Computer Physics Communications 123 (1999) 97–102

Computer Physics Communications

www.elsevier.nl/locate/cpc

Peptide folding simulations: no solvent required?

Xavier Daura, Alan E. Mark, Wilfred F. van Gunsteren *

Laboratory of Physical Chemistry, Swiss Federal Institute of Technology Zurich, ETH Zentrum, CH-8092 Zurich, Switzerland

Received 25 March 1999

Abstract

Several simulation methods are currently in use to study peptide and protein folding. They can be classified in three groups depending on how the solvent is treated. At the simplest level, the solvent is ignored. At a second level, solvent effects are implicitly represented in the atomic interaction function. At the third level, solvent degrees of freedom are treated explicitly. We have performed molecular dynamics simulations in vacuum, stochastic dynamics simulations, and molecular dynamics simulations in methanol of the folding of a β-heptapeptide to test these different levels of approximation. The results clearly show that the solvent cannot be ignored in folding simulations, and highlight the difficulties of implicitly modelling solvent effects in an atomic interaction function for a solute.

Keywords: Computer simulation; Molecular dynamics; Stochastic dynamics; β-peptides; Peptide folding; Simulation in vacuum; Simulation in methanol; Solvent effects

1. Introduction

One of the challenging unsolved problems in molecular biology is the mechanism of peptide and protein folding. Being able to predict the most stable fold of a peptide or protein in a given environment as well as the thermodynamics of folding (and unfolding) is key to the design of de novo peptides and proteins with target properties. However, the study of peptide and protein folding by experimental means is difficult. The information obtained is ambiguous and in the best cases only refers to a few metastable intermediates of folding, at conditions that are in general far from physiological. Molecular simulation methods have thus become a fundamental tool complementary to experimental observations [1,2]. Several simulation methods are currently in use to study folding [3]. They can be classified in three groups depending on how the solvent is treated [4]. At the simplest level, the solvent is ignored. Examples are molecular dynamics (MD) simulations in vacuum [5], lattice simulations [6–9], and off-lattice simulations with simplified molecular models [10–12]. Using these methods, folding is reduced to a problem of chain enthalpy and entropy. At the second level of complexity, the solvent is treated as an external field acting on the peptide or protein. This takes the form of extra terms in the potential energy function and equation of motion. Examples are stochastic dynamics (SD) simulations [13,14] and MD simulations with a mean-force potential energy term (e.g. proportional to the solvent-accessible surface area of the peptide or protein) [15–19]. Both exist with different levels of sophistication of the potential energy function. It is difficult

* Corresponding author; e-mail: wfvgn@igc.phys.chem.ethz.ch

0010-4655/99/$ – see front matter
PII: S0010-4655(99)00261-1

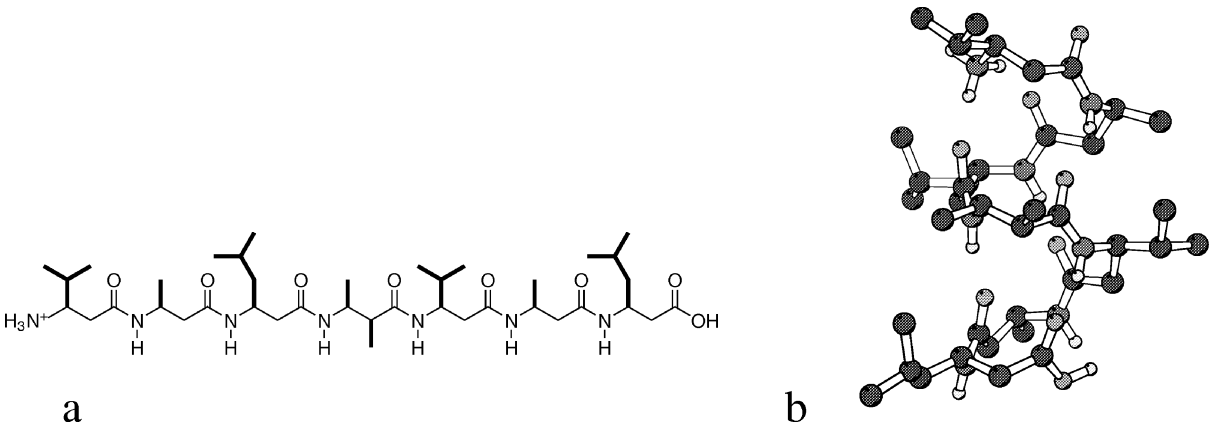

Fig. 1. (a) Structural formula of the β-heptapeptide studied. In the simulations both end groups were protonated in line with experimental data. (b) Left-handed 3_1-helical structure of the β-heptapeptide, determined experimentally by NMR in methanol at 298 K.

