

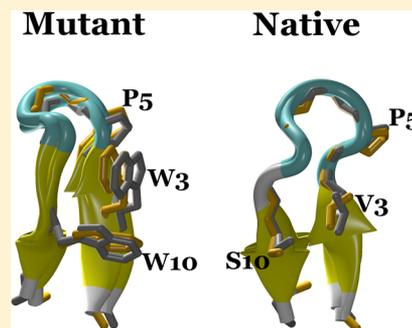
# Folding Molecular Dynamics Simulations Accurately Predict the Effect of Mutations on the Stability and Structure of a Vammin-Derived Peptide

Panagiotis I. Koukos and Nicholas M. Glykos\*

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

## Supporting Information

**ABSTRACT:** Folding molecular dynamics simulations amounting to a grand total of 4  $\mu$ s of simulation time were performed on two peptides (with native and mutated sequences) derived from loop 3 of the vammin protein and the results compared with the experimentally known peptide stabilities and structures. The simulations faithfully and accurately reproduce the major experimental findings and show that (a) the native peptide is mostly disordered in solution, (b) the mutant peptide has a well-defined and stable structure, and (c) the structure of the mutant is an irregular  $\beta$ -hairpin with a non-glycine  $\beta$ -bulge, in excellent agreement with the peptide's known NMR structure. Additionally, the simulations also predict the presence of a very small  $\beta$ -hairpin-like population for the native peptide but surprisingly indicate that this population is structurally more similar to the structure of the native peptide as observed in the vammin protein than to the NMR structure of the isolated mutant peptide. We conclude that, at least for the given system, force field, and simulation protocol, folding molecular dynamics simulations appear to be successful in reproducing the experimentally accessible physical reality to a satisfactory level of detail and accuracy.



## 1. INTRODUCTION

The increased accuracy of the latest generation of empirical force fields together with the long time scales currently available has transformed the field of molecular dynamics simulations.<sup>1</sup> Numerous studies have demonstrated the ability of folding molecular dynamics simulations to accurately reproduce the folded state of small proteins and peptides.<sup>2–13</sup> Especially in the field of peptide structure, a consensus appears to slowly emerge: the AMBER99SB family of force fields (mainly the ILDN and STAR-ILDN variants<sup>14–16</sup>) with the TIP3P water model<sup>17</sup> and full (PME-based) electrostatics<sup>18</sup> have repeatedly been shown to be able to accurately predict the structure and dynamics of peptides ranging from very stable folders,<sup>5,13–16</sup> to mostly disordered peptides,<sup>12,13,19–21</sup> and for all structural motifs from mainly helical<sup>5,12,15,16</sup> to almost exclusively  $\beta$ -hairpin-like.<sup>16,21</sup> This increased accuracy has shifted the interest to the question of how accurate and sensitive are folding simulations in predicting the effect of relatively small changes in the initial conditions.

In this spirit, we examine here the ability of folding molecular dynamics simulations—in explicit solvent and with full electrostatics—to reproduce the effects of mutations on peptide structure and dynamics. The system we selected to study is based on two variants of the vammin-derived peptides that were synthesized and thoroughly examined experimentally by the Jiménez group.<sup>22</sup> The two peptides are (a) the native peptide corresponding to loop 3 of the vammin protein (PDB entry 1WQ8, residues 69–80) hereafter referred to as the native (Nat) peptide and (b) the W3W10 double tryptophan mutant designed by Mirassou et al.,<sup>22</sup> hereafter referred to as

the mutant (2W) peptide. The sequences of these two peptides (with the mutated sites highlighted in red) are shown in the lower panel of Figure 1. The major experimental findings<sup>22</sup> with which the results from the simulations will be compared are as follows:

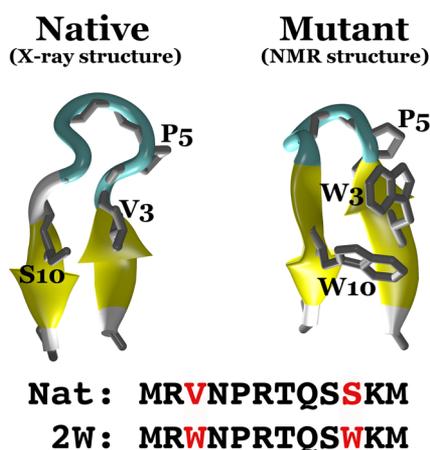
(a) The isolated native peptide is largely disordered in solution, with no discernible long-range NOEs and with  $H_{\omega}$ ,  $^{13}C_{\omega}$ , and  $^{13}C_{\beta}$  chemical shifts that are very close to their expected random coil values.<sup>22</sup> This is to be contrasted with the irregular  $\beta$ -hairpin structure that this same peptide adopts when observed as part of the complete vammin protein X-ray structure (the structure of the peptide as seen in the vammin 1WQ8 crystal structure is shown in the upper panel of Figure 1). The implication is clear: the native peptide's irregular  $\beta$ -hairpin structure (Figure 1) is only stable within the scaffold of the fully folded protein but not in isolation. The reason for that is probably that in the fully folded protein the  $\beta$ -hairpin is flanked by a third (antiparallel)  $\beta$ -strand with which it forms a three-stranded antiparallel  $\beta$ -sheet.

(b) In contrast with the native peptide, the double tryptophan mutant ( $V_3 \rightarrow W$ ,  $S_{10} \rightarrow W$ ) assumes a very stable structure in solution. The adopted structure was shown<sup>22</sup> to be almost as stable as that of a designed quadruple mutant containing a disulfide bridge connecting the termini of the

Received: May 10, 2014

Revised: August 5, 2014

Published: August 6, 2014



**Figure 1.** Peptide structures and sequences. The upper panel shows schematic diagrams of the structures of (a) the native peptide as observed in the crystal structure of the vamin protein (PDB entry 1WQ8, residues 69–80) and (b) the representative solution (NMR) structure of the 2W mutant. Please note that the isolated native peptide is disordered in solution and the structure shown here is only meaningful within the context of the fully folded vamin protein. In both diagrams, residues 3, 5, and 10 are marked to aid interpretation and cartoon representations of the secondary structure elements have been used with the following color coding of the STRIDE-derived secondary structure assignments: yellow  $\rightarrow$   $\beta$  structure, cyan  $\rightarrow$  turns, white  $\rightarrow$  coil. The lower panel shows the sequences of the native (Nat) and mutant (2W) peptides with the mutated sites highlighted in red.

peptide (the C1C12W3W10 peptide in the Mirassou et al. nomenclature<sup>22</sup>).

(c) This stable structure of the 2W mutant was determined via NMR spectroscopy and was shown to be an irregular 4:6  $\beta$ -hairpin with a non-glycine  $\beta$ -bulge, an edge-to-face stacking of the tryptophans' indole rings (with W3 at the edge position and W10 at the face position), and unusual dihedral angles for the Q8 (bulge) residue located in the  $\alpha_L$  region of the Ramachandran plot.<sup>23</sup> A schematic representation of the NMR structure of the isolated 2W mutant demonstrating the tryptophan stacking is shown in Figure 1. Note that the difference in stability between the native and mutant structures is not due to differences in the hydrogen bonding pattern which is virtually unchanged between the two structures shown in Figure 1.

