

# The Impact of Interchain Hydrogen Bonding on $\beta$ -Hairpin Stability is Readily Predicted by Molecular Dynamics Simulation

Stephan Niebling, Emma Danelius, Ulrika Brath, Sebastian Westenhoff, Máté Erdélyi  
Department of Chemistry and Molecular Biology, University of Gothenburg, Gothenburg, Sweden

Received 16 February 2015; revised 7 April 2015; accepted 4 May 2015

Published online 13 May 2015 in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/bip.22671

## ABSTRACT:

Peptides are frequently used model systems for protein folding. They are also gaining increased importance as therapeutics. Here, the ability of molecular dynamics (MD) simulation for describing the structure and dynamics of  $\beta$ -hairpin peptides was investigated, with special attention given to the impact of a single inter-strand sidechain to sidechain interaction. The MD trajectories were compared to structural information gained from solution NMR. By assigning frames from restraint-free MD simulations to an intuitive hydrogen bond on/off pattern, folding ratios and folding pathways were predicted. The computed molecular model successfully reproduces the folding ratios determined by NMR, indicating that MD simulation may be straightforwardly used as a screening tool in  $\beta$ -hairpin design. © The Authors. *Biopolymers* Published by Wiley Periodicals, Inc. *Biopolymers (Pept Sci)* 104: 703–706, 2015.

**Keywords:** beta-hairpin; hydrogen bond; molecular dynamics

This article was originally published online as an accepted preprint. The “Published Online” date corresponds to the preprint version. You can request a copy of any preprints from the

Additional Supporting Information may be found in the online version of this article

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Correspondence to: Máté Erdélyi; Department of Chemistry and Molecular Biology, University of Gothenburg, Kemivägen 10, 412 96 Gothenburg, Sweden; e-mail: mate@chem.gu.se

Contract grant sponsor: European Union Seventh Framework Programme  
Contract grant number: FP7/2007-2013 Under Grant Agreement No. 259638  
Contract grant sponsor: Swedish Research Council  
Contract grant number: 2012-3819

© 2015 The Authors. *Biopolymers* Published by Wiley Periodicals, Inc.

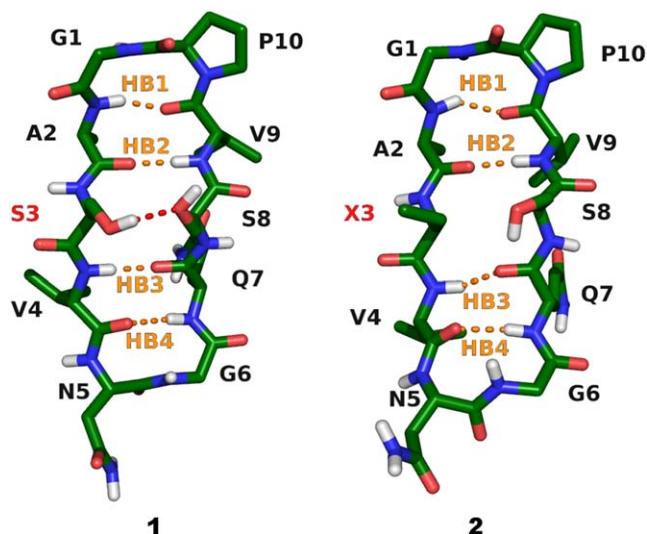
past two calendar years by emailing the *Biopolymers* editorial office at [biopolymers@wiley.com](mailto:biopolymers@wiley.com).

## INTRODUCTION

The significance of peptides as therapeutics is rapidly growing. Over one hundred peptidic drugs are marketed for the treatment of diseases such as allergy, asthma, diabetes, inflammation, cancer, as well as cardiovascular and infective illnesses,<sup>1,2</sup> whereas about 140 peptide candidates are currently in clinical and another 500–600 in preclinical development.<sup>3</sup> Most of them are small, i.e. are composed of 8-to-10 amino acids.<sup>1</sup> The present approval rate of peptide drugs is twice as compared to that of small molecules. Owing to their rapidly growing importance, the understanding and prediction of the behavior of peptides is receiving increased attention.<sup>4,5</sup> Peptide foldamers, such as  $\beta$ -hairpins, have gained further importance as suitable model systems for early stages of protein folding.<sup>6</sup> The dynamic nature of peptides makes their structural analysis a remaining challenge. The solid state structure of oligopeptides does not necessarily coincide with their conformation in solution,<sup>7</sup> and the commonly utilized NMR restraint-driven structure calculations are likely to yield averaged conformations that might be more misleading than informative.<sup>8–10</sup>

