THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

Cite This: J. Phys. Chem. B 2018, 122, 106−116

pubs.acs.org/JPCB

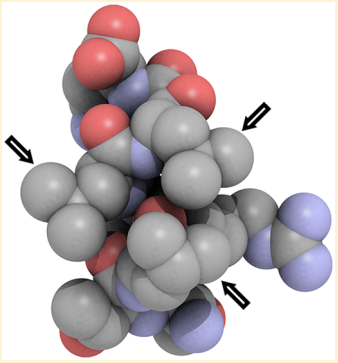
Folding Simulations of a Nuclear Receptor Box-Containing Peptide Demonstrate the Structural Persistence of the LxxLL Motif Even in the Absence of Its Cognate Receptor

Triantafyllia Adamidou, Konstantina-Olympia Arvaniti, and Nicholas M. Glykos*

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

Supporting Information

ABSTRACT: Regulation of nuclear receptors by their coactivators involves the recognition and binding of a specific sequence motif contained in the coactivator sequence. This motif is known as the nuclear receptor (NR) box and contains a conserved LxxLL subsequence, where L is leucine and x is any amino acid residue. Crystallographic studies have shown that the LxxLL motifs adopt an α-helical conformation when bound to their cognate nuclear receptors. Here we use an extensive set of folding molecular dynamics simulations to examine whether the α-helical conformation demonstrated by the LxxLL motifs in the bound state may represent a persistent structural preference of these peptides even in the absence of their cognate receptors. To this end, we have performed a grand total of 35 μs of adaptive tempering folding simulations of an NR-box-containing peptide derived from Drosophila's *fushi tarazu* segmentation gene product. Our simulations—performed using full electrostatics and an explicit representation of two different solvents (water and a TFE/water mixture)—clearly indicate the presence of a persistent helical preference of the LxxLL motif with a concomitant native-like structure and contacts between the motif's leucine residues. To lend further support to our findings, we compare the simulation-derived peptide dynamics with experimental NMR-derived nuclear Overhauser effect (NOE) measurements that had been previously obtained for the same peptide in the same two solvents. The comparison demonstrates a quantitative agreement between simulation and experiment with average upper bound NOE violations of less than 0.084 Å, thus independently validating our main conclusion concerning the intrinsic preference of NR-box motifs to form helical structures even in the absence of their cognate receptors.



1. INTRODUCTION

The LxxLL motif[1] is an extensively characterized sequence motif which participates in numerous cases of protein−protein interactions. The majority of these LxxLL-mediated interactions involves associations between transcription factors and coactivator proteins.[2] The prominent structural motif observed in these associations—as seen in several crystallographic structures of protein/LxxLL motif complexes[1]—involves the presence of an α-helical structure for the LxxLL motif.[1−4] This raises the question of whether the helical structure of the motif is mostly due to (and arising from) the energetics and structural properties of the formation of these protein complexes, or is a manifestation of an underlying helical preference of the LxxLL motif.

Here we attempt to establish an answer to this question through the application of extensive (35 μs long) folding simulations of a member of the LxxLL motif family, the nuclear receptor (NR)-box peptide derived from Drosophila's *fushi tarazu* segmentation gene product. The aim of performing a folding simulation of an NR-box peptide in solution is to establish whether even the isolated peptide shows a clear preference for adopting helical structures. Such a preference—if it indeed existed—would indicate the presence of a structurally persistent bias of the LxxLL motifs toward their native (complex-like) structure even in the absence of their respective receptors and would place the observed mode of association in a structurally (and thermodynamically) firmer ground.

The peptide that we have chosen to simulate is the NR-box sequence of Drosophila's *fushi tarazu* segmentation gene product corresponding to residues 102−120 of the full length protein (sequence VEERPST**LRALL**TNPVKKL, UniProt entry P02835). This LxxLL-containing peptide is involved in the interaction between Drosophila's orphan nuclear receptor *fushi tarazu* factor 1 (FTZ-F1) and the *fushi tarazu* segmentation gene product (FTZ).[5,6] This interaction is necessary for the regulation of genes that define the alternating segments in Drosophila's embryo.[5,7] The detailed structural basis of this interaction has been recently elucidated through the crystallographic structure determination of the FTZ-F1/FTZ complex.[8]

Having outlined the interesting biology in which this NR-box is involved, we should note that the real reason for our choice to use the specific LxxLL-containing peptide was rather

Received: October 17, 2017
Revised: November 22, 2017
Published: December 6, 2017

DOI: 10.1021/acs.jpcb.7b10292
J. Phys. Chem. B 2018, 122, 106−116

mundane and was based on the availability of experimentally derived NOE measurements for the free peptide in two different solvents [water and a 50% (v/v) TFE/water mixture].[9] We perceived the availability of published NOE data as an opportunity to not only strengthen the conclusions drawn from the simulation but also independently validate our choice of force field and simulation protocol as described in the next section (please note that, although the NOE data for this peptide have been published in printed form,[9] no associated structural models have been deposited with either the PDB or BMRB databases).

In the following paragraphs, we describe the simulation protocol and the methods used for analyzing it, we examine the issue of how close are our simulations to sufficiently sampling the folding landscape of the peptide, and then we proceed with the analysis of the simulation results with emphasis on the secondary structure preferences of the peptide, the structural stability of the LxxLL motif, and, finally, the agreement between the simulation and the experimental NOE measurements. We close by revisiting the question of the influence of pre-existing structural preferences in the context of a protein−protein complex formation.

