FEBS Letters 457 (1999) 414–418 FEBS 22524

Long-range electrostatic effects on peptide folding

Johan Åqvist*

Department of Cell and Molecular Biology, Uppsala University, Biomedical Center, Box 596, S-751 24 Uppsala, Sweden

Received 25 June 1999

Abstract The reversible folding/unfolding of a short peptide in solution is studied by molecular dynamics simulations. The effects of long-range electrostatic interactions are examined and found to be important both for the equilibrium between folded and unfolded states and the dynamics of the folding process. The neglect of long-range electrostatics leads to an increased population of unfolded states and increased structural fluctuations. When such interactions are taken into account, the peptide unfolds and folds to the experimentally determined structure several times during a 25 ns trajectory, with approximately equal populations of folded and unfolded states in the neighborhood of its proposed melting temperature. The effect of using spherical boundary conditions rather than periodic ones does not appear to have any major effect on the folding dynamics.

© 1999 Federation of European Biochemical Societies.

Key words: Peptide folding; Molecular dynamics; Electrostatic effect

1. Introduction

The possibility to attack the protein folding problem by direct 'brute force' computer simulations on the atomic scale has not yet become a reality. However, the increased computer power has started to make some of the relevant problems in protein folding amenable to simulation methods such as molecular dynamics (MD) and Monte Carlo (MC) calculations. While MC methods have been primarily used in the context of exploring simplified lattice or grid models of protein structure, MD simulation has been employed to gain insight into folding processes of real proteins [1]. The MD technique is limited by the small time-step that has to be used in integrating the equations of motion, together with the large number of interatomic interactions that have to be taken into account for a realistic description of a system containing a polypeptide in solution. Since the typical time-span that can be covered by MD simulation is on the ns or at best μs time-scale, while real protein folding rather occurs on the second time-scale, there is certainly still a long way to go. For this reason, so-called unfolding simulations instead started to attract attention in the early 90s [2–4]. The idea behind this approach is that by initiating the unfolding of a native protein structure in solution, e.g. by raising the temperature in a computer simulation, one could follow the reverse process and hopefully learn something about actual folding. The unfolding approach may, however, be criticized since there is no guarantee that thermal denaturation is similar to protein folding (or rather the reverse of it) at physiological or room temperatures with respect to, e.g., possible folding pathways, intermediates and transition states. Almost certainly at least, the structural sampling would be affected by the different entropy weight of conformations on the free energy surface at higher temperatures.

Recently, however, important steps towards more realistic simulations of protein folding have been reported [5,6]. The study by Daura et al. [5] seems particularly promising since the folding dynamics of a short β-heptapeptide was found to be essentially completely described on a 10 ns time-scale. These authors showed that, below the melting temperature, the simulated peptide in methanol spontaneously folds into the correct 3_1-helical structure determined by NMR in the same solvent [7]. By carrying out 50 ns simulations at different temperatures, an approximate melting temperature of around 340 K could also be determined in [5]. Although not, of course, a real protein, this peptide is still the largest peptidic molecule whose folded conformation has been correctly predicted by computer simulation. A number of pertinent questions for folding regarding, e.g., pathways, intermediates and correlations with various diagnostic measures of folded states (surface areas, energies, radius of gyration, etc.) could be addressed for such a simple system [5].

The other recent study by Duan and Kollman [6] addressed the folding of a 36 residue peptide, the autonomously folding headpiece subdomain of villin. This small protein domain was simulated in aqueous solution for 1 μs starting from an unfolded structure. Several features of the folding process were found to be compatible with kinetic experiments and, although the native state was not reached within 1 μs, a marginally stable structure with clear resemblance to the native one was frequently populated.

In view of the interesting and pioneering results of [5,6], it becomes important to try to assess the scope of MD simulation for studying real folding processes. Since computer time is the limiting factor, it is, e.g., highly desirable to optimize the computational model with respect to the efficiency without losing accuracy in the description of the microscopic system. Furthermore, it becomes necessary to assess how sensitive the simulated folding processes are to the exact methodology used in the calculations. In particular, one would like to be able to draw conclusions regarding stable conformations, melting temperatures, time-scales, etc. that do not depend strongly on exactly how the MD calculations are carried out. It could, e.g., be possible that different force fields, boundary conditions, interaction truncation schemes, etc. produce significantly different dynamic and average properties, in which case the information provided by folding simulations might be more uncertain than is usually believed.