Prediction of the structure of small peptides, such as  $\beta$ -hairpins, was recently attempted by molecular dynamics (MD) simulations, with a number of successful examples.<sup>4,11–13</sup> It should be noted, however, that the outcome of these computations is commonly validated by comparison to an X-ray derived solid state, or to an NMR-derived averaged conformation, thus neglecting the inherent dynamic nature of oligopeptides that is best described by a conformational ensemble. Overall, the capability of MD to reproduce dynamic solution conformational ensembles has not yet been as convincingly demonstrated as its ability to refine proteins' structure.

Here, the ability of a conventional molecular dynamic simulation protocol to reproduce the dynamic conformational



**FIGURE 1** The sequence of cyclic  $\beta$ -hairpins **1** and **2** is identical apart from an OH to CH<sub>3</sub> substitution at the sidechain of the amino acid at position 3. Whereas the folded  $\beta$ -hairpin conformation of **1** is stabilized by interstrand sidechain to sidechain hydrogen bond of S3 to S8, this interaction is prevented for **2** due to substitution of Ser-3 to aminobutyric acid (X3). The peptide backbones are shown with carbons in green, nitrogens in blue, oxygens in red, and hydrogens in white. Aliphatic hydrogens are omitted for clarity. Hydrogen bonds are abbreviated as HB1: O<sub>V9</sub>–H<sub>A2</sub>, HB2: O<sub>A2</sub>–H<sub>V9</sub>, HB3: O<sub>Q7</sub>–H<sub>V4</sub>, and HB4: O<sub>V4</sub>–H<sub>Q7</sub> corresponding to the classification used in Table I.

ensemble of a pair of closely related  $\beta$ -hairpins (Figure 1) was examined. This model system allowed evaluation of the ability of the computational technique to describe the influence of a specific interstrand sidechain to sidechain interaction on  $\beta$ -hairpin folding. Our computational output was validated against the solution ensemble that was deduced by NAMFIS (NMR Analysis of Molecular Flexibility in Solution)<sup>8</sup> from the peptides' solution NMR data.<sup>14</sup>

## METHODS

MD simulations were performed with GROMACS 4.5.5<sup>15,16</sup> using the OPLS-AA force field,<sup>17,18</sup> with parameters for the non-natural amino acid ABU derived from parameters of leucine C <sub>$\beta$</sub>  and isoleucine C <sub>$\gamma$</sub> . Parameters for the solvent dimethylsulfoxide (DMSO) were used as implemented in Gromacs 4.5.5. Initial coordinates for **1** and **2** (Figure 1) were taken from an output structure of the NAMFIS analysis of their solution NMR data.<sup>14</sup>

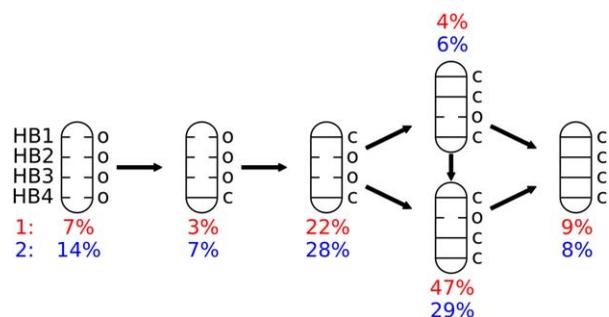
The peptides were solvated in a cubic box with periodic boundary conditions and a side length of  $\sim 36$  Å (10 Å initial minimum distance of solute to all boundaries) comprising the peptide and  $\sim 400$  DMSO molecules. For both peptides, the same molecular dynamics protocol was used. After a steepest descent energy minimization (convergence criteria 500,000 steps or maximum force  $< 10$  kJ mol<sup>-1</sup> nm<sup>-1</sup>) two 100 ps equilibration MD runs were performed. The first one in the

**Table I** The Probabilities of Conformations, their Individual and Averaged Hydrogen Bond Distances (HB1–4), and the Classification of the Overall Folding of the Most Prevalent Conformations of **1** and **2**