2. METHODS

2.1. Peptide Sequence and Residue Numbering. The peptide we simulated (hereafter referred to as **ftz**) is the NR-box sequence of Drosophila's *fushi tarazu* segmentation gene product corresponding to residues 102−120 (inclusive) of the full length protein (UniProt entry P02835). The full peptide sequence (with the LxxLL motif underlined and highlighted) is VEERPST<u>LRALL</u>TNPVKKL. To maintain consistency with the literature,[1,2] the first leucine of the **L**xx**LL** motif will be referred to as residue number **+1**, the next two ("xx") residues are numbered as **+2** and **+3**, the two C-terminal leucines of the Lxx**LL** motif are **+4** and **+5**, etc. The residues preceding the first leucine of the motif are numbered with negative residue numbers. For example, the threonine immediately before the first leucine is residue number **−1**, the preceding serine is **−2**, etc. With this numbering scheme, the peptide starts at residue **−7** and ends with residue **+12**.

2.2. System Preparation and Simulation Protocol. The preparation of the system including the starting peptide structure (in the fully extended state) together with its solvation and ionization were all performed with the program LEAP from the AMBER tools distribution.[10] For both simulations, we have used periodic boundary conditions with a cubic unit cell sufficiently large to guarantee a minimum separation between the symmetry-related images of the peptides of at least 16 Å. We followed the dynamics of the peptide's folding simulations using the program NAMD[11] for a grand total of 35.5 μs using the TIP3P water model,[12] the TFE parametrization from the R.E.D. library,[13−15] and the AMBER99SB-STAR-ILDN force field[16−19] which has repeatedly been shown to correctly fold numerous peptides.[20−32] Due to the significantly slower dynamics of the mixed TFE/water system (see section 2.4), the simulation lengths were markedly unequal: 8.9 μs for the pure water simulation, 26.6 μs for the TFE/water simulation. The proportion of TFE and water molecules needed to reproduce a 50% (v/v) mixture was determined as previously described.[13] For both simulations, adaptive tempering[33] was applied 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 280−380 K inclusive and was applied to the system through the Langevin thermostat; see below.

The simulation protocol has been previously described[30−32] and in summary was the following:

"The systems were first energy minimized for 1000 conjugate gradient steps followed by a slow heating-up phase to a temperature of 320 K (with a temperature step of 20 K) over a period of 32 ps. Subsequently, the systems were equilibrated for 10 ps under NpT conditions without any restraints, until the volume equilibrated. This was followed by the production NpT runs with the temperature and pressure controlled using the Nosè−Hoover Langevin dynamics and 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^{-1}, and the piston's oscillation period to 200 fs, with a decay time of 100 fs. The production runs were performed with the impulse Verlet-I multiple time step integration algorithm as implemented by NAMD. The inner time step was 2 fs for the water simulation and 2.5 fs for the TFE/water simulation, with short-range nonbonded interactions being 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^{-6}. A cutoff for the van der Waals interactions was applied at 9 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 systems every 0.8 and 1.0 ps for the water and TFE/water simulations, respectively.

2.3. Trajectory Analysis. The analysis of the trajectories was performed as previously described:[30−32] "The programs CARMA,[34] GRCARMA,[35] and Cluster5D[36] 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. Chemical shifts were calculated using the program SPARTA+[37] as previously described.[32] Secondary structure assignments were calculated with the program STRIDE.[38] All molecular graphics work and figure preparation were performed with the programs VMD,[39] RASTER3D,[40] PyMol,[41] and CARMA." Estimated NOE values were obtained from the trajectories using $\langle r^{-6}\rangle$-based averaging where r is the instantaneous distance between two selected protons from a trajectory.

2.4. Extent of Sampling and Statistical Significance. As will be discussed extensively in the next section, the ftz peptide is not a stable folder. During the length of our simulations, it visits and interconverts between numerous conformations. The result is that any systematic structural preference of the peptide is mostly statistical in nature and can only be deduced by studying a sufficiently long sample of the peptide's folding dynamics. This takes us back to the ever-present problem of estimating statistical significance for the results obtained from a set of folding simulations for which a

DOI: 10.1021/acs.jpcb.7b10292
J. Phys. Chem. B 2018, 122, 106−116

complete and faithful sampling of the folding landscape is impossible for any finite amount of simulation. We approach this problem by applying a recently described probabilistic method for estimating the convergence of molecular dynamics trajectories.

The method[42] is based on the application of Good–Turing statistics to estimate the extent of structural variability that has *not* been sampled by a given simulation. This "missing" (unobserved) structural variability is codified in the form of graphs of the form "probability of unobserved structures" vs "minimal RMSD of these structures from any of the conformations that have already been observed in the simulation". The general form of these diagrams can be seen in Figure 1: they all start with high probability values at low