In this paper, we examine two of the key issues in modelling of peptide/protein folding processes, namely the role of long-range electrostatic interactions in determining conformational

*Corresponding author. Fax: (46) (18) 536971.
E-mail: aqvist@xray.bmc.uu.se

0014-5793/99/$20.00 © 1999 Federation of European Biochemical Societies. All rights reserved.
PII: S0014-5793(99)01079-0

stability and the effect of different boundary conditions in the simulations. While the first of these problems is related to what forces stabilize the native state, the second issue is of considerable importance for computational accuracy and efficiency. We report MD simulations of the β-heptapeptide studied in [5], using exactly the same force field but otherwise a different type of simulation model to show that (1) the treatment of long-range electrostatics is critical to the simulated folding process while (2) significantly different boundary conditions have no discernible effect on the results.

2. Materials and methods

The MD simulations were carried out using the program Q [8,9] with the GROMOS96 force field [10]. The NMR structure of the β-heptapeptide [7] was used as the starting point. The peptide was centered in a 25 Å sphere containing 929 methanol molecules after deleting solvent too close to the peptide (a distance criterion between peptide non-hydrogen atoms and the methanol C and O atoms of 2.7 Å and 2.3 Å, respectively, was used). The methanols were initially placed randomly oriented on a cubic grid with a density of 0.79 g/cm^3. These were then subjected to radial surface restraints to maintain the correct density of the system during the simulations. The form of this restraining potential was

$$V_{\mathrm{rstr}}(r) = \begin{cases} k(r-r_0)^2 - D & \text{if } r > r_0 \\ D(1-e^{a(r-r_0)})^2 - D & \text{otherwise} \end{cases} \quad (1)$$

where $r_0 = 24.7$ Å is the target radius of the sphere, $k = 60$ kcal/(mol Å^2) is a 'half-harmonic' force constant providing repulsion for methanol molecules leaving the sphere. The Morse-like potential inside the sphere provides a very weak attraction towards the surface which is necessary for maintaining a correct density everywhere inside the sphere. This potential was calibrated for pure methanol, yielding the parameters $D = 0.4$ kcal/mol and $a = 0.4$. The calculations were initiated by a short energy minimization (1000 steps) followed by a rapid (5 ps) heating of the system to 340 K. After this, the simulations were carried out for 25 ns at 340 K with weak coupling ($\tau_T = 0.1$ ps) to a heat bath of this temperature according to the method in [11]. We did not use separate heat baths for the solute and solvent in controlling the temperature of the system. The reason for this is that for such a small solute as the β-heptapeptide, which contains only 64 atoms with the GROMOS96 extended atom force field [10], the temperature of the peptide itself fluctuates considerably. That is, while the temperature of the solvent typically fluctuates less than 1 K in each timestep, the corresponding fluctuations of the peptide can amount to as much as 10 K. This is simply due to that its average kinetic energy instantaneously can change more than the average kinetic energy of the much larger number of solvent atoms. Thus, separate temperature scaling of 64 atoms does not seem physically well-motivated.

To first examine the possible effect of different boundary conditions, one simulation was carried out with a similar long-range interaction truncation scheme as in [5] where, however, periodic boundary conditions (PBCs) were used. We thus employed dual cutoff radii of $R_{\mathrm{c,in}} = 8$ Å and $R_{\mathrm{c,out}} = 14$ Å, where all interactions within the inner radius around each charge group (i.e. small neutral fragments or net charged ones, see [10] for definitions) were calculated at each MD step. Between $R_{\mathrm{c,in}}$ and $R_{\mathrm{c,out}}$, a third order multipole expansion [12] of the electrostatic potential around each charge group center (in the laboratory coordinate frame) was used and this expansion was updated every 25 steps together with the regular non-bonded neighbor lists. No interactions of a longer range than $R_{\mathrm{c,out}}$ were considered. The third order multipole expansion is, in general, considerably more accurate [12] than the first order one used in [5], which is equivalent to assuming a constant long-range force on each particle during the updating interval. With the present approach, the fact that the long-range forces change as particles move away during the updating interval, from their position at the latest expansion update, is taken into account since the expansion for each charge group is fixed in space and its higher derivatives retained. However, in [5], a much shorter updating interval of five MD steps was used which makes the 'constant force approximation' quite reasonable, but at a significant increase in computational effort (since all interactions are calculated every five steps rather than every 25 as in our case).

