



Short communication

# As good as it gets? Folding molecular dynamics simulations of the LytA choline-binding peptide result to an exceptionally accurate model of the peptide structure

Ilias Patmanidis, Nicholas M. Glykos\*

Department of Molecular Biology and Genetics, Democritus University of Thrace, University Campus, 68100 Alexandroupolis, Greece

## ARTICLE INFO

## Article history:

Received 8 January 2013

Received in revised form 8 February 2013

Accepted 9 February 2013

Available online 17 February 2013

## ABSTRACT

Folding simulations of a choline-binding peptide derived from the *Streptococcus pneumoniae* LytA protein converged to a model of the peptide's folded state structure which is in outstanding agreement with the experimentally-determined structures, reaching values for the root mean squared deviation as low as 0.24 Å for the peptide's backbone atoms and 0.65 Å for all non-hydrogen atoms.

© 2013 Elsevier Inc. All rights reserved.

## Keywords:

Molecular dynamics simulations

Peptide folding

Choline binding repeat

LytA protein

Empirical force fields

AMBER force field

## 1. Introduction

The AMBER99SB force field [1,2] and its various variants [3] appear to be some of the most promising and successful non-polarizable empirical force fields currently available for folding simulations of proteins [4,5] and peptides [1,2,4–7]. The emphasis in the majority of these simulation studies was placed on whether folding to the native structure was actually observed and not on the accuracy of the structure as whole (i.e. including the conformation of the side chains). Here we report results from the application of the AMBER99SB-ILDN force field to the folding (in explicit solvent and with full PME-based electrostatics) of a 14mer peptide derived from the first choline-binding repeat of the LytA protein [8]. We show that not only folding is achieved under the conditions of the simulation, but also that the agreement with the experimental structure is exceptionally accurate down to the level of individual side chains, giving for all non-hydrogen atoms an RMS deviation from the experimental X-ray structure of less than 1 Å.

## 2. Methods

The peptide whose folding we simulated is the one studied by the Sanz group [8] and corresponds to residues 197–210 of the

native LytA protein with an N → D mutation at position 203. The peptide has been shown [8] to be soluble and monomeric in aqueous solutions, its structure has been determined with NMR (and found to be very similar to the X-ray structure), and it has been shown to retain its choline-binding specificity.

## 2.1. System preparation

The starting peptide structure was in the fully extended state as obtained from the program ribosome (<http://www.roselab.jhu.edu/~raj/Manuals/ribosome.html>). Addition of missing hydrogen atoms and solvation-ionization was performed with the program LEAP from the AMBER tools distribution [9]. The peptide termini were unprotected in agreement with the experimental conditions previously reported [8]. For all simulations we used periodic boundary conditions and a cubic unit cell sufficiently large to guarantee a minimum separation between the PBC-related images of the peptide of at least 16 Å. We followed the dynamics of several folding simulations using the program NAMD [10] for a grant total of 13.7 μs and for several different combinations of temperatures (320 K, 340 K, 360 K), water models (TIP3P and TIP4P-Ew) and force fields (AMBER99SB and AMBER99SB-ILDN). Here we only report results from a 1.4 μs-long trajectory at 360 K with the TIP3P water model and the AMBER-ILDN force field. This is the only one of our trajectories for which folding to the native β-hairpin structure has been observed. The fact that we only observed a single folding event with the AMBER-ILDN-TIP3P combination should not be

\* Corresponding author. Tel.: +30 25510 30620; fax: +30 25510 30620.

E-mail addresses: [glykos@mbg.duth.gr](mailto:glykos@mbg.duth.gr), [nmglykos@gmail.com](mailto:nmglykos@gmail.com) (N.M. Glykos).

considered as an indication that the other force fields cannot fold the peptide: as also discussed in Section 3, the LytA peptide is such a slow folder (possibly due to the presence of significant energetic frustration) that our simulations – even after many microseconds – are nowhere near to sufficiently sample the whole of its folding landscape.