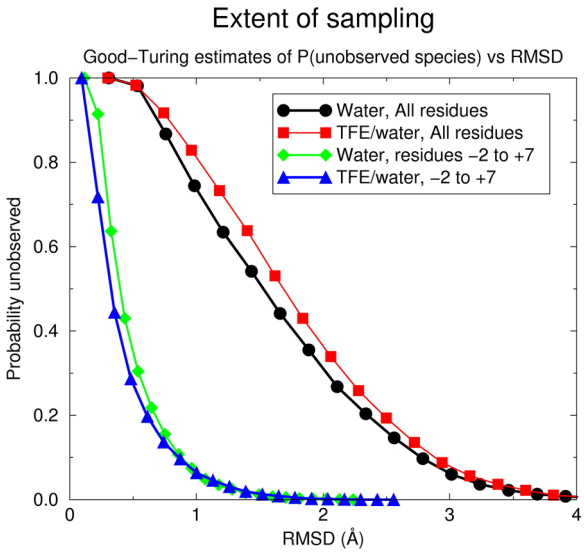


Figure 1. Extent of sampling and statistical significance. Results from the application of Good–Turing statistics to the two peptide trajectories (in water and TFE/water) for all residues (upper two curves) and for residues −2 to +7 (lower two curves). See text for details.

RMSDs and asymptotically approach low probability values for high RMSD values. The high probability values for small RMSDs signify the fact that it is very probable to observe a structure that, although different in detail, is nevertheless very similar to one of those already observed (for example, if the simulation were to last for just one additional time step, then a new structure would be observed which would be nearly—but not fully—identical with the last structure of the recorded trajectory). As the RMSDs from the already recorded structures increase, the corresponding probability values decrease. It is the exact form of these curves and how fast they approach low probability values that inform us about how significant—or otherwise—is the structural variability that we have missed due to limited sampling.

Figure 1 shows results obtained from the application of Good–Turing statistics to the two trajectories studied for the ftz peptide (in water and TFE/water respectively, see section 2.2). The results shown have been obtained from four independent calculations and are organized in two groups. The first group (upper two curves in Figure 1, colored red and black) show results from the direct application of the method to the two trajectories using the C^{α} atoms of all residues of the peptide.

Both curves (red and black in Figure 1) show the same general form, and they exhibit quite large RMSD values even for non-negligible probabilities. For example, the probabilities corresponding to an RMSD of 3 Å are of the order of 0.10. This implies that if we continued the simulations we would expect approximately 1/10th of the new (previously unobserved) structures to differ by an RMSD of at least 3 Å from the structures already observed. The probabilities maintain significant values even to very large RMSD values of the order of approximately 5 Å. The take-home message from these two curves is clear: after a total of 35 μs of simulation time, significant structural variability remains unobserved, and the sampling of the peptide's folding landscape is nowhere near convergence. As already discussed, this is not surprising: the ftz peptide is not a fast and stable folder, and thus, the full complexity arising from the dynamics of a 19-residue-long peptide is accessible to the simulations. As a side note, notice also how the curve for the TFE/water solvent is slightly translated toward higher RMSD values, although the TFE/water simulation lasted almost 3 times longer than the simulation in water (26.6 μs vs 8.9 μs). This is an indirect but clear demonstration of the much slower dynamics observed in the complex solvent environment.

The lower two curves (green and blue) in Figure 1 show the results obtained from the application of Good–Turing estimation to a subset of residues centered on the LxxLL motif (residues −2 to +7 inclusive). Clearly, the effect of limiting the residue selection to the amino acids encompassing and surrounding the LxxLL motif is rather dramatic: both curves fall quite rapidly to small probability values, reaching negligibly small values for RMSDs of the order of 2 Å. Repeating the calculation using only atoms of the motif *per se* (residues +1 to +5) showed that negligibly small probability values are reached for RMSDs as low as 0.8 Å, demonstrating again the much better sampling afforded by the simulations for the motif-containing part of the peptide. Notice also how the curve for the TFE/water simulation is now shifted to *lower* RMSD values compared with the water curve (which is the inverse of the behavior observed for the whole peptide). As will be discussed in the next section, this is the result of the significant stabilization of secondary structure induced by the presence of TFE.

Taken together, the results from the Good–Turing calculations paint the following picture: the dynamics of the peptide as a whole have not been adequately sampled, and there is significant structural variability that has not been observed in the existing simulations. However, the dynamics and structural variability of the part of the peptide centered on the LxxLL motif appear to have been sufficiently sampled with expected maximal RMSDs (of unobserved structures) ranging between 0.8 and 2 Å depending on how tight the residue selection is around the motif. This analysis already suggests that the N- and C-terminal residues are much more flexible than the central part of the peptide, as will be discussed later.

Although this section is already long enough, we would like to present one additional analysis concerning the extent and statistical significance of our simulations. The analysis that follows is again an effort to estimate the amount of structural variability that has not been observed in the simulations, but this time using as a measure the uncertainty in the peptide's secondary structure assignments. The procedure we followed to tackle the problem is outlined below.