In the following paragraphs, we describe the simulation protocol and the analyses performed, followed by a detailed comparison between the computational results and the experimental findings. We conclude by discussing the implications of these results with an emphasis on the question of how close are peptide folding simulations to the stage of serving as dependable and robust analytical tools for predicting peptide structure and dynamics.

## 2. METHODS

**2.1. System Preparation and Simulation Protocol.** The starting peptide structures were 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 were performed with the program LEAP from the AMBER tools distribution.<sup>24</sup> For both simulations, we used periodic boundary conditions and a cubic unit cell sufficiently large to guarantee a minimum separation

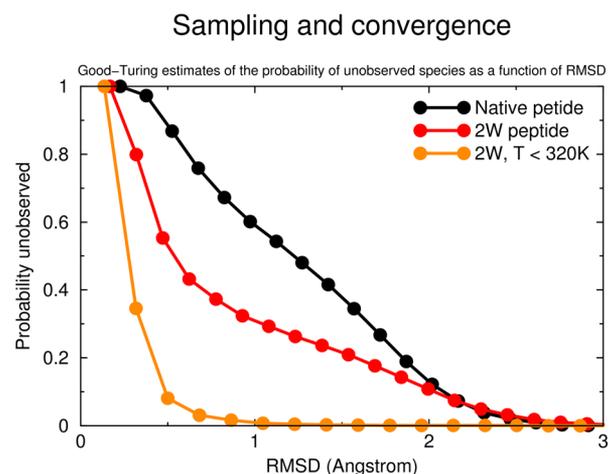
between the PBC-related images of the peptides of at least 16 Å. We followed the dynamics of the peptides' folding simulations using the program NAMD<sup>26</sup> for a grand total of 4  $\mu$ s (2  $\mu$ s for each peptide) using the TIP3P water model,<sup>17</sup> the AMBER99SB-ILDN force field,<sup>16</sup> and adaptive tempering<sup>25</sup> as implemented in the program NAMD (adaptive tempering is formally equivalent to a single-copy replica exchange folding simulation with a continuous temperature range; for our simulations, this temperature range was 300–500 K inclusive and was applied to the system through the Langevin thermostat, see below).

The simulation protocol for both peptides was the following. The system was first energy minimized for 1000 conjugate gradient steps followed by a slow heating-up phase to the final temperature of 300 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 Langevin dynamics and the Langevin piston barostat control methods as implemented by the NAMD program, with adaptive tempering applied through the Langevin thermostat, while the pressure was maintained at 1 atm. The Langevin damping coefficient was set to 1 ps<sup>-1</sup>, 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 time step integration algorithm as implemented by NAMD. The inner time step was 2 fs, short-range nonbonded interactions were calculated every one step, and long-range electrostatics interactions every two time steps using the particle mesh Ewald method with a grid spacing of approximately 1 Å and a tolerance of 10<sup>-6</sup>. A cutoff for the van der Waals interactions was applied at 8 Å through a switching function, and SHAKE (with a tolerance of 10<sup>-8</sup>) 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.

**2.2. Trajectory Analysis.** The programs CARMA<sup>27</sup> and GRCARMA<sup>28</sup> have been used for almost all 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, calculation of similarity  $Q$  values, etc. Secondary structure assignments were calculated with the program STRIDE.<sup>29</sup> All molecular graphics work and figure preparation were performed with the programs VMD,<sup>30</sup> RASTER3D,<sup>31</sup> and CARMA.

**2.3. Extent of Sampling.** Quantifying the extent of sampling of molecular dynamics trajectories is always difficult, even more so in the case of folding simulations where the configurational space is vast. There are two aspects of the problem. The first is to quantify the extent of sampling with respect to the full configurational space accessible to the system, irrespectively of whether the peptides are in the folded or unfolded state. For all reasonable simulation time scales—and considering the number of feasible unfolded conformations—this (global) coverage will almost certainly be poor. The second—more relevant—question concerns the extent of sampling of only the peptides' stable conformers. Such a

differentiation is feasible in our case because stable conformers are associated with low adaptive tempering temperatures. We tackle both of these questions through the application of a recently described approach<sup>32</sup> which is based on the application of Good-Turing statistics to estimate the probability of unobserved peptide conformations as a function of the RMSD from all already observed (in the trajectory) structures. For the calculation, we have used the computer program available via <https://github.com/pkoukos/GoodTuringMD>. The results are shown in Figure 2 in the form of three graphs.

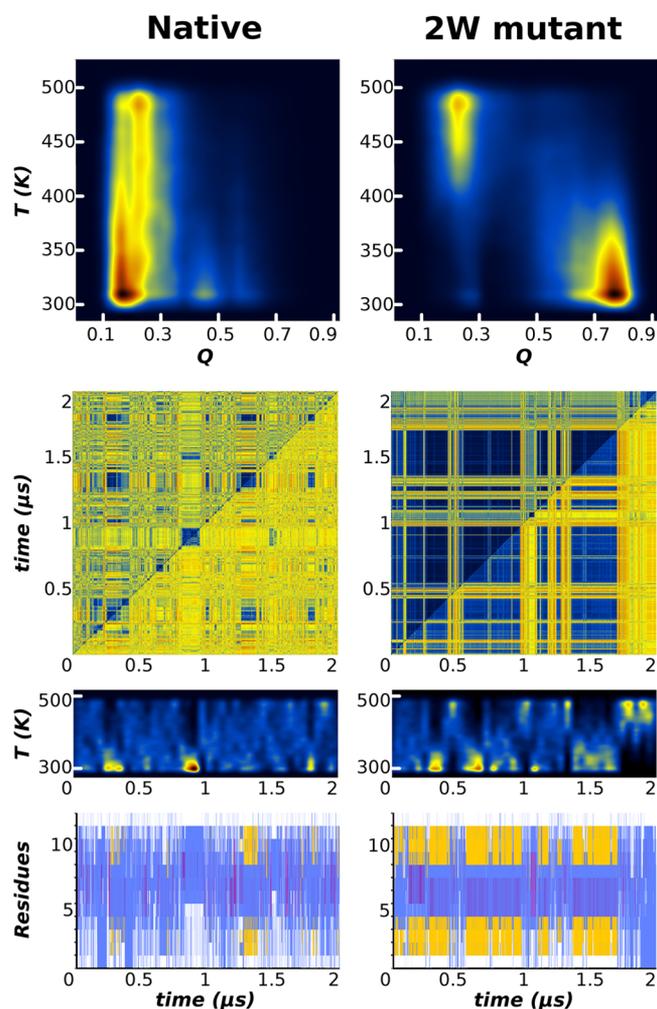


**Figure 2.** Convergence and sufficient sampling. Results from the application of Good-Turing statistics to estimate the extent of sampling of the native and mutant trajectories as produced by the GoodTuringMD program (<https://github.com/pkoukos/GoodTuringMD>). The two upper graphs (black and red) are the estimates of the probability of unobserved species (i.e., thus far unobserved peptide conformers) as a function of the RMSD from all already observed conformations of the native (black curve) and mutant (red curve) trajectories. The lower (orange) curve is the Good-Turing estimate for the 2W trajectory but only including peptide structures whose corresponding adaptive tempering temperature was less than 320 K. See section 2.3 for details.

