Vol 14 No 11, November 2005
1009-1963/2005/14(11)/2365-05

Chinese Physics

© 2005 Chin. Phys. Soc.
and IOP Publishing Ltd

Functional structures and folding dynamics of two peptides*

Sheng Yue-Biao(盛乐标)[a], Li Jing(李 菁)[a],
Ma Bao-Liang(马保亮)[a], and Wang Wei(王 炜)[a,b]†

[a] *National Laboratory of Solid State Microstructure, Institute of Biophysics, and Department of Physics, Nanjing University* 210093, *China*

[b] *Interdisciplinary Center of Theoretical Studies, Chinese Academy of Sciences, Beijing* 100080, *China*

(Received 15 March 2005; revised manuscript received 8 May 2005)

The folding dynamics and structural characteristics of peptides RTKAWNRQLYPEW (P1) and RTKQLYPEW (P2) are investigated by using all-atomic simulation procedure CHARMM in this work. The results show that P1, a segment of an antigen, has a folding motif of α-helix, whereas P2, which is derived by deleting four residues AWNR from peptide P1, prevents the formation of helix and presents a β-strand. And peptide P1 experiences a more rugged energy landscape than peptide P2. From our results, it is inferred that the antibody CD8 cytolytic T lymphocyte prefers an antigen with a β-folding structure to that with an α-helical one.

Keywords: peptide folding, molecular dynamics, protein secondary structure prediction
PACC: 8715B, 8715H, 8715K

1. Introduction

Protein folding and unfolding are essential events in cells. In the past two decades, much progress has been achieved. Nowadays, it is generally accepted that folding resembles a diffusive process on a rugged funnel-like energy landscape.[1−11] On the study of protein folding, the all-atomic simulations play a very important role since much detailed information of folding can be obtained. Despite some good agreement between the all-atomic simulations and the experiments, the gap between the timescales of experimental studies and the simulations for proteins with typical size, say 100 residues, remains unbridged. The timescale which the all-atomic simulations can reach is generally known as nanoseconds to microseconds today. However, the timescale of protein folding events is typically several milliseconds to tens of seconds.[12] Since the first molecular dynamics simulation for the bovine pancreatic trypsin inhibitor (BPTI) about 28 years ago, many efforts have been made to increase the accuracy of the various all-atomic models, to improve the methods of the computations, and to extend the timescale of the simulations.[12−17] Recently, with the advancement of experimental techniques, some small, ultra-fast folding proteins or peptides have been found. Now it seems to be possible to simulate some ultra-fast folding events.[12,18]

Among the objects behaving the ultra-fast folding behaviours, peptide is the representative and indispensable one. Peptide folding may be tractable computationally. Actually, the study on peptide folding is a first step to catch on some basic features of folding, and is also critical for understanding the molecular mechanisms of early events of protein folding. It may offer a unique opportunity to bridge the gap between theoretical and experimental understanding on protein folding.[19] However, because of their high flexibility, it is not a trivial thing to determine the structures of small peptides in solution. Thus the prediction on how peptides fold in solution becomes much important in structural biology. In addition, peptides fold very fast in experiment, which requires probing on the nanosecond time resolution. Theoretically, it demands the high-resolution simulations based on atomic details. Moreover, the study related to the stereo struc-

*Project supported by the National Natural Science Foundation of China (Grant Nos 90103031, 10474041, 90403120 and 10021001), and the Nonlinear Project (973) of the NSM.

†Corresponding author. E-mail: wangwei@nju.edu.cn

tures and biological functions of peptides has become a hot topic in the realm of drug design and discovery nowadays. Therefore, the study of peptide folding is very interesting and exciting, while also imbued with challenges.