DOI: 10.1021/acs.jpcb.7b10292
J. Phys. Chem. B 2018, 122, 106–116

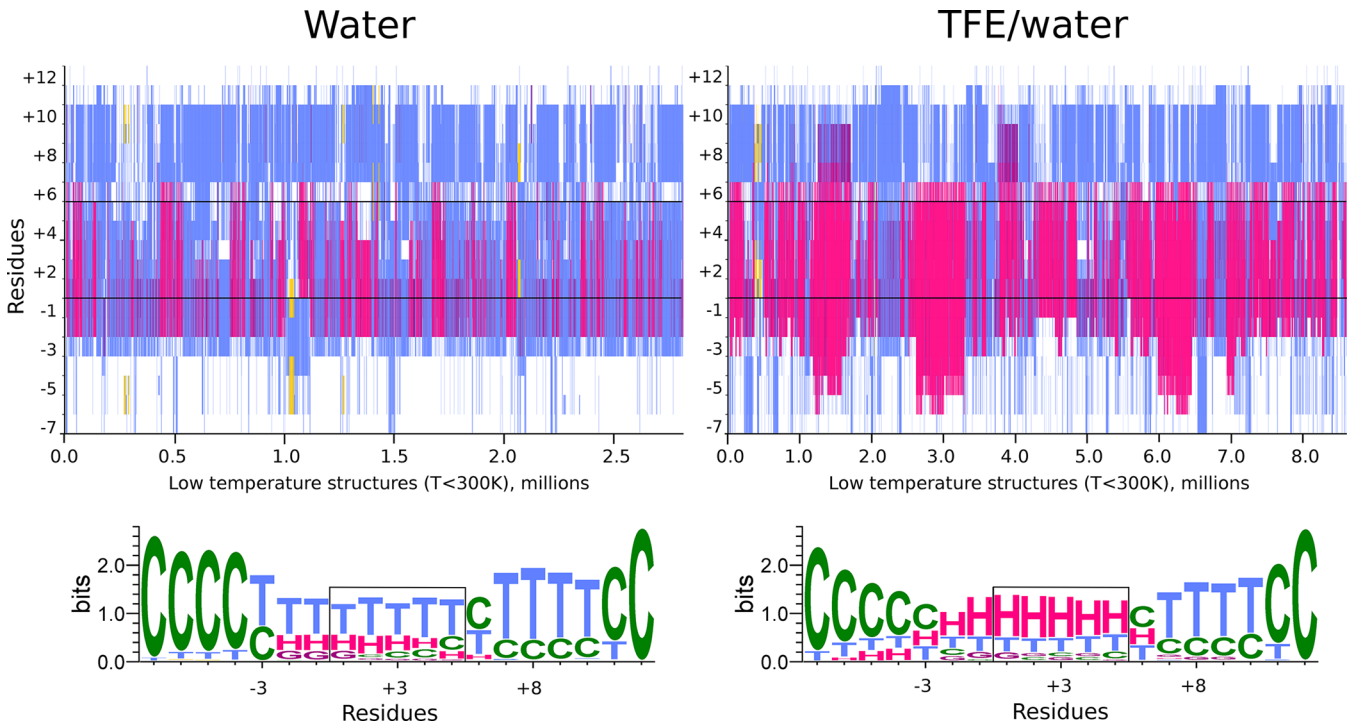


Figure 2. Secondary structure analysis overview. Side-by-side comparison between the peptide's secondary structure properties in water (left column) and TFE/water mixture (right column). For both solvents, the upper panel shows the evolution of the per-residue STRIDE-derived assignments for all structures from the trajectory whose corresponding adaptive tempering temperature was less than 300 K (2.8 and 8.6 million structures from the water and TFE/water simulations; these correspond to approximately 30% of the total number of structures observed in each of the respective trajectories). The color coding is red/magenta → $\alpha/3_{10}$ helical structure, yellow → β-structure, cyan → turns, white → coil. The lower panels show the weblogo[44]-derived distributions corresponding to the STRIDE-derived assignments with H → α helix, G → 3_{10} helix, B → β structure, T → turn, C → coil. The color coding is identical with the above except for coil which is depicted in green. In all four diagrams, thin black lines enclose the limits of the LxxLL motif.

The sole source of information for the Good–Turing estimation is a distance matrix describing relations between discrete samples taken from a trajectory. In the case of the RMSD-based analysis discussed above, the distance matrix is the trivially calculated RMSD matrix. However, *any* suitable metric can be used to construct the distance matrix, and thus, any quantity that can expressed as a distance in a metric space can be used in the analysis. What we want to estimate is the uncertainty of the peptide's secondary structure assignments (as obtained, for example, from STRIDE[38]). This could tell us how probable—or otherwise—it is to observe a completely new secondary structure motif for the ftz peptide. The problem is that secondary structure assignments are character-based entities and not numerical quantities, which makes the construction of a meaningful distance matrix not trivial. The solution we employed is based on using a modified form of the Hamming distance[43] between any two sets of secondary structure assignments as a distance metric (the Hamming distance is equal to the number of substitutions needed to convert one string to the other). The modification we applied is an empirical attempt to correct for obvious problems with the pure (unmodified) Hamming distance. For example, a transition between an α-helix and a 3_{10}-helix would normally be assigned to the same distance (equal to 1.0) as a transition between an α-helix and β-structure. We corrected this by assigning a distance of 0.35 (instead of 1.0) for transitions involving helical states. Or, for another example, the distance between two residues that have both been assigned to the coil or turn state would be zero, although both of these states are highly heterogeneous in nature. We also corrected for this by assigning a distance of 0.15 for all transitions involving a coil or turn state, even when the same residues were assigned to the same (coil or turn) state. The results from this calculation are shown in the Supporting Information Figure S1 using the same set of four graphs as in Figure 1. In general, there is good agreement between the conclusions drawn from the RMSD matrices and the secondary structure analysis: the sampling is significantly better for the residues surrounding the LxxLL motif with the probabilities dropping to less than 10^{-5} for the possibility of unobserved secondary structure changes involving more than one residue. The only consistent difference between the two analyses is that when using the secondary structure criterion the TFE/water mixture simulation demonstrates a better sampling than water even when using all peptide residues. This is the result of the significant secondary structure stabilization induced by TFE, as will be discussed below.