## 2.2. Simulation protocol

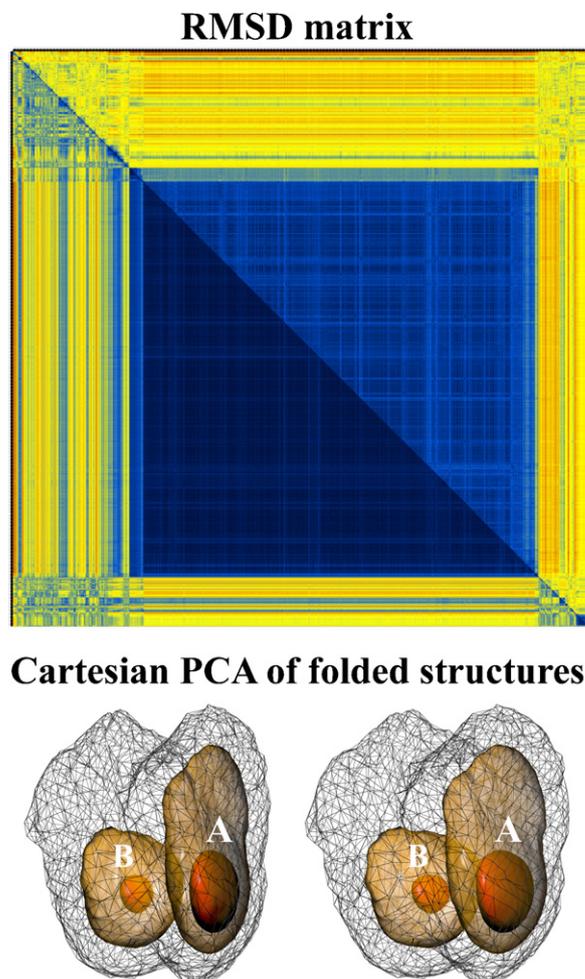
The system was first energy minimized for 1000 conjugate gradient steps followed by a slow heating-up phase to the final temperature of 360 K (with a temperature step of 20 K) over a period of 32 ps. Subsequently the system was equilibrated for 10 ps under NpT conditions without any restraints, until the volume equilibrated. This was followed by the production NpT run with the temperature and pressure controlled using the Nosè-Hoover Langevin dynamics and Langevin piston barostat control methods as implemented by the NAMD program (and maintained at a temperature of 360 K and a pressure of 1 atm). The Langevin damping coefficient was set to  $1 \text{ ps}^{-1}$ , and the piston's oscillation period to 200 fs, with a decay time of 100 fs. The production run was performed with the impulse Verlet-I multiple timestep integration algorithm as implemented by NAMD. The inner timestep was 2 fs, short-range non-bonded interactions were calculated every one step, and long-range electrostatics interactions every two timesteps using the particle mesh Ewald method with a grid spacing of approximately 1 Å and a tolerance of  $10^{-6}$ . A cutoff for the van der Waals interactions was applied at 8 Å through a switching function, and SHAKE (with a tolerance of  $10^{-8}$ ) was used to restrain all bonds involving hydrogen atoms. Trajectories were obtained by saving the atomic coordinates of the whole system every 0.8 ps.

## 2.3. Trajectory analysis

The program CARMA [11] was used for most of the analyses, including removal of overall rotations/translations, calculation of RMSDs from a chosen reference structure, calculation of the radius of gyration, calculation of the average structure (and of the atomic root mean squared fluctuations), production of PDB files from the trajectory, Cartesian space principal component analysis and corresponding cluster analysis, dihedral space principal component analysis and cluster analysis, calculation of the frame-to-frame RMSD matrices, etc. Secondary structure assignments were calculated with the programs STRIDE [12]. All molecular graphics work and figure preparation were performed with the programs VMD [13] and CARMA.