To gauge the effects of long-range forces, a second simulation was carried out using a plain 8 Å cutoff for all interactions with updating of neighbor lists every 25 steps. This type of scheme, which gives a significant improvement in speed due to the reduced number of interactions, is often used in MD simulations and was, e.g., employed in the study of [6]. The difference between our two simulations can thus be summarized such that in the first case, long-range electrostatics are taken into account within a range of 14 Å while in the second case, no interactions are considered beyond the 8 Å range.

3. Results

Before turning to the peptide folding simulations, it may be useful to briefly examine the spherical boundary model used here for methanol. It has earlier been shown that efficient spherical boundary conditions can be designed for water which are able to reproduce density, polarization, solvation energetics and local dynamic properties such as water re-orientational correlation times [9,13–15]. Nevertheless, there is a rather widespread belief that the inevitable boundary to vacuum in any non-periodic model severely affects dynamics inside the sphere and that PBCs therefore are preferable, although more time consuming. It is thus of considerable interest to gauge the influence of boundary conditions on a some more complex dynamic process such as peptide folding. For pure methanol, it is straightforward to parametrize a boundary model that gives the correct density within a spherical system. Here, we use essentially the same approach as in [9] to obtain parameters for Eq. 1 that yield a constant density within the sphere. Fig. 1 shows the average density number in 3 Å radial shells from a 20 ps MD simulation of pure methanol. It can be seen that, while some local structure is visible in the innermost shells (due to their small volume), the density rapidly settles around the experimental value for larger radii. To ensure a correct density in the peptide simulations, the target sphere radius was reset (to 24.7 Å) after deletion of methanol molecules overlapping the peptide (see above) according to the experimental atom number densities of methanol and proteins [16].

As expected, the β-heptapeptide samples a large number of different conformations during the 25 ns MD simulations at 340 K. Fig. 2 summarizes the folding/unfolding dynamics in terms of root mean square deviation (RMSD) versus time for

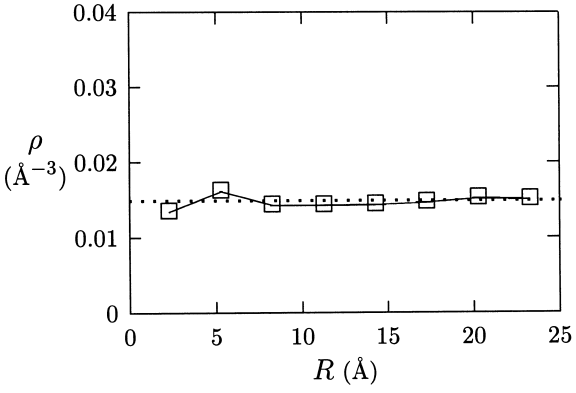

Fig. 1. Average molecular number density versus radius (in spherical shells of 3 Å resolution) calculated from a MD simulation of pure methanol in a 25 Å sphere with the boundary conditions described in the text. The experimental density number of methanol is indicated with a dotted line.

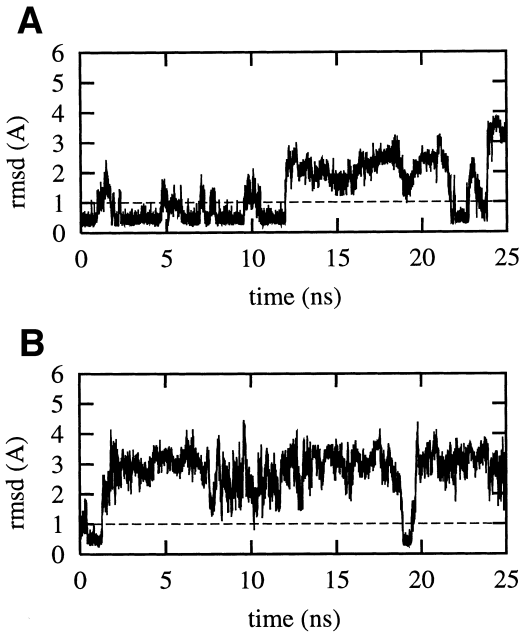

Fig. 2. Time evolution of the RMSD for the central backbone atoms (residues 2–6) with respect to the experimental structure [7] after least squares fitting of these atoms. (A) Simulation including long-range electrostatics and (B) simulation with no interactions beyond 8 Å.