3. RESULTS

3.1. The ftz Peptide in Water Is Highly Flexible but with a Detectable Helical Preference. The two graphs on the left column of Figure 2 depict results from the secondary structure analysis of the ftz peptide in water. The upper diagram shows the evolution of secondary structure (for each residue) as a function of peptide structures recorded from the trajectory. Note that for this analysis we have only used the peptide structures that correspond to stable (from the simulation's point of view) conformers. This is feasible because

DOI: 10.1021/acs.jpcb.7b10292
J. Phys. Chem. B 2018, 122, 106–116

the simulations were performed using adaptive tempering[33] which automatically adjusts the thermostat depending on the energy of the system. Having said that, no appreciable differences were observed when using for the analysis the whole trajectory. Even a cursory examination of this diagram shows that the ftz peptide in water is highly flexible with the majority of residues being assigned to coil (white) or turn (cyan) states. This is especially true for the first four (−7 to −4 inclusive) and the last six (+7 to +12) residues which are almost without exception disordered [but do note the systematic difference in assignments between the N-terminus (mostly coil) and the C-terminus (mostly turns) which implies a statistical preference for more compact structures at the C-terminus]. Assignments to β-structure are very rare, with only some very minor occurrences at approximately 0.30, 1.05, and 2.10 million structures.

Residues −2 to +5 (and to a lesser extent +6) show a different behavior: they appear to spend an appreciable amount of simulation time visiting α-helical and 3_{10}-helical states, as can be seen more clearly in the lower weblogo diagram which only shows significant helical content for residues −2 to +5. Note how the helical structure starts two residues before the LxxLL motif. We would not like to overinterpret the results, but these two residues (at positions −1 and −2) are important for the recognition of the NR-box and are used to define the classes of NR-boxes.[2] Notice also how the 3_{10}-helical state (G in the weblogo diagram) is only observed for residues −2 to +1). Unexpectedly, the presence of a 3_{10}-helical state for residues −2 to +1 appears to be anticorrelated with the existence of a helical state for any other peptide residue (including residues +2 to +5). Because this is difficult to discern from Figure 2, we compare in Figure 3 two weblogo diagrams which were calculated using two distinct subsets of the peptide structures. The first subset included only those structures for which residues −2 to +1 were assigned as 3_{10}-helical. The second subset included only those structures for which residues +2 to +5 were α-helical.

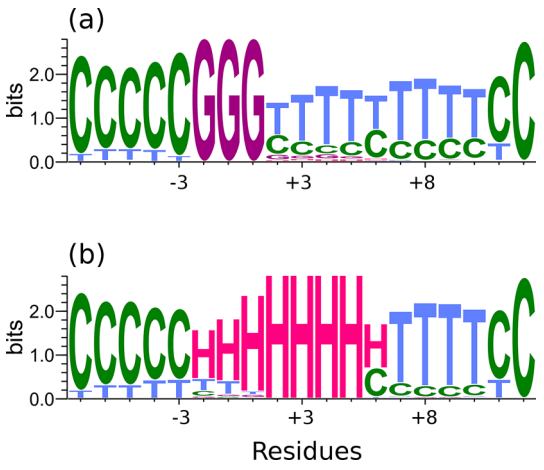


Figure 3. 3_{10}-Helix and α-helix have distinct residue preferences and are mutually exclusive. Panel a is a weblogo representation of secondary structure preferences for the peptide in water when using for analysis only structures whose residues −2 to +1 were assigned by STRIDE to a 3_{10}-helical state. Apart from the selected residues, no other detectable helical preferences are evident. Panel b shows the same analysis but this time focusing on structures for which STRIDE assigned an α-helical state to residues +2 to +5. This time a clear helical preference is evident also for residues −2, −1, +1, and +6. Notice how the two states appear to be mutually exclusive (no α-helical assignments when selecting 3_{10}-helix-containing structures, no 3_{10}-helical assignments when selecting α-helical structures).

The take-home message from the diagrams in Figure 3 is clear: when residues −2 to +1 are in the 3_{10}-helical state, the rest of the peptide is almost exclusively in either turn or coil states. Conversely, the presence of α-helical structure for residues +2 to +5 is highly correlated with the existence of an α-helical state also for residues −2, −1, +1, and +6. As will also be discussed later, this interesting finding is difficult to validate against the experiment because due to the high flexibility of the peptide in water only a handful of long-range NOEs have been reported.

Before closing this section, it is worth noting that the per-residue helical preferences obtained from the simulation (Figure 2) are in good agreement with the results obtained from sequence-based prediction algorithms. For example, the program AGADIR[45] predicts significant (∼7% at 278 K) helicity only for residues +1 to +6 inclusive. Similarly, the PSIPRED[46] server estimated significant helical content for the residues −4 to +6 inclusive.

3.2. The Presence of TFE Significantly Stabilizes the Peptide's Helical State. Before presenting the results from the TFE/water simulation, we believe that a cautionary note is in order: Proper parametrization and validation of TFE and TFE/water mixtures is an active force field development area.[13−15] Add to this the still active debate concerning the mechanism by which TFE alters biomolecular folding and dynamics,[13,14,47] and it becomes immediately obvious that the results reported below for the TFE/water simulation should be viewed more like an opportunity for additional validation of the selected force field and TFE parameters than anything else. Cautionary note aside, the analysis of the results obtained from the TFE/water simulation follows.