H-Bonds	(%)	Average Distances/Å				Average	Folded?
		HB1	HB2	HB3	HB4		
<b>Peptide 1</b>							
<i>oooo</i>	7	5.87	9.67	8.06	5.00	7.15	u
<i>ooc</i>	3	4.14	5.56	4.40	2.30	4.10	u
<i>cooc</i>	22	2.29	3.77	3.82	2.18	3.01	u
<i>cocc</i>	47	2.40	3.70	2.17	2.13	2.60	f
<i>ccoc</i>	4	2.21	2.50	3.70	2.19	2.65	f
<i>cccc</i>	9	2.25	2.42	2.18	2.16	2.25	f
<b>Peptide 2</b>							
<i>oooo</i>	14	5.92	8.87	6.95	4.09	6.46	u
<i>ooc</i>	7	5.31	7.60	5.14	2.45	5.13	u
<i>cooc</i>	28	2.26	3.72	3.83	2.19	3.00	u
<i>cocc</i>	29	2.37	3.59	2.26	2.13	2.59	f
<i>ccoc</i>	6	2.17	2.49	3.74	2.23	2.66	f
<i>cccc</i>	8	2.23	2.37	2.24	2.18	2.25	f

Hydrogen bonds (Figure 1) are characterized as open (*o*) or closed (*c*) corresponding to the cutoff threshold 3 Å. Peptide **1** is predicted to possess 66%, whereas **2** 43% folded  $\beta$ -hairpin conformation, which compare well to the experimentally determined 88% and 55%, respectively.<sup>14</sup> Here u = unfolded, f = folded. The full tables with all the theoretically possible 16 population groups is shown in the Supporting Information (Table SI).

constant particle number, volume, temperature ensemble (NVT; with modified Berendsen thermostat with velocity rescaling<sup>19</sup> at 300 K and a 0.1 ps timestep; separate heat baths for peptide and solvent); the second one in the constant particle number, pressure, temperature ensemble (NPT; Parrinello–Rahman pressure coupling<sup>20,21</sup> at 1 bar with a compressibility of  $4.5 \times 10^{-5}$  bar<sup>-1</sup> and a 2 ps time constant). During both equilibration runs, a position restraint potential with a force constant of 1000 kJ mol<sup>-1</sup> nm<sup>-2</sup> was added to all peptide atoms. To generate coordinates and velocities for the following production runs, a 100 ps simulation with position restraints (same as for the equilibration runs) was used. Coordinates and velocities at every 10 ps were used for the production runs (11  $\times$  400 ns) which resulted in a total simulation time of 4.4  $\mu$ s for each peptide. For all MD simulations the leap-frog integrator was used with a time step of 2 fs. Coordinates were saved every 2 ps. The same temperature and pressure coupling schemes as applied for the equilibration runs were used for the subsequent MD simulations. All bonds to hydrogen atoms were constrained using the linear constrained solver (LINCS)<sup>22</sup> with an order of 4 and one iteration. A grid-based neighbor list with a threshold of 10 Å was used and updated every five steps (10 fs). The particle-mesh Ewald method<sup>23,24</sup> was used for long-range electrostatic interactions above 10 Å with a fourth order interpolation and a maximum spacing for the FFT grid of 1.6 Å. Lennard–Jones interactions were cutoff above 10 Å. A long range dispersion correction for energy and pressure was used to compensate for the Lennard–Jones interaction cutoff.<sup>16</sup>



**FIGURE 2** Folding pathway of peptides 1 and 2. The most probable folding route from the completely unfolded *oooo* state to the completely folded *cccc* conformation was derived from the population change maps shown in Table II. The probability of each state at room temperature is denoted below the hydrogen bond schemes. Probabilities of the less populated states are given in the Supporting Information.

## RESULTS AND DISCUSSION

### Population Analysis

MD frames were classified according to the presence or absence of the possible interchain hydrogen bonds of the peptides' folded  $\beta$ -hairpin conformation, using the on/off scheme for HB1-4 shown in Figure 1. Hydrogen bonds were detected with the software MDAnalysis<sup>25</sup> with a distance threshold of 3 Å and an angle lower limit of 120°. If these criteria were met, a hydrogen bond was labelled as *c* (closed), otherwise it was labelled as *o* (opened). The four possible hydrogen bonds yield  $2^4 = 16$  theoretically available hydrogen bond patterns, yet only around half of the combinations were observed to be significantly populated. The contribution of the six most abundant hydrogen bond patterns is shown in Table I, whereas the populations of all conformations are shown in Supporting Information Table SI.