or even impossible, however, to include all potential solvent effects on the dynamics of a peptide or protein in a model of this sort. For example, changes in solvent entropy in the first solvation shells upon folding and unfolding cannot be directly modelled in a potential energy term. At the third level of complexity, the solvent degrees of freedom are treated explicitly, as in MD simulations with explicit solvent [20–22]. Because of computational limitations, studies of protein folding are essentially restricted to the first two levels, i.e. without treating the solvent explicitly. MD simulations including solvent explicitly have, nevertheless, reached a state in which it is possible to simulate the reversible folding of small peptides in solution [20,21]. This allows one to test the different levels of approximation. In this article we present results of MD simulations in vacuum, SD simulations, and MD simulations in methanol of a β-heptapeptide [20] (Fig. 1) at four different temperatures in order to address the question of the title: is explicit treatment of solvent degrees of freedom in peptide folding simulation studies necessary?

2. Methods

Twelve 50 ns MD simulations were performed using the GROMOS96 package of programs [23]. The dynamics of the β-heptapeptide (Fig. 1) was studied at a series of temperatures, 298 K, 340 K, 350 K, and 360 K, by MD simulation in methanol [20], by SD simulation, and by MD simulation in vacuum. In the MD simulations in methanol the GROMOS96 43A1 force field (for solvated biomolecules) was used [23]. The pressure of the system was 1 atm and periodic boundary conditions were applied. The temperature and pressure were maintained by weak coupling to an external bath [24]. The initial structure of the peptide for the simulations at 298 K, 340 K, and 350 K was the left-handed 3_1-helical fold shown in Fig. 1b. The system contained the β-heptapeptide and 962 methanol molecules in a rectangular box. 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 1778 methanol molecules in a truncated octahedron. A twin-range cutoff of 0.8 nm/1.4 nm was used for all non-bonded interactions. The shortest distance peptide-wall was initially 1.4 nm. Full details of these simulations have been described previously [20]. In the SD simulations and the MD simulations in vacuum the GROMOS96 43B1 force field was used [23]. This force field is the one corresponding to the 43A1 force field, but adapted in order to be used for simulations of biomolecules in vacuum. The initial structure of the peptide was as above, i.e. left-handed 3_1-helical at 298 K, 340 K, and 350 K, and extended at 360 K. The atomic friction coefficients used in the SD

simulations (60 ps^{-1} at 298 K, 40 ps^{-1} at 340 K, 35 ps^{-1} at 350 K, and 30 ps^{-1} at 360 K) were calculated using Stokes' law based on stick boundary conditions [14], an approximate value of 0.28 nm for the Stokes radius of the methanol model, and the viscosity of liquid methanol (approximately 0.6 cP at 298 K, 0.4 cP at 340 K, 0.35 cP at 350 K, and 0.3 cP at 360 K). A single friction coefficient was used for all atoms in the peptide. In the MD simulations in vacuum the temperature was maintained by weak coupling to an external bath, with a relaxation time of 0.1 ps. In the SD simulations and the MD simulations in vacuum no cutoff was used for the non-bonded interactions.

To find clusters of similar structures in a trajectory, the atom-positional root-mean-square deviation (RMSD) between all pairs of structures was determined for a pool of 5000 structures taken at 0.01 ns intervals from the trajectory [25]. For each structure the number of other structures for which the backbone atom-positional RMSD was $\leqslant$ 0.1 nm (residues 2–6) (neighbour conformations) was calculated. The structure with the highest number of neighbours was taken as the centre of a cluster, and formed together with all its neighbours a (first) cluster of similar structures. The structures of this cluster were thereafter eliminated from the pool of structures. The process was repeated until the pool of structures was empty. In this way, a series of non-overlapping clusters of similar structures was obtained.

