9	2	16	1.13	−72,91	−70,53	−65,54	−62,116	−72,−42	4	0	0	9	1	40	11
10	6	20	0.51	−67,10	−71,−5	−81,23	−78,111	−78,118	5	0	0	29	4	43	14
11	8	18	0.41	−69,79	−73,24	−79,95	−60,−36	−89,113	2	0	0	15	1	40	12
12	1	13	1.52	−74,121	−73,121	−69,119	−56,−46	−87,−36	3	0	0	26	1	37	10
13	1	19	1.41	−74,120	−74,117	−60,115	−67,119	−74,−43	1	0	0	5	0	42	13
14	8	23	0.35	−73,78	−66,45	−67,70	−66,110	−77,117	2	0	0	11	2	42	16
15	6	26	0.58	−72,118	−72,117	−67,118	−50,−36	−90,116	0	0	0	15	3	41	17
16	5	24	0.69	−70,116	−74,117	−64,116	−60,111	−81,117	1	0	0	9	1	40	18

†Shown are the population probability P (in %), the average lifetime τ (in ps), the free energy ΔG (in kcal/mol) relative to the minimum-energy state 3, the average dihedral angles (ϕ_i, ψ_i), as well as the occurrences (in %) of various hydrogen bonding types: n_α denote the occurrence of n α-helical (i,i + 4) bonds, while $n_{3_{10}}$ and n_t refer to 3_{10}-helical (i,i + 3) bonds and to (i,i + 2) turns, respectively.

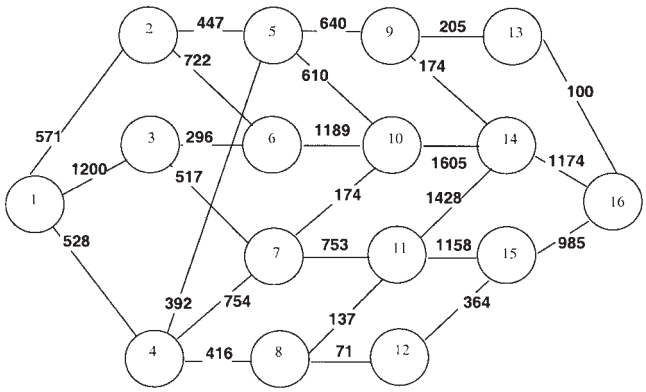

Fig. 4. The map of connectivity obtained from the MD simulation of penta-alanine. Each conformational state is represented by a circle, lines between circles account for direct conformation transitions, and the number of transitions during the 100 ns trajectory are written next to the line.

of these findings for the folding mechanism of longer peptides. Pathways A and B which start helix nucleation from either the N- or C-terminus side clearly may also occur in larger systems. Furthermore, a recent MD study on a 15-residue polyalanine revealed an α-helical H-bond forming at the center of the peptide.[45]

Let us finally consider two representative examples of the complete folding pathway from state 16 to state 1 as shown in Figure 2(d). Although the process essentially takes the above described folding pathways A and B, the figure clearly elucidates the diffusive character of the folding, involving numerous excursions to side states.

CONCLUSIONS

We have performed a 100-ns MD simulation of A_5 in explicit water and studied the reversible folding and unfolding of the peptide. Introducing a new PCA based on peptide dihedral angles, we have analyzed in detail the free energy landscape of A_5. The dPCA free energy surfaces shown in Figure 2 combined with the corresponding connectivity map in Figure 4 and the conformational analysis of Table I represent the central results of this work, which can be summarized as follows.

First, it has been demonstrated by Figure 2 that a correct separation of internal and overall motion is essential for the construction and interpretation of the energy landscape of a biomolecule undergoing large structural rearrangements. Internal coordinates fulfill this requirement in a natural way, and we believe that the proposed transformation in Eq. (3) of the peptide dihedral angles is advantageous compared to previous representations using internal coordinates.[13,19–24] This is because (1) the dPCA vector space is metric and therefore avoids problems associated with the periodicity of dihedral angles, (2) the dPCA vectors are of low dimension (unlike when using a PCA of the atom distances), and (3) the eigenvectors may be readily visualized in Cartesian space (unlike to an isotropically distributed ensemble analysis).