## 3. Results

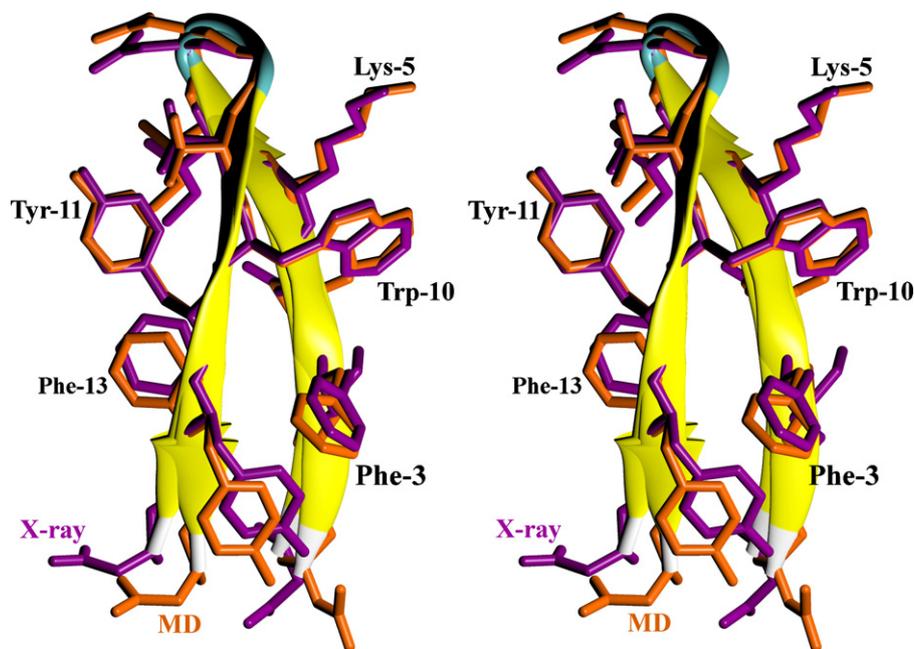
The upper panel of Fig. 1 shows a color representation of the RMS deviations between all pairs of structures recorded from the molecular dynamics trajectory. The yellow–red areas of the diagram indicate high RMSD values and correspond to unfolded, fast-converting peptide conformations. The prominent dark blue square (extending from  $\sim 0.3 \mu\text{s}$  to  $1.3 \mu\text{s}$  of our trajectory) is the folding event and comprises stably and correctly folded  $\beta$ -hairpin structures. An unbiased selection of a representative structure for the folded state of the peptide was performed as follows. At the first stage all folded structures as judged by both the RMSD matrix (Fig. 1, upper panel) and dihedral PCA-based cluster analysis were selected. These  $\sim 1.1$  million structures corresponded to 61% of the total number of structures recorded. The folded structures were subsequently analysed using Cartesian PCA-based cluster analysis as implemented by CARMA [11]. This analysis (see Fig. 1, lower panel) showed that in the folded state the peptide inter-converts between two conformations, with the major one



**Fig. 1.** RMSD matrix and principal component analysis. The upper panel is a color representation of the trajectory's RMSD matrix with the horizontal and vertical axis corresponding to simulation time and ranging from 0 to  $1.45 \mu\text{s}$ . The lower half of the matrix has been calculated using the  $\text{C}\alpha$  atoms only, the upper half using all non-hydrogen atoms. The origin is at the top-left-hand corner and the linear color scale ranges from dark blue (corresponding to an RMSD of zero), through yellow (for intermediate RMSDs), to dark red (large RMSDs). The lower panel shows a wall-eyed stereodiagram of the projection of the trajectory's folded structures on the space defined by the top three eigenvectors as obtained from Cartesian principal component analysis. Three isosurfaces are shown at increasingly higher density of the corresponding distribution. The major (A) and minor (B) molecular conformations are marked. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

accounting for  $\sim 75\%$  of the folded population. Representative structures for the two conformations were identified by calculating an average structure for each cluster and then selecting the frame from the trajectory with the lowest RMS deviation from the corresponding average structure. The structures of these two conformers are very similar having an RMSD over all backbone atoms of only  $0.46 \text{ \AA}$ . Their only significant difference is the rotation of the side chain of Tyr-11 by 60 degrees about the  $\chi_1$  angle. If Tyr-11 is excluded from the calculation, the RMS deviation between the two conformers using all non-hydrogen atoms is  $1.0 \text{ \AA}$ . Note that the selection of the molecular-dynamics-derived structures as outlined above is totally agnostic with respect to the experimentally determined structure, i.e. the experimentally known peptide structure has not in any way been used for the selection of the representative molecular dynamics structures.

The stereodiagram shown in Fig. 2 is a direct all-atom comparison between the peptide's representative molecular dynamics structure (calculated as described above) and its X-ray



**Fig. 2.** Experiment vs. simulation: structural comparison. This wall-eyed stereodiagram compares the representative simulation-derived structure (orange) with the X-ray crystallographic structure (magenta). A cartoon representation of the backbone trace (colored according to its secondary structure assignment with STRIDE) is also shown to aid interpretation. The choline-binding site is located between the side-chains of Phe-3 and Trp-10 (residues 199 and 206 in the *LytA* numbering).

crystallographic structure (PDB entry 1gvm). The agreement between the two structures is exceptionally accurate even at the level of individual side chains: The RMS deviation (excluding the first and last residues) is 0.54 Å for all backbone atoms and 1.10 Å for all heavy atoms. If residues 3–12 are used for the calculation, the RMSD's become 0.37 and 0.99 Å only. To our knowledge, the structural agreement shown in Fig. 2 is the best reported so far from a peptide folding simulation. This is more so given that the molecular dynamics structure shown in this figure was based on an unbiased selection (see previous paragraph) and is *not* the structure that best agrees with the experimental structure as discussed in the next paragraph.