The two upper (black and red) graphs in this figure quantify and compare the extent of sampling for the two peptides, without differentiating between stable or unstable and folded or unfolded configurations. These graphs clearly show the better sampling for the more stable 2W peptide, as will be discussed in the next section. The lower (orange) curve in Figure 2 is the Good-Turing estimate for the 2W trajectory but only including peptide structures whose corresponding adaptive tempering temperature was less than 320 K. Because these structures mostly correspond to stable peptide conformations, the lower curve is the answer to the question “what is the probability that if we continue the simulation we will observe a stable peptide conformation with an RMSD (from all other already observed stable peptide conformations) higher than X Angstrom?” Clearly, the analysis suggests that all major stable conformers of 2W have been sampled. The corresponding analysis (i.e., with a temperature cutoff) for the native peptide is meaningless, since the native peptide is disordered (see next section).

### 3. RESULTS

**3.1. The Native Peptide Is Mostly Disordered, the Mutant Peptide Has a Well-Defined Structure.** The upper row of graphs in Figure 3 compares the log density



**Figure 3.** Comparative analysis of the native and mutant trajectories. The four rows compare from top to bottom (a) the  $Q$ - $T$  diagrams, (b) the RMSD matrices, (c) the adaptive tempering temperature distributions, and (d) the secondary structure assignments. See section 3.1 for a detailed discussion of this figure.

distributions of ( $T_{\text{adaptive}}$  vs  $Q$ ) for the two simulations, where  $T_{\text{adaptive}}$  is the adaptive tempering temperature,  $Q$  is the similarity index<sup>33</sup> to the experimentally known peptide structures (with a value of 1.0 corresponding to a structure identical with the experimentally known), and hot colors in the graph indicate high log density (i.e., that a large number of structures from the trajectory have the corresponding values of  $Q$ - $T$ ). Because these graphs contain a wealth of information but are not very common in the literature, we will discuss them in some detail. We start by presenting the expected form of these graphs in the case of an ideally disordered peptide (meaning a peptide with a completely flat energy landscape for any and all of the temperatures visited during a given simulation) and compare it with what would be expected from a peptide with an ideally perfect folding funnel. An ideally perfect disordered peptide would—at least for all reasonably available simulation time scales—give a density distribution that would resemble a vertical line close to zero on the  $Q$ -values and covering the whole range of temperatures. The reason is, of course, that a disordered peptide will spend almost all of the simulation time away from any given structure irrespectively of the temperature of the system. On the other hand, a peptide

with an ideally perfect folding funnel would give a graph resembling a section through a half-funnel: structures with low  $Q$  (disordered-unfolded configurations) would be associated with high temperatures, and structures with high  $Q$  (native-like, stable structures) would be associated with low temperatures. As is obvious from Figure 3, the behavior of the two peptides follows closely the descriptions above. The mutant (2W) peptide shows a clear funnel-like distribution with two major concentrations, one at the high( $Q$ )–low( $T$ ) part of the diagram (lower right, corresponding to the basin of the folding funnel) and the other at the low( $Q$ )–high( $T$ ) end of the diagram (upper left, corresponding to unfolded conformations). In contrast with the mutant, the native peptide lacks any significant features at the high( $Q$ )–low( $T$ ) part, and the distribution closely resembles a vertical line at low  $Q$  values, as would be expected from a largely disordered peptide. Even at this early level of the analysis, the differences between the two peptides are so pronounced that they leave little doubt: the native peptide is disordered, and the 2W mutant has a funnel-like folding landscape with a single deep minimum that corresponds to a stable structure very similar to the one determined experimentally.

The RMSD matrices shown in the second row of Figure 3 corroborate with and expand the conclusions drawn from the  $Q$ – $T$  diagrams. These matrices represent an all-to-all comparison between the structures recorded from each of the corresponding trajectories, and are presented in this figure with a color coding in which dark blue corresponds to low RMSD values (very similar structures), going through yellow (RMSDs of approximately 6 Å) to dark red which corresponds to high RMSD values (dissimilar structures, RMSDs of approximately 12 Å). The upper half of these matrices was calculated using the  $C_\alpha$  atoms only, and the lower half, all non-hydrogen atoms. Dark blue boxes centered on the diagonal of those diagrams indicate stable peptide conformations that persist in time, yellow-red areas indicate disordered fast-interconverting peptide conformations, and off-diagonal dark blue areas indicate that the simulation repeatedly and independently visited the same peptide structure. Clearly, the native peptide hardly takes any long-lasting stable conformations and spends almost all of the 2  $\mu$ s long simulation interconverting between numerous different peptide conformations. In contrast, the 2W mutant spends most of the simulation time having a stable structure (blue areas of the diagram) which is being recurrently visited through several unfolding and refolding events. The unfolding events are associated with high adaptive tempering temperatures, as can be seen in the graphs immediately below the RMSD matrices. These diagrams (third row of Figure 3) show the log density distribution of the adaptive tempering temperatures as a function of simulation time and are in a one-to-one correspondence with the RMSD matrices in the vertical direction. Comparison of the RMSD matrix with the temperature distribution for the 2W peptide clearly indicates that unstable fast-converting peptide structures (yellow-red areas of the RMSD matrix) correspond to the high temperature range of adaptive tempering.

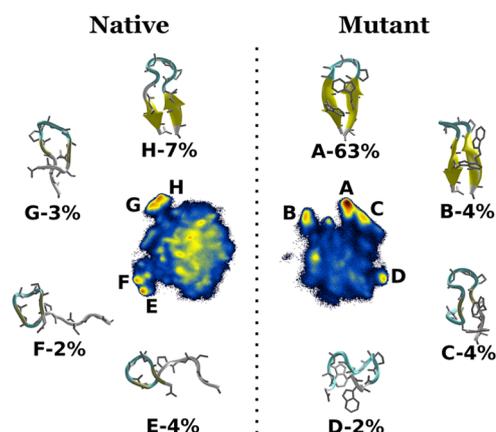
Finally, the graphs in the last row of Figure 3 show the per-residue secondary structure assignments for the two simulations as a function of time. The assignments are as obtained by STRIDE with yellow corresponding to  $\beta$  structure, cyan-magenta to turns and  $3_{10}$ -like turns, and white to coil. Taken together with the RMSD matrices (with which they are in a one-to-one correspondence), a clear picture emerges: the 2W

peptide takes up a stable  $\beta$ -hairpin ( $\beta$ -turn- $\beta$ ) structure that persists for the great majority of the simulation and is only given up at the high temperature regime of adaptive tempering. The native peptide interconverts between numerous unstable conformations, with only a very minor  $\beta$ -hairpin population which was briefly visited at approximately 0.3, 1.4, and 1.8  $\mu$ s (this conformation corresponds to the very light blue area centered at  $Q \sim 0.6$  in the native peptide's  $Q$ – $T$  diagram).