Peptide ARG - THR - LYS - ALA - TRP - ASN - ARG - GLN - LEU - TYR - PRO - GLU - TRP (termed P1, single-letter abbreviation: RTKAWNRQLYPEW) is a segment of the antigen of CD8 cytolytic T lymphocyte (CTL 14, which is derived by stimulating blood lymphocytes of a melanoma patient with autologous tumour cells), and the antigen is presented by human lymphocyte antigen (HLA)-A32.[20] P1 contains 13 amino acids, which are mostly hydrophilic except the residues ALA4 and LEU9. Peptide ARG-THR-LYS-GLN-LEU-TYR-PRO-GLU-TRP (termed P2, single-letter abbreviation: RTKQLYPEW) is a nonomer which is produced from peptide P1 by removal of residues ALA-TRP-ASN-ARG (single-letter abbreviation: AWNR). P2 contains only one hydrophobic residue LEU5 and cannot form hydrophobic core, thus different motifs between peptides P1 and P2 can be expected. In experiment, Vigneron *et al* considered that the peptide antigenicity would depend on a post-translational modification. Then they verified such a guess by introducing synthetic peptides including P1 inside HLA-A32$^+$ target cells by electroporation before testing for CTL recognition, and they observed that peptide P1 could be recognized by CTL 14 after electroporation into autologous Epstein-Barr virus transformed B cells, but could not be recognized after exogenous loading on these cells. They also found that peptide P2 could be efficiently recognized by CTL 14 after exogenous loading. Thus they came to a conclusion that P1 had been modified to be P2 by splicing AWNR in the target cells in order to perform its function, and P2 was the very peptide identical to the one presented by melanoma cells.[20] Therefore, it is easy to image that there are some differences in structure between peptides P1 and P2, and these differences must result from the specific features of the sequences and the intra-interactions of these two peptides, and also the interactions between the peptides and CTL 14.

In this paper, we aim to investigate the folding pathways and structural characteristics of these two peptides. Based on all-atomic simulations using CHARMM,[21] the folded structures for the two peptides are obtained. To verify our simulations, the DSC (discrimination of protein secondary structure class) secondary protein structure prediction[22] is also made. In our work, we deem that the antigenic activity is highly related to the formed structure, and a β-folding structure for P2 is argued to fit the biological function of an antigen of the antibody CTL 14.

2. Methods

Molecular dynamics simulations are performed by using the program CHARMM[21] with the all-hydrogen parameter set PARAM22. Both peptides P1 and P2 are solvated by the TIP3P explicit water molecules. The SHAKE algorithm is used to fix the lengths of bonds involving hydrogen atoms. The folding simulations of these two peptides are performed by using constant pressure/temperature dynamic technique. Four runs of simulations for each peptide are carried out at 300K for the isobaric-isothermal ensemble. The initial structures of the two peptides are both extended chains. For each run, the peptide is solvated in a rhombic dodecahedral box of explicit water molecules. Periodic boundary conditions are used to minimize the boundary effects. Particle Mesh Ewald algorithm is used for calculating the electrostatic interactions.[23] The time step of integration for the potential energy is taken as 2 fs. Structures of the peptide molecule obtained from simulations are saved every 1 ps for analysis. The simulations are performed for 20ns each at 300K, and the former 15ns trajectories are used for analysis. The final structures for P1 and P2 are shown in Fig.1. The structures at $t = 15$ns for P1 and $t = 14$ ns for P2 are used respectively as the reference structures for latter analysis.

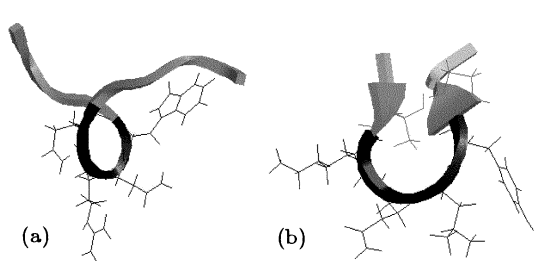


Fig.1. (a) The final folding structure of peptide P1. The backbone is shown in cartoon, and the side-chains of the helix composed of residues WNRQ are shown in wireframe. The helical content of the peptide is coloured in black, with the other parts in grey. (b) The final folding structure of peptide P2. The backbone is shown in cartoon, and the side-chains of residues 2 to 7 are shown in wireframe. The turn is coloured in black, while the β-fold is shown in dark grey.

3. Results and discussion

To characterize the folding process, three main quantities are employed, namely the root mean square deviation (RMSD) of the C_α atoms, the radius of gyration (R_g), and the solvent accessible surface area (SASA). RMSD denotes the dissimilarity of two structures since it takes the whole details of protein structures. R_g is a good measure of the dynamics for the collapse of a protein, which represents the compactness of a conformation quantitatively. SASA is a factor to measure the hydrophobic effect. The hydrophobic force is considered to be one of the most important one among various forces which determine the tertiary structures of proteins. SASA is traced out by the probe sphere centre as it rolls over the protein. An atom or a group of atoms is defined as accessible if a solvent (water) molecule of specified size can be brought into the van der Waals' contact.