Second, the results obtained from the dPCA have characterized the free energy surface of a small peptide in explicit solvent in unprecedented detail. Figure 2 reveals that the energy landscape of A_5 is quite rugged, which reflects the fine counter-balance of the various enthalpic and entropic contributions of the folding process. The free energy surface contains numerous minima of comparable energy ($\lesssim$ 1 kcal/mol) and life times ($\approx$ 20 ps), which correspond to well-defined conformational states of the peptide. Furthermore, it has been shown that neighboring states are connected by low barriers of similar height and that each state typically makes transitions to only three or four neighbor states in the dPCA space. The typical time scale of such a transition involving the making or breaking of structural hydrogen bonds is about 100 ps, while the

overall folding process takes about 1 ns. It is interesting to note that the overall appearance of the dPCA energy landscape in Figure 2 resembles more a muffin tin than a funnel, although there is a slight overall slope towards the folded state. The funnel shape would become more apparent, if conformational states of higher energy were included (e.g., for a simulation at higher temperature) and if the folded state were lower in energy (i.e., for a longer peptide with a native fold).

Third, we have identified several principal pathways for helix nucleation, including paths leading to an initial α-helical H-bond at the N-terminal, at the C-terminal, as well as at the center of the peptide. Furthermore a folding pathway was found that involves a global hairpin structure. We have also considered representative examples of the complete folding pathway from the all-extended state 16 to the all-helical state 1, which revealed the diffusive character of the folding process. Extending this work to larger peptides, we hope to gain new insight on various topics of peptide folding theory, including, e.g., Flory's independent pair hypothesis and the ideas of helix-coil theory.[39,46]

Finally, we have found that also in the unfolded part of the free energy landscape there are well-separated minima corresponding to specific conformational states, e.g., structures 13, 14, 15, and 16. Although these conformations are often referred to as "random coil," the MD results show that the backbone dihedral angles are not randomly distributed in these unfolded conformations. For example, the most extended state, ensemble 16, maintains an average structure that is quite similar to the proline II conformation. We note that the proline II state was recently found to represent the dominant conformation of trialanine[47] and hepta-alanine.[48] These finding supports recent theoretical[49,50] and experimental[48,51] investigations which suggest that the unfolded state of peptides and proteins may be more structured and less random than previously thought.

ACKNOWLEDGMENTS

We thank Rainer Hegger and Harald Schwalbe for inspiring and helpful discussions. Financial support by the Deutsche Forschungsgemeinschaft, the National Natural Foundation of China (No. 90203013), and the Lee Kuan Yew Fellowship from Nanyang Technological University, Singapore, is gratefully acknowledged. The simulations were partially performed on the supercomputer of BIRC in NTU.