**3.2. The Major Conformation for Both Peptides Is a  $\beta$ -Hairpin but with Large Differences in Stability.** If molecular dynamics is to be used as a stand-alone analysis tool aiming to identify from first-principles whether a given peptide has a stable structure, and if yes, which one this stable structure is, it is important that the selection of a representative structure from the molecular dynamics trajectories is unbiased, i.e., made completely agnostic of any experimental information available for the systems being studied. For this work, unbiased representative structures were selected as follows. In the first stage, dihedral principal component analysis (dPCA) was performed,<sup>34,35</sup> and an initial set of clusters was identified by three-dimensional dPCA cluster analysis as performed by the program CARMA. In the second stage, these dPCA-derived clusters were used as input to a three-dimensional Cartesian PCA clustering but using only the peptides' backbone atoms. The result from these two stages is a set of prominent clusters with distinct backbone conformations but without any differentiation with respect to putative heterogeneity in the side chains' conformations. In the final step, these clusters were further analyzed using another round of Cartesian PCA, but this time using all of the peptides' non-hydrogen atoms. Representative structures for these final clusters 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.

Figure 4 is a collage showing schematic diagrams of the representative structures corresponding to the four most highly populated clusters of the native and mutant peptides as identified by the procedure described above. The guiding diagrams in the middle of this figure are the log density projections of the trajectories on the first two principal components derived from dPCA, demonstrating again the large differences between the folding behavior of the two peptides. The marked peaks (A through D for the mutant, E through H for the native peptide) are in a one-to-one correspondence with the structural diagrams at the periphery of the diagram (noting again that the actual cluster analysis was performed in the three-dimensional space defined by the top three principal components, and not in two dimensions). The numbers below each structure schematic are the percentage of simulation time that each of the corresponding dPCA-derived clusters occupied. On the whole, Figure 4 places all observations up to now on a structurally solid ground: The mutant peptide has a very stable and persistent  $\beta$ -hairpin structure (marked A in Figure 4) that occupies the great majority of the trajectory. This is more so if we note that the structures marked as B and C have essentially the same  $\beta$ -hairpin characteristics with a stabilizing Trp-Trp edge-to-face stacking.

The most highly populated cluster recorded from the native peptide simulation is again a  $\beta$ -hairpin structure (marked H in Figure 4), which, however, is quite distinct from the mutant structure, and occupies only 7% of the trajectory, clearly indicating again the absence of a stable and persistent



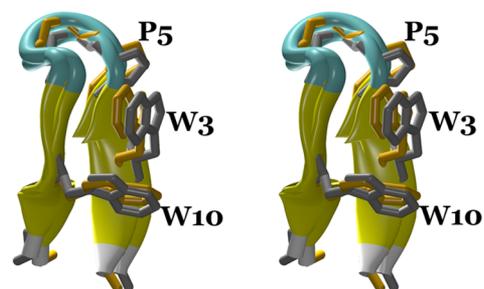
**Figure 4.** Clusters and their frequencies. The two diagrams in the center of the image are the log density projections of the corresponding trajectories on the top two principal components. Prominent clusters are marked on these diagrams (A–D for the mutant peptide, E–H for the native). The structure diagrams in the periphery of the image are the corresponding unbiased representative structures which were calculated as described in section 3.2, with cartoon representations of the secondary structure elements and color coded according to the STRIDE-derived secondary structure assignments. The numbers below the structure diagrams are the cluster frequencies expressed as a percentage of the total simulation time for each peptide.

conformation for this peptide. The percentage for this cluster is actually as low as the sum of the E and F conformers, both of which display a clearly non-native conformation with a disordered N-terminal region and an ordered C-terminal turn (these two peptide structures both correspond to the stable conformation that can be seen in Figure 3 diagrams at approximately  $0.9 \mu\text{s}$ ).

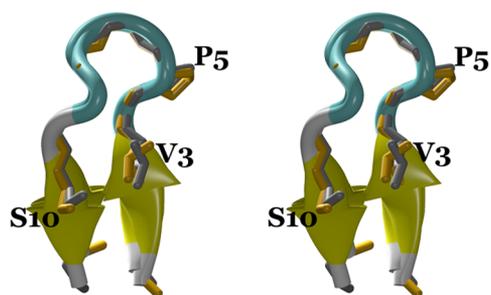
**3.3. The Structures Derived from Molecular Dynamics Are in Excellent Agreement with the Experimentally Determined Ones.** The wall-eyed stereodiagrams in Figure 5 are a direct comparison between the representative molecular-dynamics-derived peptide structures with the experimentally determined (NMR and X-ray) structures. The upper stereodiagram in Figure 5 compares the unbiased representative structure of the major cluster from the mutant simulation (marked A in Figure 4) with the representative NMR structure as determined by Mirassou et al.<sup>22</sup> and shown in Figure 1. The agreement between the least-squares superimposed structures is excellent down to the level of the conformation of individual side chains, especially the tryptophans' edge-to-face stacking. To put this in numbers, we calculated the RMSD between the backbone atoms of the two structures (excluding the two hyper-mobile terminal residues) and this was found to be only  $0.79 \text{ \AA}$ , comparable with the experimental uncertainty of the atomic positions. If instead of using the unbiased molecular dynamics structures for calculating the RMSD we select the structure (from the trajectory) that best agrees with the experimental NMR structure, the RMSDs drop down to  $0.23 \text{ \AA}$  using the  $C_{\alpha}$  atoms and to  $0.33 \text{ \AA}$  using all backbone atoms, demonstrating again the excellent agreement between experiment and simulation. What is more important, however, is the ability of the folding simulations to reproduce the  $\beta$ -bulge and the associated unusual dihedral angles for the bulge (Q8) residue. This is discussed in more detail in the next section.