Firstly we report the results of the folding process for P1. At the beginning, a rapid decrease in C_α RMSD and R_g corresponding to the process of quick structural reorganization at the folding initial stage is observed (Fig.2). Then at about 3 ns, the two hydrophobic residues ALA4 and LEU9 begin to exclude the water molecules and form the hydrophobic core. At about $t = 10$ ns, the hydrophobic core forms completely. Before this, a sharp peak is observed in the plot of SASA at about 6.5 ns, and the related peaks can also be seen in the plots of C_α RMSD and R_g. This implies that a significant barrier is encountered during the rotation of the side-chain of the LEU9. Once the peptide overcomes this barrier, the hydrophobic core forms rapidly and the structure becomes more compact. After $t = 10$ ns, the peptide begins its readjustment to continue its folding. At last, the peptide reaches the folded state at about 14 ns. Structurally, the final conformation of peptide P1 has an obvious hole formed by residues TRP5, ASN6, ARG7 and GLN8, with ALA4 and LEU9 as the seal of the hole. Residues 5 to 8 constitute a small piece of helix as shown in black in Fig.1(a).

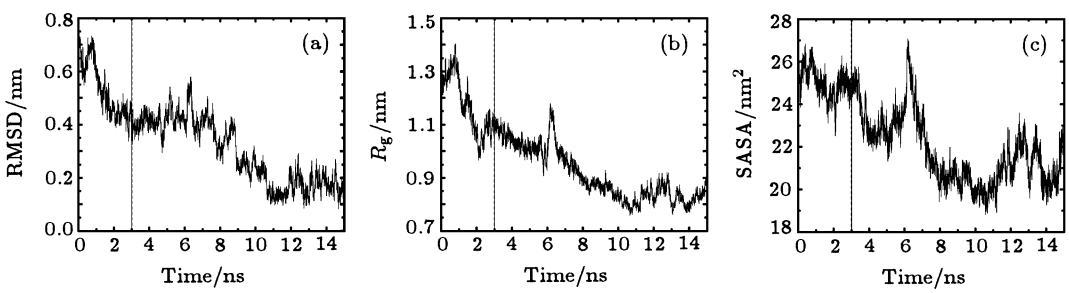


Fig.2. Global properties of the peptide P1. (a) C_α RMSD versus time. (b) R_g versus time. (c) SASA versus time. And the time of 3 ns is marked with a dotted line.

Peptide P2 is the nonomer produced by removal of four residues AWNR from peptide P1. Since the hydrophobic residue ALA is deleted, P2 contains only one hydrophobic residue. Thus different from the folding of P1, no hydrophobic core appears during its folding. As can be clearly seen in Fig.3, the small variation of SASA indicates that no special hydrophobic secondary structure forms during the whole process. After the phase of structural reorganization, peptide P2 begins to shrink under the driving of energy and entropy at about 3 ns, which is revealed by the decreasing of C_α RMSD and R_g in Fig.3. The decreasing continues until an abruption occurs at 9.3 ns. Such an abruption implies that the forming of the β-strand between THR2 and PRO7 encounters an inevitable barrier. After overcoming this barrier, the peptide arrives at its folding end-point rapidly and achieves a balance between the energy and the entropy at 13.8 ns. From the final structure, an obvious β-strand formed by THR2 and PRO7 is observed [cf. Fig.1(b)].

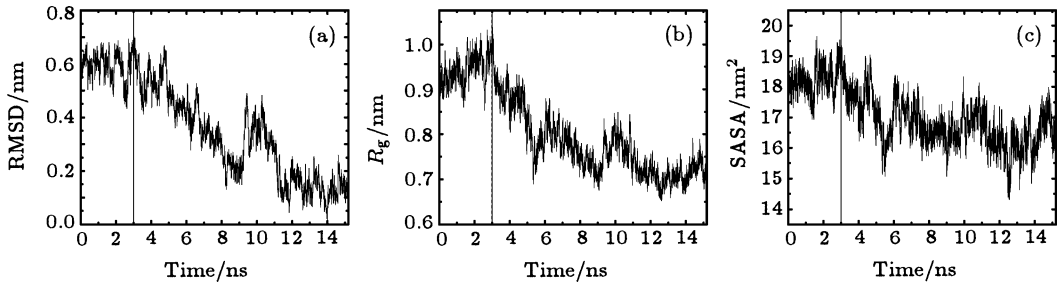


Fig.3. Global properties of the peptide P2. (a) C_α RMSD versus time. (b) R_g versus time. (c) SASA versus time. And the time of 3ns is marked with a dotted line.