Comparison of the water and TFE/water diagrams in Figure 2 immediately shows that the main effect of TFE is the one already expected: the presence of the organic solvent greatly stabilizes the peptide's helical structure (but note that there are documented cases where TFE stabilized nonhelical structures[48,49]). The observed stabilization has three aspects. The first concerns the persistence of helical states as a function of simulation time. In the case of the pure water simulation, the percentage of structures with an α-helical assignment for residues −2 to +5 (inclusive) is slightly less than 10% of the trajectory. In the presence of TFE, this percentage is 4 times higher, reaching a value of approximately 42%.

The second aspect concerns the actual extent of helicity as a function of the number of residues in a helical state. Whereas in water no helical residues were observed outside the range −2 to +6, in the presence of TFE, significant excursions of helicity toward the peptide's N-terminus are observed four times (centered at 1.5, 3.0, 6.2, and 7.0 million structures in Figure 2). These intervals of extended helical structure represent a significant portion of the trajectory. To put this in numbers, we calculated the proportion of the trajectory for which all residues between −6 → +5, −5 → +5, −4 → +5, and −3 → +5 were exclusively α-helical. These percentages were found to be 3.8, 12.7, 16.7, and 25.5%, respectively.

The third aspect is that TFE appears to selectively stabilize the α-helical state but not the 3_{10}-helical. The percentage of 3_{10}-helical residues in the water trajectory is 3.5% (this percentage was calculated over the sum of all peptide residues over all recorded peptide structures). Repeating the calculation

DOI: 10.1021/acs.jpcb.7b10292
J. Phys. Chem. B 2018, 122, 106−116

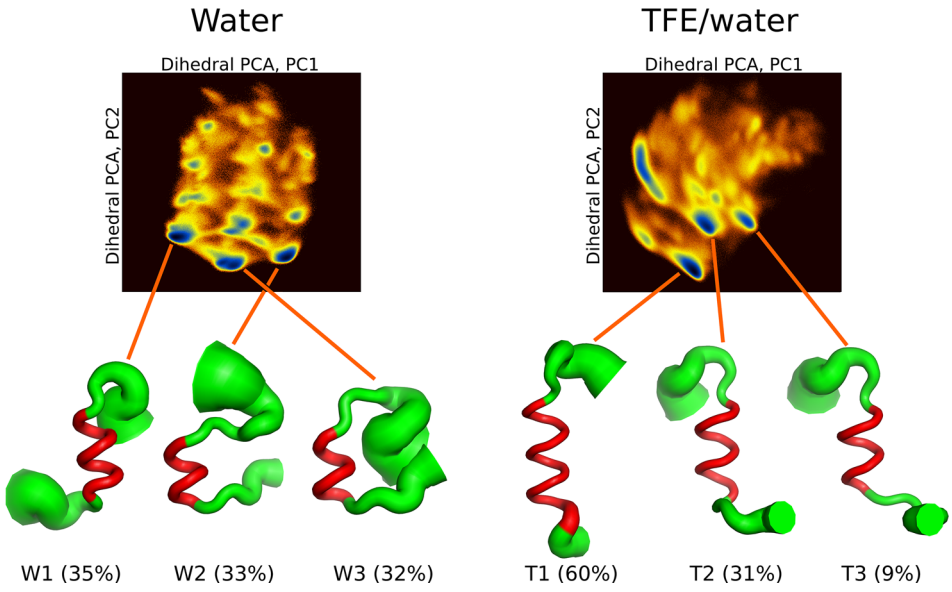


Figure 4. Dihedral principal component analysis and representative structures. The top two diagrams depict the logarithm of the density distribution of the projection of the two trajectories on the plane defined by the top two eigenvectors obtained from a dihedral principal component analysis of the corresponding simulations (blue is high density). The structure schematics below the projections are representative structures derived from the indicated peaks (clusters) of the density distributions. The radius of the tube representing the peptide's backbone is proportional to the atomic RMS fluctuations of the respective atoms calculated over all structures belonging to the same cluster. The color coding is red for helical structure, green for coil/turns/unassigned. In all diagrams, the N-terminus is toward the lower part of the figure and the structures have been vertically aligned in such a way as to place residue −2 at the same level. The percentages below the structures are the *relative* populations of the corresponding clusters.

with the TFE/water trajectory gave an essentially identical percentage of 4.0%. There is, however, one difference concerning the 3_{10}-helical structure: in the presence of TFE, α-helical and 3_{10}-helical segments can coexist in the same structure and are not mutually exclusive, as was shown for the water simulation in Figure 3. This can be seen (upper graph of Figure 2) in the 3_{10}-helical regions (colored dark magenta) of residues +7 to +9 at approximately 1.5 and 4.0 million structures. Notably, when α-helical and 3_{10}-helical segments coexist in the same structure, there is again a clear preference for their relative positions on the peptide: the α-helical region is N-terminally located and usually reaches up to residue +6, whereas the 3_{10}-helix is toward the C-terminus, usually occupying residues +7 to +9.

To summarize the main results up to now, the ftz peptide in water is mostly disordered but with a clear helical preference for residues −2 to +5. In the presence of TFE, this helical preference is becoming so pronounced that it essentially constitutes a stably folded peptide structure for residues −2 to +5 but with significant extensions of helicity toward the peptide's N-terminus.