What was rather unexpected was the finding that the rather unstable  $\beta$ -hairpin structure adopted by the native peptide in its

## Mutant (MD vs NMR)



## Native (MD vs X-ray)

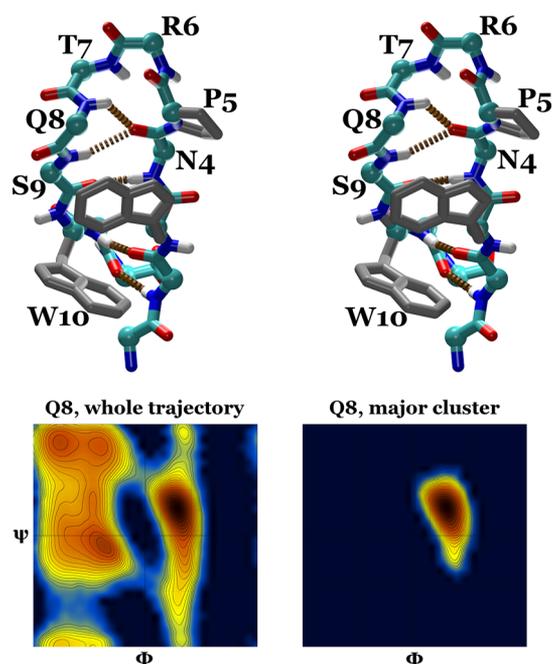


**Figure 5.** Experiment vs simulation: Structural comparison. The upper wall-eyed stereodiagram compares for the 2W mutant the representative molecular-dynamics-derived structure (colored orange, same as the structure “A” in Figure 4) with the experimentally (NMR) determined one (colored gray, same with Figure 1). The lower stereodiagram shows the same comparison for the native peptide but using the experimental structure observed in the vammim protein crystal structure (as shown in Figure 1). For both diagrams, cartoon representations of the secondary structure elements are shown to aid interpretation with the coloring scheme described in the legend of Figure 1.

folding simulation turned out to be more similar to the structure of this same peptide as seen in the complete vammim protein crystal structure than the structure seen in the NMR determination of the peptide's mutant form. Actually, and as can be discerned from the lower stereodiagram of Figure 5, the agreement between the least-squares superimposed X-ray and molecular-dynamics-derived structures is even better than that of the mutant. The backbone atom RMSD between the representative structures was only  $0.69 \text{ \AA}$ , reducing even further to  $0.35 \text{ \AA}$  ( $C_{\alpha}$  atoms) and  $0.44 \text{ \AA}$  (all backbone atoms) when the best-agreeing structure from the trajectory was selected. Having made that observation, the question naturally arises as to its significance. It is clearly tempting to invoke the well-known argument concerning pre-existing structural preferences of peptides which are manifested and maintained in the structures they adopt inside proteins. For the specific case considered here, we believe that the evidence is circumstantial and inconclusive: the native peptide has been studied experimentally using NMR<sup>22</sup> and no stable structure could be discerned within the detection limits of the method. The very small  $\beta$ -hairpin population indicated by molecular dynamics agrees with the absence of a stable structure, but clearly, this agreement does not validate any short-lived structure that could

have been observed during the simulation. Having said that, we shall not resist the temptation of noting the improbability that it is only by pure coincidence that the following two statements turned out to be concurrently true, i.e., that (a) the known NMR structure of the mutant agrees much better with the structure obtained from the mutant simulation but not with the structure obtained from the native simulation and (b) the structure of the native peptide from the vamin X-ray structure agrees much better with the structure obtained from the native simulation but not with the structure obtained from the mutant simulation.

**3.4. The Simulation Faithfully Reproduces the  $\beta$ -Bulge, the Tryptophans' Edge-to-Face Stacking, and the Unusual Dihedral Angles for the Q8 Residue.** With such an excellent agreement between the structures derived from simulation and experiment (see section 3.3 and Figure 5), it is not surprising that the unusual geometry of the mutant  $\beta$ -hairpin is faithfully reproduced by the folding simulation. This is shown in Figure 6 (upper wall-eyed stereodiagram). The



**Figure 6.**  $\beta$ -bulge, Trp stacking, and Q8's dihedral angles. The wall-eyed stereodiagram (upper panel) is a skeletal model of the backbone of the mutant's representative molecular dynamics structure. To aid interpretation, the  $C_{\alpha}$  atoms are shown as spheres and the side chains of W3, P5, and W10 are drawn in full (colored gray). The broken lines (colored brown) represent hydrogen bonds, with the bifurcated bond involving N4, Q8, and S9 clearly visible. The classical edge-to-face stacking of the tryptophans' indole rings is clearly visible. The lower panel compares the Ramachandran plots (in log density units) of residue Q8 when using (a) the whole of the mutant simulation (left diagram) and (b) only the structures of the major ( $\beta$ -hairpin) cluster (right diagram). The major peak in both diagrams coincides with the  $\alpha_L$  region of the Ramachandran plot.

structure in this figure is again the representative 2W structure, but this time with all backbone atoms and their hydrogen bonds (broken lines) drawn. The view is from the top (loop region) of the peptide structure and clearly demonstrates the bulge and the bifurcated hydrogen bond, in excellent agreement with the experimental studies.

The same figure also demonstrates the tryptophan stacking as obtained from the simulation. The arrangement is a classical edge-to-face stacking of the indole rings with W3 at the edge position and W10 at the face position (seen also in Figure 5), again in full agreement with the experimental findings.

Finally, the lower panel of Figure 6 shows an analysis of the dihedral angles of the mutant peptide's Q8 residue which was shown by Mirassou et al.<sup>22</sup> to occupy the  $\alpha_L$  region of the Ramachandran plot. The diagram on the left is the (log density) Ramachandran plot of Q8 using all peptide structures recorded in the whole of the 2W trajectory. This is to be compared with the distribution (right diagram) obtained when only the major ( $\beta$ -hairpin) cluster was used for the analysis (cluster A in Figure 4). Again, the simulation faithfully reproduces the experimental observations.