the backbone atoms of residues 2–6 (after least squares fitting these atoms) between the MD structures and the experimental NMR structure [7]. As noted by Daura et al. [5], configurations with this RMSD below 1 Å are unambiguously folded while those above it represent more or less unfolded states.

In the simulation including long-range electrostatics (Fig. 2A), the peptide can be seen to unfold and refold several times during the trajectory. The first unfolding event occurs after about 1.5 ns and the corresponding structure is shown in Fig. 3A. This conformation is representative of several unfolded structures with the RMSD near 2 Å in that it retains the central native H-bond between NH(3) and CO(5). After about five rather rapid unfolding/folding events in the first 12 ns, during which period the folded state is mostly populated, the system enters a longer unfolded phase of around 10 ns. A snapshot of the unfolded conformation at 18.4 ns is shown in Fig. 3B, where it can be seen that, although the structure is fairly compact, none of the H-bonds between the central residues 2–6 are formed. The system then returns to the folded state and Fig. 3C shows an example of a typically folded structure (at 21.8 ns) superimposed on the NMR conformation.

Overall, the behavior of the simulation with long-range electrostatics turned on is clearly similar to that observed by Daura et al. [5]. That is, one observes both rapid (partial) unfolding events with life-times on the ns or sub-ns time-scale and longer unfolding periods on roughly the 10 ns time-scale. The more rapid events are characterized by RMSDs in the order of 1–2 Å while longer unfolding periods typically have RMSDs above 2 Å. Although our statistics are a bit limited for assessing whether 340 K corresponds approximately to the melting temperature of the peptide, as concluded in [5], the data at least appear to be compatible with [5]. Hence, there is no evidence for a significant effect of the different boundary conditions on the folding dynamics.

The situation is quite different for the 'cutoff simulation' excluding long-range electrostatics. The time evolution of the RMSD is shown in Fig. 2B, from which it is immediately clear that the equilibrium between folded and unfolded states has been drastically shifted towards the latter due to the neglect of long-range forces. Thus, the peptide spends most of