3.3. Dihedral Principal Component Analysis Allows Visualization of the Peptide's Flexibility. Secondary structure analysis (Figure 2) established the fundamental characteristics of the two simulations and set a clear framework for the subsequent analyses. The aim of this section is to place these observations in a more structurally oriented context by identifying prominent peptide conformations for the two simulations.

Structural analysis of highly flexible systems is not trivial. A series of papers from the Stock group established that dihedral principal component analysis (dPCA) is a powerful and high resolution method for studying such systems.[50−52] We have performed dPCA on both the water and TFE/water trajectories as described in section 2.3. On the basis of the results obtained from sections 3.1 and 3.2, we have limited the dPCA analysis to only include residues −3 to +7, thus reducing the amount of noise that incorporation of disordered residues would have introduced into the calculation. The results are shown in the top row of Figure 4 in the form of the log density distributions corresponding to the top two principal components of the respective trajectories. High density (dark blue) peaks in these distributions correspond to populations of structures with similar principal component values, and thus similar dihedral angles and backbone structures. The large number of peaks in these diagrams are indicative of flexible systems that visit several distinct conformations during the simulation. As expected, the number of distinct peaks is higher for the water simulation.

Having obtained the distribution of the principal components from dPCA, it is possible to associate high density peaks from these distributions with distinct peptide conformers [but please note that the actual clustering is performed in the three- or five-dimensional principal component space,[36] and not in two dimensions (as shown here for clarity)]. The lower part of Figure 4 shows schematic (cartoon) diagrams of the representative structures obtained from the top three clusters of each trajectory (the "representative structure" is this structure from a cluster that is closest to the average of all structures that belong to the given cluster). To allow visualization of the structural variability that these conformers encompass, the radius of the tube representation used in Figure 4 is proportional to the RMS fluctuations of the respective atoms as calculated from all members of a cluster. The important thing to note here is that these representations are only suggestive and that they do not adequately convey the amount of structural variability that is actually present in the clusters, nor—especially for the less well ordered coil parts—

DOI: 10.1021/acs.jpcb.7b10292
J. Phys. Chem. B 2018, 122, 106−116

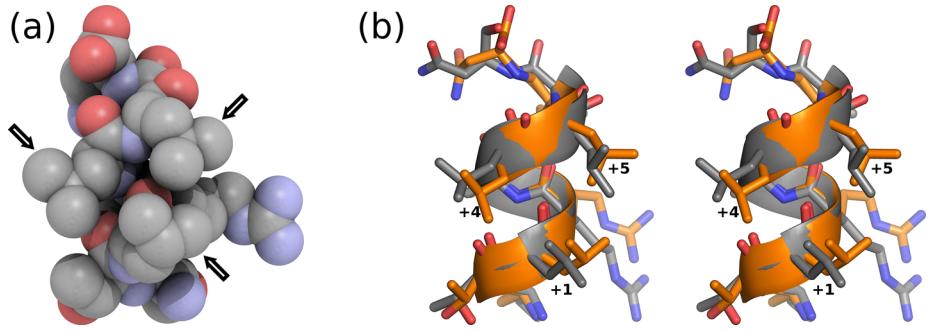


Figure 5. Experiment vs simulation, structural comparison. Panel a shows a space-filling model of residues −2 to +7 in a helical conformation demonstrating the hydrophobic patch created by the leucines of the LxxLL motif (marked with arrows). Panel b is a wall-eyed stereodiagram comparing the crystallographic structure of the ftz peptide bound to the receptor (colored gray) with the representative simulation-derived structure in water (colored orange). See text for details of how the simulation-derived structure was obtained. In both panels, the structures are on the same scale, in the same orientation (with the peptide's N-terminus toward the lower part of the figure), and leucines +1, +4, and +5 of the LxxLL motif have been labeled to aid interpretation.

the actual path (in the three-dimensional space) that the peptide backbone follows.

To make sure that this point is not missed, we show in the Supporting Information Figure S2 an image from the actual superposition of 500 structures that belong to the smallest cluster from the TFE/water simulation (marked as T3 in Figure 4). Clearly, the tube representations used here, although useful as simplifying approximations, do not adequately describe the amount of structural variability that is actually present in the dPCA-derived clusters. The take-home message from this analysis is that the structures shown in Figure 4 should not be viewed as proposed (or even, existent) "peptide structures". The ftz peptide is so flexible that only superpositions such as the one shown in the Supporting Information represent a fair depiction of the dPCA-derived clusters. Unfortunately, such superpositions are so complex and noisy that they obscure the structural context that is actually present in the clusters.