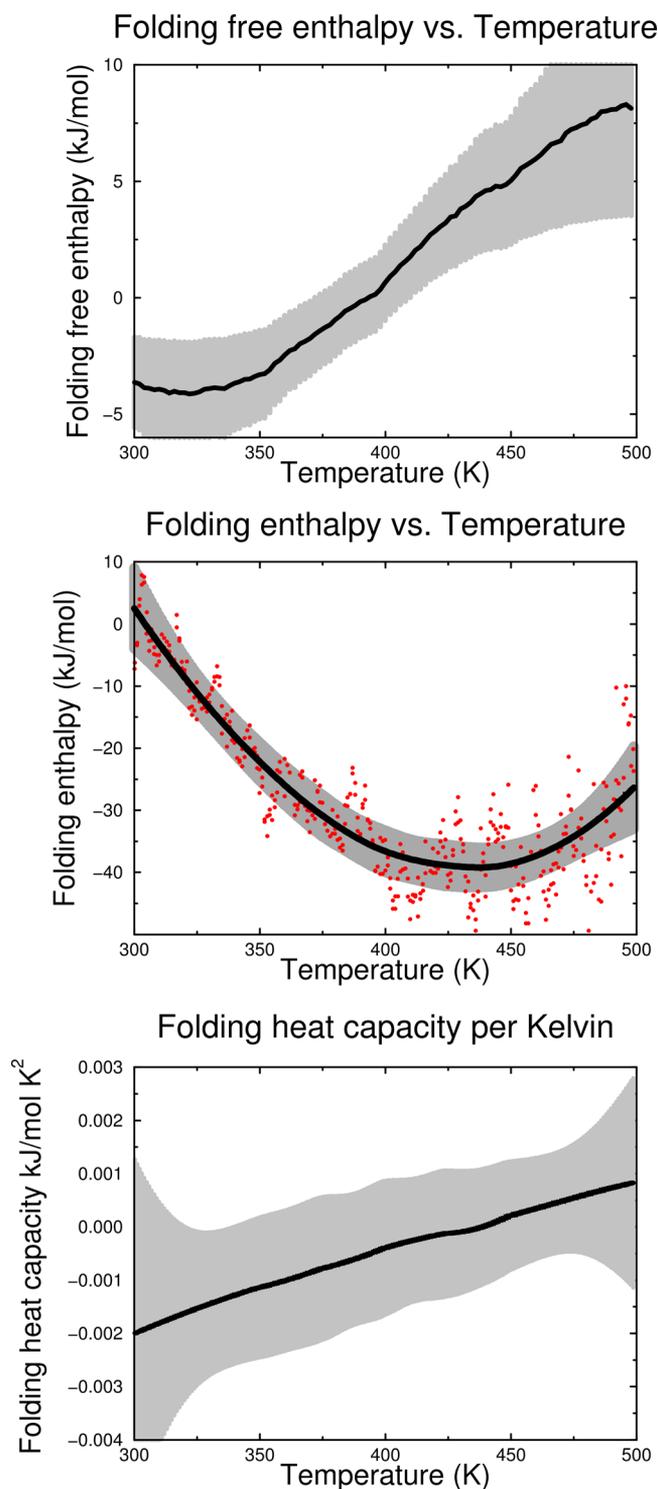
**3.5. Folding Thermodynamics Calculations Indicate That the Mutant Peptide Is Extremely Stable and That Its Folding Is Entropy-Driven at Low Temperatures.** The discussion up to now focused almost exclusively on the structural aspects of the native and mutant peptides' behavior, systematically ignoring the underlying folding thermodynamics. There are two reasons for that. The first—and by far the most important—is that there are no direct experimental data available concerning the peptides' folding thermodynamics and kinetics, which precludes a meaningful comparison with the simulation. The only qualitative indication of the peptides' native-state stability comes from the ranking (based on NMR chemical shifts) of the six peptides that Mirassou et al.<sup>22</sup> studied which—in order of decreasing stability—is the following: C1C12W3W10 > W3W10  $\gg$  C3C10 > C1C12  $\gg$  native, where the C3C10 and C1C12 are double cysteine mutants containing a disulfide bridge each. The second reason for placing the emphasis of the structures *per se* is our belief that it is indeed the ability of folding simulations to correctly identify stable structures—if they exist—that would be the most convincing indication of the maturity and accuracy of the method. Having said that, information concerning the peptides' folding thermodynamics is available directly from the trajectories and in this section we present a collection of pertinent thermodynamics quantities obtained from the mutant (2W) peptide simulation.

The whole treatment in this section follows closely the approach described by Boned, van Gunsteren, and Daura<sup>36</sup> with two significant deviations. The first concerns the ever-present question of how to select which trajectory structures should be classified as folded or unfolded. Instead of a pure RMSD cutoff, we have classified folded structures as being those that lie at the intersection of three sets. The first set was obtained through the application of an (RMSD vs native structure) cutoff which was calculated from the histogram of all RMSDs recorded in the trajectory. The second set was based on using the  $Q$ -values as criteria, implementing again a cutoff obtained from the corresponding histogram. The last set was obtained through the application of a more "fuzzy" structure-based selection as follows. Structures were classified as folded if and only if all three of the following criteria were true: (a) At least four of the residues 2–4 and 9–11 (corresponding to the two  $\beta$ -strands) were classified by STRIDE as belonging to  $\beta$  structure. (b) At least two hydrogen bonds existed between the backbone atoms of residues 2–4 and 9–11 (noting that STRIDE also uses hydrogen bonding patterns for its assignments). (c) Finally, there was a contact between the side chains of the two tryptophans, with contact being defined as a distance

of less than 5 Å between any of the non-hydrogen atoms of the two side chains. The agreement between the assignments obtained from these RMSD-,  $Q$ -, and structure-based criteria was of the order of 90% which also allowed us to obtain an estimated standard error for the derived (from the trajectory) frequencies of the folded and unfolded populations.

The second important deviation from the Boned et al.<sup>36</sup> approach arises from differences in the simulation protocol *per se*. For our simulations, we have used adaptive tempering<sup>25</sup> which is formally equivalent to a reweighted single-copy replica exchange folding simulation with a continuous temperature range. The continuity of temperature has two consequences. The first is that it is necessary to establish that the trajectory-derived thermodynamic quantities do converge for a sufficiently small temperature integration interval. That this is indeed the case is shown in Figure 1 of the Supporting Information, which uses the estimated folding enthalpy  $\Delta H_{\text{Folded-Unfolded}}$  (denoted  $\Delta H_{\text{FU}}$  hereafter) to demonstrate that convergence is achieved once the temperature integration interval reaches a value of approximately 3 K or less. The second consequence of having a continuous temperature space (and with such a small temperature integration interval) is that the statistical noise per data point increases significantly, but simultaneously, the sampling along temperature is much finer. We believe that this can be turned to an advantage, since it allows both the smooth fitting of the primary data and the estimation of standard errors. This is shown clearly in the middle graph of Figure 7. In this diagram, the scatter plot (red circles) is the primary  $\Delta H_{\text{FU}}$  data as obtained from the trajectory using temperature integration intervals of 2 and 3 K plus their offsets (the scatter plot is identical with the five upper curves shown in Figure 1 in the Supporting Information). The black line in this same diagram is the corresponding local polynomial regression fit of the primary data (obtained via  $R$ 's loess function), and the shaded area is the standard error obtained from the regression. The availability of smooth and almost continuous estimates for the thermodynamic quantities greatly simplifies their integration and differentiation, as well as the propagation of their errors. We should note here that the sampling of the high temperature regime (higher than  $\sim 420$  K) is rather incomplete in our simulation, as testified by the much higher noise level seen in the scatter plot of the middle diagram of Figure 7. Although we have decided, for completeness, to keep all data for the calculations reported below, the derived quantities for temperatures higher than  $\sim 420$  K should probably be considered unreliable. Finally, it should be mentioned here that no additional histogram-based reweighting is needed for simulations performed with adaptive tempering. As demonstrated by Zhang and Ma<sup>25</sup> using an Ising model and a Lennard-Jones system, adaptive tempering allows the correct calculation of thermodynamic quantities for the entire temperature spectrum without additional reweighting.