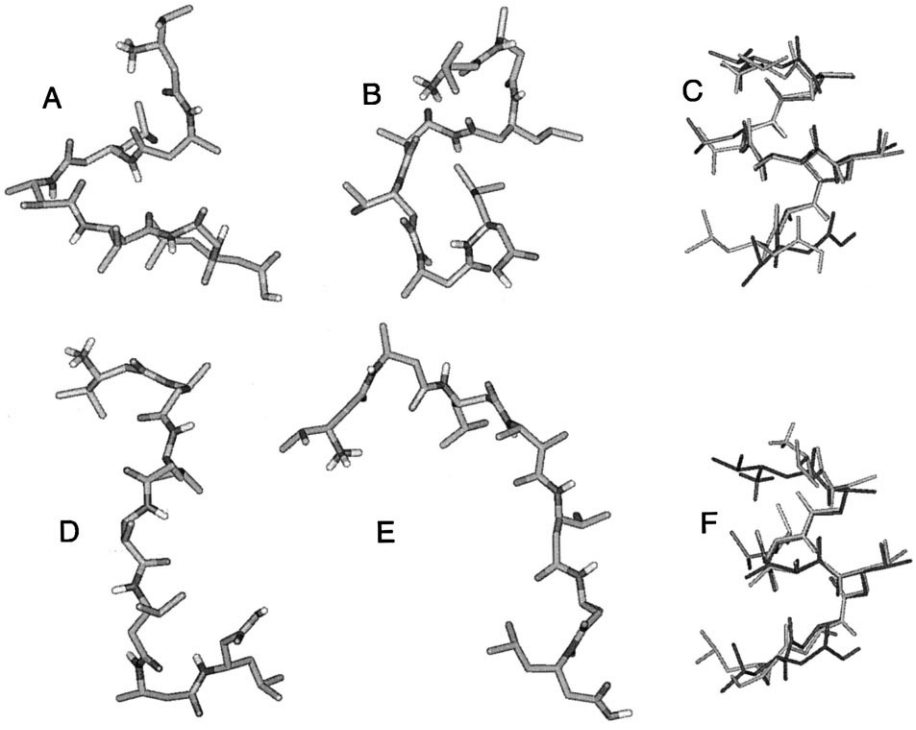

Fig. 3. MD snapshots from the simulations with (A–C) and without (D–F) long-range electrostatics. A, B, D and E show typical unfolded conformations while C and F depict folded MD structures (light lines) superimposed on the experimental one (dark lines).

the time in unfolded conformations and shows only two brief visits to the folded state (at around 1.5 and 19 ns). The unfolded conformations sampled in the cutoff simulation are also generally characterized by a larger radius of gyration and larger RMSD values with respect to the NMR structure. This turns out to be related to the fact that the unfolded conformations tend to be considerably more extended in this case. Fig. 3D and E shows two examples (at 9.6 and 19.8 ns) of typical unfolded states where no internal H-bonds are present and the structures are essentially of an extended nature. It can also be noted from Fig. 2B that the fluctuations of RMSD along the trajectory are significantly higher than for the simulation including long-range electrostatics. This behavior is presumably related to the noise introduced by only including short-range forces within 8 Å, where atoms entering and leaving the cutoff sphere of a given group (and generating noise) now will be much closer to the group. The fact that the unfolded structures tend to be more extended may also in itself, of course, cause larger RMSD fluctuations.

Even though the unfolded state appears to be considerably more favored in the cutoff simulation, the system manages to refold at around 19 ns. A corresponding folded structure from this period, superimposed on the NMR conformation, is shown in Fig. 3F. It can be seen that the agreement with the experimental structure is similar to that of Fig. 3C and the 3_1-helix is clearly revisited with the central part of the backbone (2–6) in perfect agreement with the 'correct' conformation.

4. Discussion

We have carried out two relatively long MD simulations of the β-heptapeptide whose solution structure in methanol has been determined by NMR [7]. Our main objective with this study is to investigate the influence of long-range electrostatic forces in the peptide folding process and to examine the effect of using spherical boundary conditions rather than PBCs, which may speed up this type of 'long-time' dynamics calculations considerably.

As far as boundary conditions are concerned, there appears to be no significant effect on the peptide folding dynamics resulting from using a spherical rather than periodic model, as long as the former is calibrated to give the correct density of the system. Although we used a somewhat different treatment of long-range electrostatics, with a higher multipole expansion but less frequent updating of it, and did not include van der Waals interactions beyond 8 Å, our results seem to be fully compatible with those of Daura et al. [5] for the same system. The neglect of van der Waals forces beyond 8 Å is not expected to cause large effects since these are of a relatively short range and, for larger distances, approximately spherically symmetric. Irrespective of what type of boundary conditions is used, it is, of course, necessary to use a sufficiently large system. The β-heptapeptide of Seebach and coworkers [7] in methanol is a useful system for studying simple folding dynamics since only around 3000 particles need to be treated (our calculations typically take 15 h/ns on a DEC Alpha 433MHz processor).

While the peptide is seen to unfold and refold several times during the simulation including long-range electrostatics, it is clear that truncation of these interactions beyond 8 Å causes significant effects on both the equilibrium and dynamic properties of the solvated peptide. The equilibrium is largely shifted towards the unfolded state, where also typical structures are more extended than when long-range electrostatics are taken into account. The fluctuations of the RMSD with respect to the experimental reference conformation are also considerably higher for the cutoff simulation. In view of these results, it would appear highly undesirable to use such short cutoffs in studies of peptide and protein dynamics. It has also been shown earlier that the energetics of ionic solvation processes are drastically affected by such truncation schemes [14,17].