A direct RMSD-based comparison between the simulation and the experimental structures shows that the simulation visits conformations that are in even better agreement with the X-ray and NMR determinations than the one shown in Fig. 2. To put this in numbers, Table 1 records the lowest RMS deviations observed by comparing all of the trajectory-derived structures with the experimental (NMR and X-ray) structures. The RMSD values quoted in this table (at the ~0.5 Å range) are not very different from the crystallographic positional errors expected on the basis of Luzzati plots.

A comparison between the NMR and X-ray entries in Table 1 shows that the simulation appears to be in better agreement with the X-ray-derived structure than with the NMR structure. We connect this finding with the computational indication that the peptide is a slow folder (we have observed only one folding event in more than 13 μs of simulation time), and with the observation that the peptide is visiting very many transiently stable but non-native

**Table 1**  
Lowest RMSDs (in Å) observed between the molecular dynamics structures and the peptide's experimental (X-ray and NMR) structures with and without the termini.

	RMSD, C $\alpha$ atoms	RMSD, all heavy
X-ray, residues 1–14	0.25	0.91
X-ray, residues 2–13	0.17	0.65
NMR, residues 1–14	0.53	1.59
NMR, residues 2–13	0.43	1.45

conformations (data not shown). A possible interpretation of these results may be based on the presence of significant frustration in the energy landscape for this peptide which makes difficult the consistent interpretation of the NMR data in terms of a single family of native  $\beta$ -hairpin-like structures.

We should close this section by noting that the excellent agreement in terms of the three dimensional structure of the peptide's native state offers no evidence whatsoever concerning the ability of the simulation to correctly (or otherwise) predict the peptide's folding dynamics. Indeed, with only one folding event recorded it is not even possible to confidently characterize the peptide's simulation dynamics, so much so to meaningfully compare them with the experimental data.

#### 4. Discussion

The take-home message of this communication is not just based on Fig. 2, it is Fig. 2: for the given peptide, simulation protocol and force field, the folded structure derived from the simulation is not just validated through the comparison with the experimentally determined structures of the peptide's native state, it is practically indistinguishable from them. Although no attempt will be made to generalize our results, we shall not resist the temptation of noting our enthusiasm for the accuracy with which pure physics-based methods and the newest generation of force fields can predict the structures of complex biological molecules.

#### Acknowledgements

We should like to thank Prof. Jesus Sanz for providing the peptide's coordinates and for useful discussions.

#### References

- [1] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, *Proteins* 65 (2006) 712–725.
- [2] L. Wickstrom, A. Okur, C. Simmerling, *Biophysical Journal* 97 (2009) 853–856.
- [3] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J.L. Klepeis, R.O. Dror, D.E. Shaw, *Proteins* 78 (2010) 1950–1958.

- [4] D.E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R.O. Dror, M.P. Eastwood, J.A. Bank, J.M. Jumper, J.K. Salmon, Y. Shan, W. Wriggers, *Science* 330 (2010) 341–346.
- [5] C. Zhang, J. Ma, *Journal of Chemical Physics* 132 (2010) 244101.
- [6] P.S. Georgoulia, N.M. Glykos, *Journal of Physical Chemistry B* 115 (2011) 15221–15227.
- [7] K.K. Patapati, N.M. Glykos, *Biophysical Journal* 101 (2011) 1766–1771.
- [8] B. Maestro, C.M. Santiveri, M.A. Jiménez, J.M. Sanz, *Protein Engineering, Design & Selection* 24 (2011) 113–122.
- [9] D.A. Case, T.E. Cheatham 3rd, T. Darden, H. Gohlke, R. Luo, K.M. Merz Jr., A. Onufriev, C. Simmerling, B. Wang, R.J. Woods, *Journal of Computational Chemistry* 26 (2005) 1668–1688.
- [10] L. Kale, R. Skeel, M. Bhandarkar, R. Brunner, A. Gursoy, N. Krawetz, J. Phillips, A. Shinozaki, K. Varadarajan, K. Schulten, *Journal of Computational Physics* 151 (1999) 283–312.
- [11] N.M. Glykos, *Journal of Computational Chemistry* 27 (2006) 1765–1768.
- [12] D. Frishman, P. Argos, *Proteins* 23 (1995) 566–579.
- [13] W. Humphrey, A. Dalke, K. Schulten, *Journal of Molecular Graphics* 14 (1996) 33–38.