3. Results

The twelve trajectories of the β-heptapeptide are presented in Fig. 2 in terms of atom-positional root-mean-square deviation (RMSD) from the model left-handed 3_1-helix (Fig. 1b) as a function of time. A backbone RMSD $\leqslant$ 0.1 nm (residues 2 to 6) from the model structure has been used as a criterion for left-handed 3_1 helicity [20]. We will refer to this type of conformation as folded, as it corresponds to the predominant conformation in methanol solution at room temperature determined experimentally [26]. Figs. 2a–d correspond, respectively, to the MD simulations in methanol at 298 K, 340 K, 350 K, and 360 K. At 298 K the peptide is for most of the simulation time in its folded conformation. At 340 K, 350 K, and 360 K the peptide undergoes a series of folding and unfolding transitions. The percentage of folded structures in these simulations decreases with increasing temperature, as it is expected. The melting temperature of the peptide in the force field, defined as the temperature at which the free energy of folding is zero, is around 340 K [20]. Figs. 2e–h correspond, respectively, to the SD simulations at 298 K, 340 K, 350 K, and 360 K. At 298 K, 340 K, and 350 K, where the initial structure was left-handed 3_1-helical, the peptide unfolds after about 2.5 to 7.5 ns (depending on the temperature) and never folds again. At 360 K, where the initial structure was totally extended, the peptide never folds to the left-handed 3_1-helical form. Figs. 2i–l correspond, respectively, to the MD simulations in vacuum at 298 K, 340 K, 350 K, and 360 K. At 298 K the peptide unfolds after about 2.5 ns and never folds again. At 340 K and 350 K the peptide unfolds at the beginning of the simulation and folds briefly only twice and once, respectively, during the 50 ns. At 360 K, starting from the extended structure, the peptide folds briefly only once during the 50 ns.

In order to compare the regions of conformational space that have been sampled at each of the temperatures with each of the types of simulation, cluster analysis has been carried out. In the MD simulations in methanol a total of 9 clusters at 298 K, 158 clusters at 340 K, 137 clusters at 350 K, and 219 clusters at 360 K were found [25]. The first (i.e. most populated) cluster corresponds to the left-handed 3_1-helical fold, and incorporates approximately 97% of the ensemble at 298 K, 50% at 340 K, 39% at 350 K, and 25% at 360 K. The central structure of the first cluster has a backbone atom-positional RMSD (residues 2 to 6) from the NMR model structure of 0.05 nm at 298 K, 0.06 nm at 340 K, 0.05 nm at 350 K, and 0.06 nm at 360 K. In the SD simulations a total of 11 clusters at 298 K, 17 clusters at 340 K, 35 clusters at 350 K, and 93 clusters at 360 K were found. The first cluster incorporates approximately 79% of the ensemble at 298 K, 72% at 340 K, 49% at 350 K, and 23% at 360 K. The central structure of the first cluster has a backbone atom-positional RMSD (residues 2 to 6) from the NMR model structure of 0.23 nm at 298 K, 0.33 nm at 340 K, 0.33 nm at 350 K, and 0.18 nm at 360 K. The most populated conformations (first clusters) in the four SD simulations represent, in terms of secondary structure and by analogy to α-peptides, (imperfect)