REFERENCES

1. Ball, K D, Berry, RS, Kunz, RE, Li, F-Y, Proykova, A, Wales DJ. From topographies to dynamics on multidimensional potential energy surfaces. Science 1996;271:963–965.
2. Onuchic, JN, Schulten, ZL, Wolynes, PG. Theory of protein folding: the energy landscape perspective. Annu Rev Phys Chem 1997;48:545–600.
3. Wales, DJ, Doye, JPK, Miller, MA, Mortenson, PN, Walsh, TR. Energy landscapes: From clusters to biomolecules. Adv Chem Phys 2000;115:1–111.
4. Gruebele M. Protein folding: the free energy surface. Curr Opin Struct Biol 2002;12:161–168.
5. Daura X, Gademann K, Jaun B, Seebach D, van Gunsteren WF, Mark AE. Peptide folding: when simulation meets experiment. Angew Chem Int Ed Engl 1999;38:236–240.
6. Karplus M, McCammon JA. Molecular dynamics simulations of biomolecules. Nat Struct Biol 2002;9:646–652.
7. Brooks CL III Protein and peptide folding explored with molecular simulations. Acc Chem Res 2002;35:447–454.
8. Gnanakaran S, Nymeyer H, Portman J, Sanbonmatsu KY, Garcia AE. Peptide folding simulations. Curr Opin Struct Biol 2003;13: 168–174.
9. Ichiye T, Karplus M. Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. Proteins 1991;11:205–217.
10. Garcia AE. Large-amplitude nonlinear motions in proteins. Phys Rev Lett 1992;68:2696–2699.
11. Amadei A, Linssen ABM, Berendsen HJC. Essential dynamics of proteins. Proteins 1993;17:412–425.
12. Hayward S, Kitao A, Hirata F, Go N. Effect of solvent on collective motions in globular proteins. J Mol Biol 1993;234:1207–1217.
13. Becker OM. Geometric versus topological clustering: an insight into conformation mapping. Proteins 1997;27:213–226.
14. de Groot BL, Daura X, Mark AE, Grubmüller H. Essential dynamics of reversible peptide folding: memory-free conformational dynamics governed by internal hydrogen bonds. J Mol Biol 2001;309:299–313.
15. Tournier AL, Smith JC. Principal components of the protein dynamical transition. Phys Rev Lett 2003;91:208106–1-208106-4.
16. McLachlan AD. Gene duplications in the structural evolution of chymotrypsin. J Mol Biol 1979;128:49–79.
17. Eckart, C. Some studies concerning rotating axes and polyatomic molecules. Phys Rev 1935;47:550–552.
18. Hünenberger PH, Mark AE, van Gunsteren WF. Fluctuation and cross-correlation analysis of protein motions observed in nanosecond molecular dynamics simulations. J Mol Biol 1995;252:492–503.
19. Prompers JJ, Brüschweiler R. Dynamics and structural analysis of isotropically distributed molecular ensembles. Proteins 2002;46: 177–189.
20. Schieborr U, Rüterjans H. Bias-free separation of internal and overall motion of biomolecules. Proteins 2001;45:207–218.
21. Prompers JJ, Brüschweiler R. General framework for studying the dynamics of folded and nonfolded proteins by NMR relaxation spectroscopy and MD simulation. J Am Chem Soc 2002;124:4522–4534.
22. van Aalten DMD, de Groot BL, Finday JBC, Berendsen HJC, Amadei A. A comparison of techniques for calculating protein essential dynamics. J Comput Chem 1997;18:169–181.
23. Elmaci N, Berry RS. Principal coordinate analysis on a protein model. J Chem Phys 1999;110:10606–10622.
24. Abseher R, Nilges M. Are there non-trivial dynamic cross-correlations in proteins? J Mol Biol 1998;279:911–920.
25. Hummer G, Garcia AE, Garde S. Helix nucleation kinetics from molecular simulations in explicit water. Proteins 2001;42:77–84.
26. Margulis CJ, Stern HA, Berne BJ. Helix unfolding and intramolecular hydrogen bond dynamics in small α-helices in explicit solvent. J Phys Chem B 2002;106:10748–10752.
27. Berendsen HJC, van der Spoel D, van Drunen R. Gromacs: a message-passing parallel molecular dynamics implementation. Comp Phys Comm 1995;91:43–56.
28. Lindahl E, Hess B, van der Spoel D. Gromacs 3.0: a package for molecular simulation and trajectory analysis. J Mol Mod 2001. 7:306–317.
29. Schuler LD, Daura X, van Gunsteren WF. An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase. J Comput Chem 2001;22:1205–1218.
30. Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J. Intermolecular forces. Dortrecht: Reidel; 1981. p 331–342.
31. Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical-integration of cartesian equations of motions of a system with constraints-molecular dynamics of n-alkanes. J Comput Phys 1977;23:327–341.
32. van Gunsteren WF, Berendsen HJC. Computer simulation of molecular dynamics-methodology, applications, and perspectives in chemistry. Angew Chem Int Ed Engl 1990;29:992–1023.
33. Berendsen HJC, Postma JPM, van Gunsteren WF, Dinola JRH. Molecular dynamics with coupling to an external bath. J Chem Phys 1984;81:3684–3690.
34. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM Jr,

Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J Am Chem Soc 1995;117: 5179–5197.
35. Jorgensen WL, Maxwell DS, Tirado-Rives J. Development and testing of the OPLS all atom force field on conformational energetics and properties of organic liquids. J Am Chem Soc 1996;118: 11225–11236.
36. Hu H, Elstner M, Hermans J. Comparison of a qm/mm force field and molecular mechanics force fields in simulations of alanine and glycine "dipeptide" (ace-ala-nme and ace-gly-nme) in water in relation to the problem of modeling the unfolding peptide backbone in solution. Proteins 2003;50:451–463.
37. Mu YG, Kosov DS, Stock G. Conformational dynamics of trialanine in water.2.comparison of amber, charmm, gromos, and opls force field to nmr and infrared experiments. J Phys Chem B 2003;107:5064–5073.
38. Gnanakaran S, Garcia AE. Validation of an all-atom protein force field: from dipeptides to larger peptides. J Phys Chem B 2003;107: 12555–12557.
39. Zaman MH, Shen MY, Berry RS, Freed KF, Sosnick TR. Investigations into sequence and conformational dependence of backbone entropy, inter-basin dynamics and the Flory isolated-pair hypothesis for peptides. J Mol Biol 2003;331:693–711.
40. MacKerell AD, Feig JM, Brooks CL. Improved treatment of the protein backbone in empirical force fields. J Am Chem Soc 2004;126:698–699.
41. Graf J, Nguyen P, Mu Y, Stock G, Schwalbe H. Structure and dynamics of small alanine peptides: a joint molecular-dynamics/ nmr study.2004. Submitted for publication.
42. Kabsch W, Sander C. Dictionary of protein secondary structure: Pattern recognition of hydrogen bonded and geometrical features. Biopolymers 1983;22:2577–2637.
43. Zimm BH, Bragg JK. Theory of the phase transition between helix and random coil in polypeptide chains. J Chem Phys 1959;31:526–535.
44. Lifson S, Roig A. Theory of helix-coil transition in polypeptides. J Chem Phys 1961;34:1963–1974.
45. Takano M, Yamato T, Higo J, Suyama A, Nagayama K. Molecular dynamics of a 15 residue poly(l-alanine) in water: helix formation and energetics. J Am Chem Soc 1999;121:605–612.
46. Ohkubo YZ, Brooks CL III. Exploring Flory's isolated-pair hypothesis: statistical mechanics of helix-coil transitions in polyalanine and the C-peptide from RNase A. Proc Natl Acad Sci USA 2003;100:13916–13921.
47. Woutersen S, Pfister R, Hamm P, Mu YG, Kosov DS, Stock G. Peptide conformational heterogeneity revealed from nonlinear vibrational spectroscopy and molecular-dynamics simulations. J Chem Phys 2002;117:6833–6840.
48. Shi Z, Olson CA, Rose GD, Baldwin RL, Kallenbach NR. Polyproline II structure in a sequence of seven alanine residues. Proc Natl Acad Sci USA 2002;99:9190–9295.
49. van Gunsteren WF, Ḃrgi R, Peter C, Daura X. The key to solving the protein-folding problems lies in an accurate description of the denatured state. Angew Chem Int Ed Engl 2001;40:351–355.
50. Zagrovic B, Snow CD, Khaliq S, Shirts MR, Pande VS. Native-like mean structure in the unfolded ensemble of small proteins. J Mol Biol 2002;323:153–164.
51. Klein-Seetharaman J, Oikawa M, Grimshaw SB, Wirmer J, Duchardt E, Ueda T, Imoto T, Smith LJ, Dobson CM, Schwalbe H. Long-range interactions within a nonnative protein. Science 2002; 295:1719–1722.