We start the presentation of the results by showing in the upper diagram of Figure 7 the variation of the folding free enthalpy  $\Delta G_{\text{FU}}$  vs temperature as obtained from counting statistics (relative frequencies of folded vs unfolded configurations in a given temperature interval, see Boned et al.<sup>36</sup>). The shaded area corresponds to the estimated standard error and was obtained by assigning an uncertainty equal to 0.10 to each of the observed frequencies (see second paragraph of this section). The derived sigmoidal curve appears to be fairly typical for a foldable peptide, but what is definitely untypical is how stable this mutant peptide appears to be: with a more or



**Figure 7.** Folding thermodynamics. From top to bottom, these graphs show the variation of the folding free enthalpy as a function of temperature (upper graph), of the folding enthalpy as a function of temperature (middle graph), and, finally, the variation of the folding heat capacity per Kelvin as a function of temperature (lower graph). See section 3.5 for details and an extensive discussion.

less constant  $\Delta G_{\text{FU}}$  of approximately  $-4.1$  kJ/mol for temperatures up to  $\sim 340$  K and an estimated melting temperature of  $\sim 390$  K, the mutant peptide would appear to be structurally more persistent than some very stable thermophilic proteins. This is more so if we consider the

rather strict selection criteria we applied in classifying structures as folded (discussed in the second paragraph of this section). In the absence of direct experimental evidence, it is clearly impossible to estimate how much force-field-dependent the observed stabilization of the native state is. It should be noted, however, that Mirassou et al.<sup>22</sup> showed that this mutant is significantly more stable than two mutants (C3C10 and C1C12) in which the two  $\beta$ -strands were covalently bonded through a disulfide bridge. Additionally, the force-field-dependent stabilization—if present—would clearly not be in the form of a secondary structure bias of the force field, since the native peptide is under identical simulation parameters almost completely disordered.

The middle diagram in Figure 7 shows the variation of  $\Delta H_{\text{FU}}$  as a function of temperature. The scatter plot (red circles) corresponds to the primary data as calculated from the trajectory's total energies,<sup>36</sup> the solid black line is the local polynomial regression fit, and the shaded area is the estimated standard error of the regression. The important observation here is that enthalpy appears to favor folding at higher temperatures, with an almost linear relationship between  $\Delta H_{\text{FU}}$  and  $T$  for all temperatures up to  $\sim 370$  K. What this finding implies is that for temperatures below the peptide's melting temperature the loss of enthalpy with increasing temperature is faster for the unfolded state than for the folded state.<sup>36</sup> The observation that enthalpy promotes folding at higher temperatures may sound surprising at first, but as it turns out, the behavior of  $\Delta H_{\text{FU}}$  as observed in Figure 7 has also been observed and is in excellent agreement with the findings of Boned et al.<sup>36</sup> on the folding of a completely unrelated helical  $\beta$ -heptapeptide in methanol (and with a completely different force field and simulation protocol). Whether this behavior may represent a consistent trend for fast and stably folded peptides remains to be seen.

The last graph in Figure 7 shows the variation of the folding heat capacity per Kelvin as a function of temperature. The standard errors shown in this diagram were estimated through a direct local polynomial regression fit of the primary data. The diagram clearly indicates what was already expected given the behavior of enthalpy: the entropy appears to be more favorable to folding at lower temperatures and to dominate the peptide's folding process. This, naturally, is also in full agreement with the findings of Boned et al.<sup>36</sup> Note also how the folding heat capacity per Kelvin reaches zero simultaneously with the stabilization of the values of the folding enthalpy at  $\sim 440$  K. Having obtained folding free enthalpies both through counting statistics (upper diagram of Figure 7) and from the trajectory derived total energies (in the form of differences), we can compare the corresponding estimates as an internal consistency check. The value of  $\Delta\Delta G_{\text{FU}}$  between 390 and 320 K was found to be 4.1 kJ/mol based on frequencies and 5.2 kJ/mol based on total energies, dropping to 1.9 and 2.2 kJ/mol, respectively, when the temperatures 370 and 340 K were considered. The good agreement and consistency between these estimates justifies, we believe, our choice to apply the Boned et al.<sup>36</sup> approach to a trajectory obtained via the adaptive tempering method.

#### 4. DISCUSSION

The take-home message of this communication is clear: folding molecular dynamics simulations were so accurate that they could have successfully substituted the experiment in characterizing the stability and structure of the two peptides studied

here. The simulations correctly identified which peptide is disordered and which one has a stable and persistent structure, and they correctly identified both the general characteristics of the major conformer as well as its detailed and uncommon structural properties (i.e., that it is an irregular  $\beta$ -hairpin with a non-glycine  $\beta$ -bulge, an edge-to-face stacking of the tryptophans' indole rings, and unusual dihedral angles for the Q8 residue). Additionally, the simulations indicated that the disordered (native) peptide does take up a irregular  $\beta$ -hairpin structure, and that this structure is more similar to the conformation observed in the protein's crystal structure than the structure observed in the mutant's NMR structure (note, however, that there is no direct experimental evidence that this prediction is indeed correct). Clearly, and for the case examined here, molecular dynamics performed as a robust, dependable, and sensitive method for characterizing peptide structure and dynamics. This is more so if we recall that the two peptides studied here share 83% sequence identity.

Having said that, it is this large difference in the behavior of the two peptides that also constitutes the major weakness of this work: the peptides we have selected to study have very pronounced differences in their folding behavior, so pronounced that their comparison may not provide the sensitivity needed for further development of the force fields (note, however, that this statement makes the tacit assumption that other popular empirical force fields would also reproduce the experimental findings at the level of accuracy afforded by the AMBER99SB-ILDN force field, an assumption that may not hold true).

In summary, we have added one more example to the growing gallery of very successful applications of folding molecular dynamics simulations to predict peptide structure and dynamics. The combination of the AMBER99SB-ILDN (or STAR-ILDN) force field with the TIP3P water model and full electrostatics appears to be one of the best combinations currently available for studying peptide structure.

#### ■ ASSOCIATED CONTENT

##### 📄 Supporting Information

Figure showing the variation of folding enthalpy as a function of the temperature integration interval. This material is available free of charge via the Internet at <http://pubs.acs.org>.

#### ■ AUTHOR INFORMATION

##### Corresponding Author

\*Phone: +30-25510-30620. Fax: +30-25510-30620. Home page: <http://utopia.duth.gr/~glykos/>. E-mail: [glykos@mbg.duth.gr](mailto:glykos@mbg.duth.gr).

##### Notes

The authors declare no competing financial interest.

#### ■ REFERENCES

- (1) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y.; Wriggers, W. Atomic-level Characterization of the Structural Dynamics of Proteins. *Science* **2010**, *330*, 341–346.
- (2) Gnanakaran, S.; García, A. E. Validation of an All-Atom Protein Force Field: From Dipeptides to Larger Peptides. *J. Phys. Chem. B* **2003**, *107*, 12555–12557.
- (3) Dagget, V.; Kollman, P. A.; Kuntz, I. D. A Molecular Dynamics Simulation of Polyalanine: An Analysis of Equilibrium Motions and Helix-Coil Transitions. *Biopolymers* **1991**, *31*, 1115–1134.