An important result of this study is then that long-range interactions (>8 Å) seem to be crucial to the folding process. This would not be unexpected for the formation of β-sheet structures where the characteristic interactions are intrinsically of long-range in terms of amino acid sequence, i.e. residues that are non-adjacent in the sequence need to come within H-bonding distance of each other. It might appear more surprising that helix formation, in this case, the β-peptide 3_1-helix with H-bonds between NH(i) and CO(i+2), is also dependent on longer range interactions. One reason for this is probably that with an 8 Å cutoff, the polar backbone groups of the peptide (NH and CO) only 'see' their polar counterparts in adjacent residues if the peptide is in an extended conformation. Hence, if backbone interactions between, say, i and i+2 (such as the experimentally observed backbone H-bonds) constitute a driving force for helix formation from an extended structure, this driving force would not be present with an 8 Å cutoff. Another effect which may be of importance as soon as the peptide contains ionized groups is that truncation of long-range interactions between solvent molecules increases their affinity for charged groups [14,17]. That is, intermolecular interactions between the solvent molecules are reduced and, consequently, also their resistance to polarization by charged groups. This leads to an overpolarization of the solvent around charges and to a too negative free energy of solvation of the charge [14,17]. It has, e.g., been observed that solvent accessible ion-pairs in proteins tend to break in MD simulations with short cutoffs, instead exposing separated charges to the solvent [18]. Since the β-heptapeptide has one ionized group (NH_3) at its N-terminus, its tendency to form the experimentally observed charge dipole type H-bond with CO(3) will inevitably be reduced by truncation of solvent-solvent interactions beyond 8 Å.

While MD computer simulations are beginning to emerge as a promising tool for studying real peptide/protein folding processes, a lesson to be learned from the present work is that both equilibrium and dynamic properties may depend significantly on the simulation procedure employed. It is not either trivial to guess the influence of different factors, such as simulation parameters, force fields, boundary conditions, etc., without actually carrying out several simulations and comparing them. In this respect, the fact that this work and that of [5] both yield reversible folding to the experimentally determined stable structure, using the same force field but different boundary conditions, is rather encouraging.

Acknowledgements: Support from the TFR and NFR is gratefully acknowledged.

References

[1] Karplus, M. and Sali, A. (1995) Curr. Opin. Struct. Biol. 5, 58–73.
[2] Mark, A.E. and van Gunsteren, W.F. (1992) Biochemistry 31, 7745–7748.
[3] Daggett, V. and Levitt, M. (1992) Proc. Natl. Acad. Sci. USA 89, 5142–5146.
[4] Caflisch, A. and Karplus, M. (1993) Proc. Natl. Acad. Sci. USA 91, 1746–1750.
[5] Daura, X., Jaun, B., Seebach, D., van Gunsteren, W.F. and Mark, A.E. (1998) J. Mol. Biol. 280, 925–932.
[6] Duan, Y. and Kollman, P.A. (1998) Science 282, 740–744.
[7] Seebach, D., Ciceri, P.E., Overhand, M., Jaun, B., Rigo, D., Oberer, L., Hommel, U., Amstutz, R. and Widmer, H. (1996) Helv. Chim. Acta 79, 2043–2066.
[8] Åqvist, J. (1998) Q Version 2.4, Uppsala University.
[9] Marelius, J., Kolmodin, K., Feierberg, I. and Åqvist, J. (1999) J. Mol. Graph. Model. (in press).
[10] van Gunsteren, W.F., Billeter, S.R., Eising, A.A., Hünenberge, P.H., Krüger, P., Mark, A.E., Scott, W.R.P. and Tironi, I.G. (1996) Biomolecular Simulation: The GROMOS96 Manual and user Guide, vdf Hochschulverlag AG, ETH, Zürich.
[11] Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A. and Haak, J.R. (1984) J. Chem. Phys. 81, 3684–3690.
[12] Lee, F.S. and Warshel, A. (1992) J. Chem. Phys. 97, 3100–3107.
[13] King, G. and Warshel, A. (1989) J. Chem. Phys. 91, 3647–3661.
[14] Åqvist, J. (1994) J. Phys. Chem. 98, 8253–8255.
[15] Essex, J.W. and Jorgensen, W.L. (1995) J. Comput. Chem. 16, 951–972.
[16] Creighton, T.E. (1983) Proteins, W.H. Freeman, New York.
[17] Åqvist, J. (1996) J. Comput. Chem. 17, 1587–1597.
[18] Hansson, T. and Åqvist, J. (1995) Prot. Eng. 8, 1137–1144.