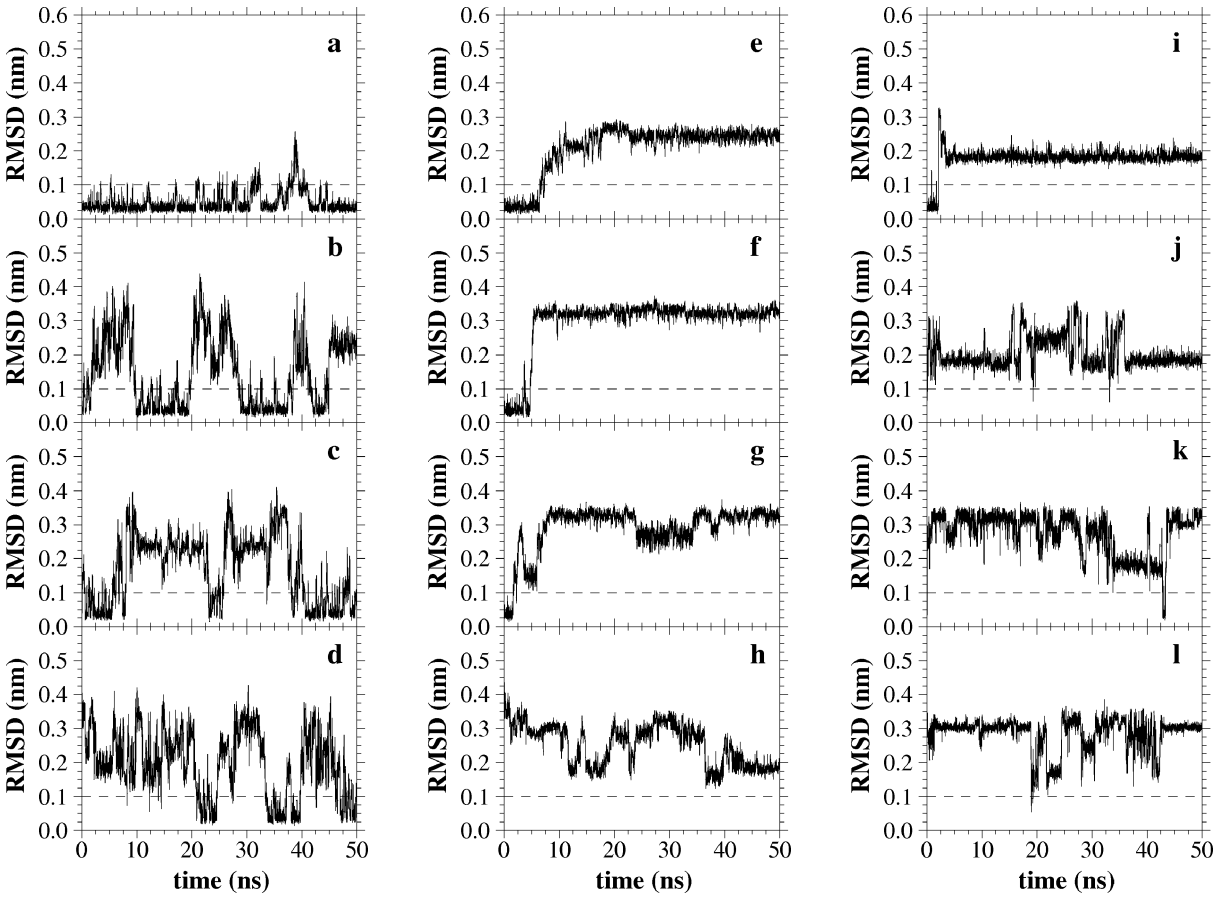

Fig. 2. Backbone atom-positional root-mean-square deviation (RMSD) from the left-handed 3_1-helical model structure (Fig. 1b) for residues 2 to 6 as a function of time, for the three types of simulation at four different temperatures: (a) Molecular Dynamics (MD) simulation in methanol at 298 K; (b) MD simulation in methanol at 340 K; (c) MD simulation in methanol at 350 K; (d) MD simulation in methanol at 360 K; (e) Stochastic dynamics (SD) simulation at 298 K; (f) SD simulation at 340 K; (g) SD simulation at 350 K; (h) SD simulation at 360 K; (i) MD simulation in vacuum at 298 K; (j) MD simulation in vacuum at 340 K; (k) MD simulation in vacuum at 350 K; (l) MD simulation in vacuum at 360 K.

hairpin-like folds. In particular, the most populated conformations at 340 K and 350 K represent the same fold, with a backbone RMSD between the central structures of the respective first clusters of 0.04 nm (residues 2 to 6). In the MD simulations in vacuum a total of 16 clusters at 298 K, 90 clusters at 340 K, 113 clusters at 350 K, and 150 clusters at 360 K were found. The first cluster incorporates approximately 93% of the ensemble at 298 K, 59% at 340 K, 31% at 350 K, and 42% at 360 K. The central structure of the first cluster has a backbone atom-positional RMSD (residues 2 to 6) from the NMR model structure of 0.18 nm at 298 K, 0.18 nm at 340 K, 0.33 nm at 350 K, and 0.30 nm at 360 K. The most populated conformations (first clusters) in the MD simulations in vacuum at 298 K, 340 K, and 350 K represent again hairpin-like folds. In particular, the most populated conformations at 298 K and 340 K represent the same fold, with a backbone RMSD between the central structures of the respective first clusters of 0.03 nm (residues 2 to 6). The most populated conformation in the MD simulation in vacuum at 360 K is a random-coil type of conformation.

4. Discussion

The three types of simulations performed produce widely different results. The MD simulations in methanol correctly predict a left-handed 3_1-helix as the most stable fold of the β-heptapeptide in methanol at room temperature. The left-handed 3_1-helix is the predominant fold at any of the four temperatures studied, with its probability decreasing as the temperature increases. The SD simulations and the MD simulations in vacuum predict, in contrast, a hairpin-like conformation as the most stable fold of the β-heptapeptide at room temperature. The predominant conformation, however, changes with temperature. The left-handed 3_1-helix is rarely sampled in the MD simulations in vacuum, and never sampled (starting from a random conformation) in the SD simulations. The number of different conformations (clusters) sampled is remarkably lower in the MD simulations in vacuum than in the MD simulations in methanol, and lower again in the SD simulations. In the simulations without explicit solvent the peptide is more easily trapped in local energy minima. The critical problem is, however, that the free energy hypersurfaces of the three types of systems appear to be quite different.