As we know, it is nearly impossible to obtain the detailed folding pathway thermodynamically. Nevertheless, it would be achievable to take a glance into the outline of the folding pathway in dynamics. Here a trick is made to characterize the folding pathway, that is to replace the free energy, the third dimensional variable in the conventional contour plot of free energy with RMSD. And the two reaction coordinates are R_g and number of contacts respectively. A contact is defined when the distance between any pair of heavy atoms of two residues i and j ($j > i + 2$) is less than 0.54nm. The contour plots of C_α RMSDs for peptides P1 and P2 are both shown in Fig.4. The pathways of peptide P1 seem to be more rugged than that of P2. In the figure, two major pathways are observed for peptide P1, while only one is observed for peptide P2. For P1, 14 contacts are formed in the folded structure, whereas 10 contacts are formed in the folded structure of P2 due to the shrink of the size. Also we observed the decrease of R_g due to the shortened length of the sequence for P2. The discreteness of the regions in the plots is mainly due to the inadequate sampling of the conformational space, which is inevitable at the present time, as it is still beyond the current computer capacity to gain an accurate and perfect contour plot.

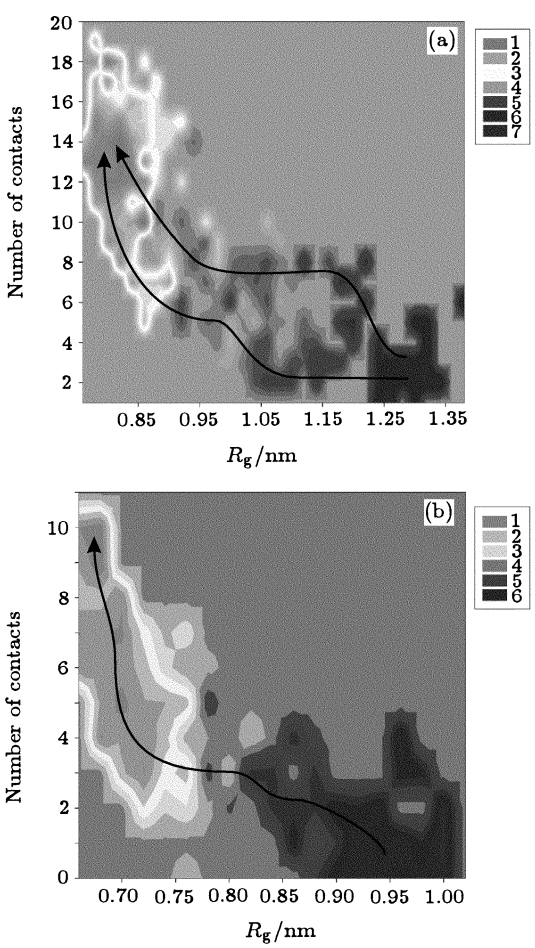


Fig.4. Contour plots of RMSDs for peptides P1 (a) and P2 (b) respectively.

Now, let us compare the structures obtained from simulations with that from the protein secondary structure prediction based on the sequences of peptides P1 and P2. The prediction is made at the NPS server.[24] The structures from the simulations are well consistent with those from the secondary structure prediction based on DSC algorithm. The detailed results are shown in Table 1. At the NPS server, a BLASTP[25] search of the sequence is performed against the SWISS-PROT database. These results are filtered and then aligned by CLUSTALW.[26] The resulting alignment is then input for DSC prediction. DSC is a protein secondary structure prediction method from multiply-aligned homologous sequences with a three-state (helix, sheet and coil) accuracy of $\sim$70%. For each residue position the following are calculated in this method: mean GOR potential, distance to end of chain, mean moment of hydrophobicity, existence of insertions and deletions, and the mean moment of conservation. Then these attributes are smoothed and a linear discrimination function is applied to make a prediction for each residue position. The fraction of residues predicted to be α-helix and

β-strand per protein are then recalculated, which is used to make a final prediction by a second linear discrimination function.