Overall folding of 1 and 2 was analysed by determining the average distances of the interchain hydrogen bonds HB1-4 (Figure 1) and assigning the structures that possess an average HB distance below the 3 Å threshold as folded, whilst those with an average distance above this threshold as unfolded (last column in Table I). This characterization of the MD trajectory yields an overall folding of 66% for 1 and 43% for 2, which are in agreement with the experimentally derived 88% versus 55%, respectively, within the limitations of the applied techniques.<sup>14</sup>

The hydrogen bond distances of the folded conformations *cccc*, *ccoc*, and *cocc* of 1 and 2, shown in Table I, were comparable. This confirmed that the OH to CH<sub>3</sub> substitution at amino acid 3 does not lead to any distortion of the overall backbone conformation of the peptide. Interestingly, barely 3% of the frames in the trajectory of 1 display the S3-S8 interstrand

**Table II** Population Change Maps for the Conformations of Peptides 1 and 2

Peptide 1						
From	To					
	<i>oooo</i>	<i>ooc</i>	<i>cooc</i>	<i>cocc</i>	<i>ccoc</i>	<i>cccc</i>
<i>oooo</i>	94%	4%	0%	0%	0%	0%
<i>ooc</i>	9%	49%	23%	8%	3%	1%
<i>cooc</i>	0%	3%	72%	14%	7%	2%
<i>cocc</i>	0%	1%	6%	79%	1%	8%
<i>ccoc</i>	0%	2%	39%	8%	37%	9%
<i>cccc</i>	0%	0%	6%	39%	3%	46%

Peptide 2						
From	To					
	<i>oooo</i>	<i>ooc</i>	<i>cooc</i>	<i>cocc</i>	<i>ccoc</i>	<i>cccc</i>
<i>oooo</i>	87%	10%	0%	0%	0%	0%
<i>ooc</i>	21%	65%	8%	2%	1%	0%
<i>cooc</i>	0%	2%	75%	11%	8%	2%
<i>cocc</i>	0%	1%	11%	73%	1%	9%
<i>ccoc</i>	0%	1%	39%	6%	40%	7%
<i>cccc</i>	0%	0%	8%	34%	5%	47%

Classification and quantification of these groups are given in Table I. Population change maps for all the theoretically possible 16 population groups are shown in the Supporting Information (Table SIV).

hydrogen bond that may not explain the higher  $\beta$ -hairpin propensity of 1 as compared to 2. The S3 sidechain of 1, however, can additionally form intrastrand hydrogen bond with the carbonyl oxygen of S3 (18%) and that of A2 (10%), which may stabilize an extended  $\beta$ -strand and thereby the  $\beta$ -hairpin conformation (for additional details see Supporting Information Table SII). These hydrogen bonds are predicted to preferentially occur in combination with the *cocc* hydrogen bond pattern (Supporting Information Table SIII), explaining its higher prevalence for 1 (47%) as compared to 2 (29%), and possibly the higher overall folding rate.

### Population Change Analysis

Molecular dynamics trajectories may provide a stochastic model for biological processes, and were previously used to characterize folding and molecular recognition events.<sup>26-29</sup> Population change maps, shown in Table II, were generated by counting the number of transitions, in percentage, in the MD trajectory between the hydrogen bond patterns shown in Table I. The line in Table II specifies from which population the transition starts and the column in which population it ends. The population analysis of 1 and 2 reveals that a majority of transitions return to the same population group they started from,

as indicated by the highest probabilities belonging to the diagonals. The most probable folding pathway is indicated by the highest values in respective rows of Table II, ignoring the diagonal value, and is shown in Figure 2.

Both peptides follow the same order of hydrogen bond formation. Thus, starting from the fully unfolded state, HB4 is formed first yielding structure *oooc*. This is followed by formation of HB1 that leads to the structure *cooc*. Next, either HB2 or HB3 is closed with somewhat different probabilities, giving first *cocc* or *ccoc* and then leading to the fully folded  $\beta$ -hairpin, *ccc*. The sequence of hydrogen bond formation upon folding may offer valuable information for designing peptides with specific structural properties. It should be noted that folding pathways may be sensitive towards force field parameters,<sup>30</sup> and the above proposed folding pathway should therefore be interpreted with care.