The results demonstrate that correct treatment of solvent is essential to realistic simulation of peptide folding. There exist much more sophisticated potential energy functions with implicit solvent representation for use in SD and MD simulations than the one used in this work, which may produce quantitatively different results. However, the process of peptide folding in solution is driven by small free energy differences. These free energy differences involve the system as a whole. Separating one part of the system from the other and replacing it by a mean-force potential, while still getting comparable free energy hypersurfaces, will be difficult if not impossible. Mean-force potentials mimicking solvent effects currently used in SD and MD simulations are in general functions of the solvent-accessible surface area of the atoms of the peptide or protein. We have previously shown that there is no correlation between the solvent-accessible surface area and the proximity to the left-handed 3_1-helical fold for the β-heptapeptide studied here [20]. Certainly, solvent effects cannot be reduced simply to a question of solvent-accessible surface areas. It seems, therefore, mandatory to explicitly treat solvent degrees of freedom when realistic simulation of peptide folding is to be achieved.

References

[1] C.M. Dobson, A. Šali, M. Karplus, Angew. Chem. Int. Ed. 37 (1998) 868.
[2] A.G. Ladurner, L.S. Itzhaki, V. Daggett, A.R. Fersht, Proc. Natl. Acad. Sci. USA 95 (1998) 8473.
[3] M. Karplus, E. Shakhnovich, in: Protein Folding, T.E. Creighton, ed. (Freeman, New York, 1992) p. 127.
[4] W.F. van Gunsteren, F.J. Luque, D. Timms, A.E. Torda, Annu. Rev. Biophys. Biomol. Struct. 23 (1994) 847.
[5] R. Czerminski, R. Elber, J. Chem. Phys. 92 (1990) 5580.
[6] Y. Ueda, H. Taketomi, N. Gō, Biopolymers 17 (1978) 1531.
[7] K.F. Lau, K.A. Dill, Macromolecules 22 (1989) 3986.
[8] J. Skolnick, A. Kolinski, J. Mol. Biol. 221 (1991) 499.
[9] A. Šali, E. Shakhnovich, M. Karplus, J. Mol. Biol. 235 (1994) 1614.
[10] J.D. Honeycutt, D. Thirumalai, Biopolymers 32 (1992) 695.
[11] H. Nymeyer, A.E. García, J.N. Onuchic, Proc. Natl. Acad. Sci. USA 95 (1998) 5921.
[12] P. Derreumaux, J. Chem. Phys. 109 (1998) 1567.
[13] W.F. van Gunsteren, H.J.C. Berendsen, J.A.C. Rullmann, Mol. Phys. 44 (1981) 69.
[14] S. Yun-Yu, W. Lu, W.F. van Gunsteren, Mol. Simul. 1 (1988) 369.
[15] F. Fraternali, W.F. van Gunsteren, J. Mol. Biol. 256 (1996) 939.
[16] W. Ooi, M. Oobatake, G. Némethy, H.A. Scheraga, Proc. Natl. Acad. Sci. USA 84 (1987) 3086.
[17] W.C. Still, A. Tempczyk, R.C. Hawley, T. Hendrickson, J. Am. Chem. Soc. 112 (1990) 6127.
[18] P.F.W. Stouten, C. Frömmel, H. Nakamura, C. Sander, Mol. Simul. 10 (1993) 97.
[19] S.-S. Sung, X.-W. Wu, Proteins: Struct. Funct. & Genet. 25 (1996) 202.
[20] X. Daura, B. Jaun, D. Seebach, W.F. van Gunsteren, A.E. Mark, J. Mol. Biol. 280 (1998) 925.
[21] X. Daura, K. Gademann, B. Jaun, D. Seebach, W.F. van Gunsteren, A.E. Mark, Angew. Chem. Int. Ed. 38 (1999) 236.
[22] F.B. Sheinerman, C.L. Brooks III, J. Mol. Biol. 278 (1998) 439.
[23] W.F. van Gunsteren, S.R. Billeter, A.A. Eising, P.H. Hünenberger, P. Krüger, A.E. Mark, W.R.P. Scott, I.G. Tironi, Biomolecular Simulation: The GROMOS96 Manual and User Guide (Vdf Hochschulverlag AG an der ETH Zürich, Zürich, 1996).

[24] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. DiNola, J.R. Haak, J. Chem. Phys. 81 (1984) 3684.
[25] X. Daura, W.F. van Gunsteren, A.E. Mark, Proteins: Struct. Funct. & Genet. 34 (1999) 269.
[26] D. Seebach, P.E. Ciceri, M. Overhand, B. Jaun, D. Rigo, L. Oberer, U. Hommel, R. Amstutz, H. Widmer, Helv. Chim. Acta 79 (1996) 2043.