Table 1. Comparison of structures from the simulations and the sequence prediction of secondary structure. The letter H represents α-helix, E represents β-strand, and C for random coil.

Sequence	RTKAWNRQLYPEW	RTKQLYPEW
Simulation	CCCCHHHHCCCCC	CECCCCECC
Prediction	CCCHHHHHCCCCC	CECCCCECC

From the analysis above, we can easily gain the knowledge that after removing residues AWNR from the peptide P1, the hydrophobic core cannot form, thus the helical structure stabled by the two hydrophobic residues of course will not appear. In other words, the major contribution to the formation of helix is from residues AWNRQL, and slicing the segment AWNR will prevent the formation of helix. Therefore, the difference in structures might be the main cause for the efficient recognition of P2 but not P1 by CTL 14. Once the peptide with specific structure combines to the corresponding active site of CTL 14, it would be stabled and act as an antigen to interactive with the antibody, i.e., the antigenic activity is highly related to the peptide structure formed.

4. Conclusions

In summary, we have studied the folding processes of two peptides P1 and P2 by using all-atomic simulations. The folding pictures for these two peptides are obtained. The differences in structures are well distinguished. The folded structures from the simulations are well consistent with those derived from the DSC secondary protein structure prediction. It should be noted that since the lengths of these two peptides are short and highly flexible, the final folded structures of the peptides could not keep stable all along. That is to say, only when the peptides combine to the special protein or antibody, they can form long-lasting stable structures and perform their biological function. To conclude, we infer that the antibody CTL 14 is not sensible to α-helical structures of P1 but prefer β-folding ones of P2.

References

[1] Wolynes P G, Onuchic J N and Thirumalai D 1995 *Science* **267** 1619
[2] Onuchic J N, Schulten Z L and Wolynes P G 1997 *Annu. Rev. Phys. Chem.* **48** 545
[3] Dill K A and Chan H S 1997 *Nature Struct. Biol.* **4** 10
[4] Socci N D, Onuchic J N and Wolynes P G 1998 *Proteins* **32** 136
[5] Chan H S and Dill K A 1998 *Proteins* **30** 2
[6] Plotkin S S and Onuchic J N 2002 *Q. Rev. Biophys.* **35** 111
[7] Plotkin S S and Onuchic J N 2002 *Q. Rev. Biophys.* **35** 205
[8] Chen J, Wang J and Wang W 2004 *Proteins* **57** 153
[9] Sheng Z M and Luo J W 2003 *Acta Phys. Sin.* **52** 2342
[10] Li W *et al* 2003 *Chin. Phys.* **12** 226
[11] Ji G F, Xue B and Wang W 2004 *Chin. Phys.* **13** 1574
[12] Daggett V and Fersht A 2003 *Nature Rev. Mol. Cell Biol.* **4** 497
[13] Karplus M 2003 *Biopolymers* **68** 350
[14] Radford S E 2000 *Trends Biochem. Sci.* **25** 611
[15] Duan Y and Kollman P A 1998 *Science* **282** 740
[16] Duan Y, Wang L and Kollman P A 1998 *Proc. Natl. Acad. Sci. USA* **95** 9897
[17] Pande V S *et al* 2003 *Biopolymers* **68** 91
[18] Mayor U *et al* 2003 *Nature* **421** 863
[19] Gnanakaran S *et al* 2003 *Curr. Opin. Struc. Biol.* **13** 168
[20] Vigneron N *et al* 2004 *Science* **304** 587
[21] Brooks B R *et al* 1983 *J. Comp. Chem.* **4** 187
[22] King R D and Sternberg M J 1996 *Protein Sci.* **5** 2298
[23] Essmann U *et al* 1995 *J. Chem. Phys.* **103** 8577
[24] Combet C *et al* 2000 *Trends Biochem. Sci.* **25** 147
[25] Samuel K and Altschul S F 1990 *Proc. Natl. Acad. Sci. USA* **87** 2264
[26] Higgins D G and Sharp P M 1988 *Gene* **73** 237