## SUMMARY

The ability of a simple, standard MD simulation protocol for prediction of  $\beta$ -hairpin folding was demonstrated. A straightforward hydrogen bond analysis of the frames in an MD trajectory was shown to yield folding ratios that are in good agreement with experimental (NMR) observations.<sup>14</sup> Moreover, we have shown that the applied MD simulation can predict the importance of a specific secondary interaction, here S3–S8, on folding ratios. It may also provide an improved understanding of the impact of amide to amide interstrand hydrogen bonds and the dynamics of peptide folding.

## REFERENCES

1. Craik, D. J.; Fairlie, D. P.; Liras, S.; Price, D. *Chem Biol Drug Des* 2013, 81, 136–147.
2. Vlieghe, P.; Lisowski, V.; Martinez, J.; Khrestchatisky, M. *Drug Discov Today* 2010, 15, 40–56.
3. Sun, L. *Modern Chem Appl* 2013, 1, e103.
4. Koukos, P. I.; Glykos, N. M. *J Phys Chem B* 2014, 118, 10076–10084.
5. Shen, Y.; Maupetit, J.; Derreumaux, P.; Tuffery, P. *J Chem Theor Comput* 2014, 10, 4745–4758.
6. Lewandowska, A.; O Idziej, S.; Liwo, A.; Scheraga, H. A. *Biophys Chem* 2010, 151, 1–9.
7. Balaram, H.; Prasad, B. V.; Balaram, P. *J Am Chem Soc* 1983, 105, 4065–4071.
8. Cicero, D.; Barbato, G.; Bazzo, R. *J Am Chem Soc* 1995, 117, 1027–1033.
9. Cumming, D. A.; Carver, J. P. *Biochemistry* 1987, 26, 6664–6676.
10. Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E. *Angew Chem Int Ed* 1999, 38, 236–240.
11. Kanyalkar, M.; Srivastava, S.; Coutinho, E. *Biomaterials* 2002, 23, 389–396.
12. Johnston, J. M.; Cook, G. A.; Tomich, J. M.; Sansom, M. S. *Biophys J* 2006, 90, 1855–1864.
13. Yang, C.; Li, J.; Li, Y.; Zhu, X. *J Mol Struct Theochem* 2009, 895, 1–8.
14. Danelius, E.; Brath, U.; Erdélyi, M. *Synlett* 2013, 24, 2407–2410.
15. Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. *Bioinformatics* 2013, 29, 845–854.
16. Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J Chem Theor Comput* 2008, 4, 435–447.
17. Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. *J Phys Chem B* 2001, 105, 6474–6487.
18. Rizzo, R. C.; Jorgensen, W. L. *J Am Chem Soc* 1999, 121, 4827–4836.
19. Bussi, G.; Donadio, D.; Parrinello, M. *J Chem Phys* 2007, 126, 014101.
20. Nosé, S.; Klein, M. *Mol Phys* 1983, 50, 1055–1076.
21. Parrinello, M.; Rahman, A. *J Appl Phys* 1981, 52, 7182–7190.
22. Hess, B.; Bekker, H.; Berendsen, H. J.; Fraaije, J. G. E. M. *J Comput Chem* 1997, 18, 1463–1472.
23. Darden, T.; York, D.; Pedersen, L. *J Chem Phys* 1993, 98, 10089–10092.
24. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J Chem Phys* 1995, 103, 8577–8593.
25. Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. *J Comput Chem* 2011, 32, 2319–2327.
26. Schwantes, C.; McGibbon, R.; Pande, V. *J Chem Phys* 2014, 141, 090901.
27. Park, S.; Klein, T. E.; Pande, V. S. *Biophys J* 2007, 93, 4108–4115.
28. Lane, T. J.; Bowman, G. R.; Beauchamp, K.; Voelz, V. A.; Pande, V. S. *J Am Chem Soc* 2011, 133, 18413–18419.
29. Buch, I.; Giorgino, T.; De Fabritiis, G. *Proc Natl Acad Sci USA* 2011, 108, 10184–10189.
30. Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. *Biophys J* 2011, 100, L47–L49.