- (4) Gnanakaran, S.; García, A. E. Helix-Coil Transition of Alanine Peptides in Water: Force Field Dependence on the Folded and Unfolded Structures. *Proteins* **2005**, *59*, 773–782.
- (5) Best, R. B.; Hummer, G. Optimized Molecular Dynamics Force Fields Applied to the Helix-Coil Transition of Polypeptides. *J. Phys. Chem. B* **2009**, *113*, 9004–9015.
- (6) Matthes, D.; de Groot, B. L. Secondary Structure Propensities in Peptide Folding Simulations: A Systematic Comparison of Molecular Mechanics Interaction Schemes. *Biophys. J.* **2009**, *97*, 599–608.
- (7) Lindorff-Larsen, K.; Maragakis, P.; Piana, S.; Eastwood, M. P.; Dror, R. O.; Shaw, D. E. Systematic Validation of Protein Force Fields Against Experimental Data. *PLoS One* **2012**, *7*, e2131.
- (8) Piana, S.; Klepeis, J. L.; Shaw, D. E. Assessing the Accuracy of Physical Models Used in Protein-folding Simulations: Quantitative Evidence from Long Molecular Dynamics Simulations. *Curr. Opin. Struct. Biol.* **2014**, *24*, 98–105.
- (9) Kaur, H.; Sasidhar, Y. U. For the Sequence YKGQ, the Turn and Extended Conformational Forms Are Separated by Small Barriers and the Turn Propensity Persists Even at High Temperatures: Implications for Protein Folding. *J. Phys. Chem. B* **2012**, *116*, 3850–3860.
- (10) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. How Fast-folding Proteins Fold. *Science* **2011**, *334*, 517–520.
- (11) Best, R. B. Atomistic Molecular Simulations of Protein Folding. *Curr. Opin. Struct. Biol.* **2012**, *22*, 52–61.
- (12) Patapati, K. K.; Glykos, N. M. Three Force Fields' Views of the 310 Helix. *Biophys. J.* **2011**, *101*, 1766–1771.
- (13) Georgoulia, P. S.; Glykos, N. M. Using J-Coupling Constants for Force Field Validation: Application to Hepta-alanine. *J. Phys. Chem. B* **2011**, *115*, 15221–15227.
- (14) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins* **2006**, *65*, 712–725.
- (15) Wickstrom, L.; Okur, A.; Simmerling, C. Evaluating the Performance of the ff99SB Force Field Based on NMR Scalar Coupling Data. *Biophys. J.* **2009**, *97*, 853–856.
- (16) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved Side-chain Torsion Potentials for the Amber ff99SB Protein Force Field. *Proteins* **2010**, *78*, 1950–1958.
- (17) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (18) Koehl, P. Electrostatics Calculations: Latest Methodological Advances. *Curr. Opin. Struct. Biol.* **2006**, *16*, 142–151.
- (19) Georgoulia, P. S.; Glykos, N. M. On the Foldability of Tryptophan-Containing Tetra- and Pentapeptides: An Exhaustive Molecular Dynamics Study. *J. Phys. Chem. B* **2013**, *117*, 5522–5532.
- (20) Patapati, K. K.; Glykos, N. M. Order through Disorder: Hyper-Mobile C-Terminal Residues Stabilize the Folded State of a Helical Peptide. A Molecular Dynamics Study. *PLoS One* **2010**, *5*, e15290.
- (21) Patmanidis, I.; Glykos, N. M. 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. *J. Mol. Graphics Modell.* **2013**, *41*, 68–71.
- (22) Mirassou, Y.; Santiveri, C. M.; Pérez de Vega, M. J.; González-Muñiz, R.; Jiménez, M. A. Disulfide Bonds versus TrpTrp Pairs in Irregular Beta-hairpins: NMR Structure of Vammin Loop 3-derived Peptides as a Case Study. *ChemBioChem* **2009**, *10*, 902–910.
- (23) Hollingsworth, S. A.; Karplus, P. A. A Fresh Look at the Ramachandran Plot and the Occurrence of Standard Structures in Proteins. *Biomol. Concepts* **2010**, *1*, 271–283.
- (24) Case, D. A.; Cheatham, T. E., III; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber Biomolecular Simulation Programs. *J. Comput. Chem.* **2005**, *26*, 1668–1688.
- (25) Zhang, C.; Ma, J. Enhanced Sampling and Applications in Protein Folding in Explicit Solvent. *J. Chem. Phys.* **2010**, *132*, 244101.
- (26) Kale, L.; Skeel, R.; Bhandarkar, M.; Brunner, R.; Gursoy, A.; Krawetz, N.; Phillips, J.; Shinozaki, A.; Varadarajan, K.; Schulten, K. NAMD2: Greater Scalability for Parallel Molecular Dynamics. *J. Comput. Phys.* **1999**, *151*, 283–312.
- (27) Glykos, N. M. CARMA: A Molecular Dynamics Analysis Program. *J. Comput. Chem.* **2006**, *27*, 1765–1768.
- (28) Koukos, P. I.; Glykos, N. M. gcarma: A Fully Automated Task-Oriented Interface for the Analysis of Molecular Dynamics Trajectories. *J. Comput. Chem.* **2013**, *34*, 2310–2312.
- (29) Frishman, D.; Argos, P. Knowledge-Based Protein Secondary Structure Assignment. *Proteins* **1995**, *23*, 566–579.
- (30) Humphrey, W.; Dalke, A.; Schulten, K. VMD—Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33–38.
- (31) Merritt, E. A.; Bacon, D. J. Raster3D Photorealistic Molecular Graphics. *Methods Enzymol.* **1997**, *277*, 505–524.
- (32) Koukos, P. I.; Glykos, N. M. On the Application of Good-Turing Statistics to Quantify Convergence of Biomolecular Simulations. *J. Chem. Inf. Model.* **2014**, *54*, 209–217.
- (33) Cho, S. S.; Levy, Y.; Wolynes, P. G. P versus Q: Structural reaction coordinates capture protein folding on smooth landscapes. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 586–591.
- (34) Mu, Y.; Nguyen, P. H.; Stock, G. Energy Landscape of a Small Peptide Revealed by Dihedral Angle Principal Component Analysis. *Proteins* **2005**, *58*, 45–52.
- (35) Altis, A.; Nquyen, P. H.; Hegger, R.; Stock, G. Dihedral Angle Principal Component Analysis of Molecular Dynamics Simulations. *J. Chem. Phys.* **2007**, *126*, 244111.
- (36) Boned, R.; van Gunsteren, W. F.; Daura, X. Estimating the Temperature Dependence of Peptide Folding Entropies and Free Enthalpies from Total Energies in Molecular Dynamics Simulations. *Chem.—Eur. J.* **2008**, *14*, 5039–5046.