Having defined the limitations of the diagrams shown in Figure 4, it is instructive to see how these structures place the findings from the secondary structure analysis on a firm ground. As can be deduced from the large RMS fluctuations (proportional to the tubes' radii), the peptide is highly flexible in water which leads to a significantly reduced number of helical residues. These range from −2 → +6 for cluster W1 down to −2 → +4 (and even +3) for members of clusters W2 and W3. The residues before and after the helical parts are so flexible that no conclusions can safely be drawn other than the possibly more extended state of the N-terminus and the appearance of more compact structures for the C-terminal residues. We have tested this observation by calculating the value of the radius of gyration for two subsets of residues, the first −7 to −3 (inclusive) and the second +7 to +11. The results were found to be 4.38 ± 0.27 Å for the N-terminal residues and 3.80 ± 0.47 Å for the C-terminus, indicating that indeed the C-terminal residues of the peptide in water show a detectable preference for more compact structures. This is also in agreement with the (coil vs turn) secondary structure assignments of these residues, as shown in Figure 2. Comparison of the percentages shown in the lower part of Figure 4 indicates that the three water-derived clusters are almost equally populated with no significant preference for structures with higher helical content, demonstrating again the marginal stability of the helical state in water. One final observation concerns the presence of contacts between the termini of the peptide to form more compact structures. This can be deduced from the structures of clusters W2 and, especially, W3 in Figure 4. We have tested this idea by calculating (as a function of simulation time) the minimal distance observed between two distinct sets of C^{α} atoms. The first set comprised the C^{α} atoms of residues −7 to −3 (inclusive), the second set, residues +7 to +12. We define a "contact" between the termini as a minimal distance of less than 6 Å (which is a rather strict choice). With this definition, we observe contacts between the peptide's termini for approximately 22% of the recorded low temperature (less than 300 K) structures in water (this number becomes 29% with a 7 Å cutoff). Cluster and secondary structure analysis of these structures showed that approximately half of them contained helical residues, thus indicating that there is a detectable correlation between the formation of helical structures and the presence of contacts between the peptide's termini. The relatively high percentage of collapsed peptide structures with contacts between the termini may also be related to the established bias of this and related families of force fields toward compact conformations.[26−29]

Turning our attention to the results obtained from the TFE/water simulation (Figure 4, right panel), it is clear that the presence of TFE significantly simplified things from the structural point of view: not only the extent of the helices and their stability increased significantly, but also the three dPCA-derived clusters account for the greatest part of the observed variance (remembering that the PC1−PC2 diagrams in this figure are on a logarithmic scale). The major cluster (T1 in Figure 4) is not only the one with the longest continuous helical stretch (residues −6 to +7 inclusive) but is also the most stable conformation accounting for 60% of all clustered peptide structures. Even the smallest cluster (T3) is as helical as the most helical cluster from the water simulation (W1). The observation concerning the systematic difference between the extended coil-like N-terminus and the more compact C-terminus appears to hold true for the TFE/water simulation as well, but due to the stability of the helical portion of the peptide, its contribution to overall peptide structure is possibly insignificant. The possibility of contacts between the peptide termini was also examined, as described above for the water simulation. With the same residue selection and distance cutoff (6 Å), contacts were identified for only 4.3% of the low temperature structures, and thus, no further analysis was deemed necessary.

DOI: 10.1021/acs.jpcb.7b10292
J. Phys. Chem. B 2018, 122, 106−116

3.4. Persistent Structural Preferences Can Be Detected Even at the Level of Side Chains. The preceding analysis established the presence of persistent secondary structure preferences for the ftz peptide (Figure 2) and identified the backbone structures corresponding to major peptide conformers (Figure 4). The identification of prominent clusters in section 3.3 was based on a dihedral principal component analysis and, thus, completely ignored the conformation of the peptide's side chains. In this section, we take the analysis one final step further by examining the structural preferences of the side chains. Our aim here is to fully compare the structure of the peptide as seen in its complex with the cognate receptor (PDB entry 2XHS) with an all-atom representative structure obtained from the simulations. Since the aim of the analysis is to compare with the crystallographic structure, only the water simulation was examined.

To obtain an all-atom representative structure from the water simulation, we proceeded as follows. In the first step, and starting from the major dPCA-derived cluster (W1 in Figure 4), we performed a Cartesian principal component analysis (CPCA) of the respective structures but only using backbone atoms. This resulted in a new set of peptide structures with well-defined backbone conformation for residues −2 to +7. In the second stage, these structures were submitted to a second round of CPCA but this time using all non-hydrogen peptide atoms in the analysis. The representative structure selected for further analysis was the structure closest to the average structure obtained from this second stage of CPCA.

A space-filling model of the simulation-derived representative structure is shown in Figure 5a, and it clearly demonstrates the presence of the expected hydrophobic patch comprising the three leucines (+1, +4, and +5) of the LxxLL motif. Panel b in the same figure shows in the form of a wall-eyed stereodiagram a direct comparison between the crystallographic receptor-bound structure of the peptide (colored gray, PDB entry 2XHS) and the representative simulation-derived structure (colored orange). The agreement between the two structures is excellent down to the level of individual side chains. To put this in numbers, we have calculated the RMS deviation between the two structures (residues −2 to +7 inclusive). For the C^{α} atoms, we obtained an RMSD of 0.59 Å, which increased to 0.65 Å when all backbone atoms were used for the calculation. Using all non-hydrogen atoms (including all of the peptide's side chain atoms) gave an RMSD of 1.56 Å. For completeness, we have also performed the same analysis using the TFE/water trajectory with more or less identical results (C^{α} RMSD was 0.65 Å, backbone 0.65 Å, all non-hydrogen atoms 1.64 Å). Given that this is a comparison between the receptor-bound and free peptide, the structural agreement is surprisingly good. Having said that, it should not be forgotten that the peptide is highly flexible and that the simulation-derived structure shown in Figure 5 only represents about 10% of the water trajectory as discussed previously and shown in Figure 2.

3.5. Experiment and Simulation Are in Quantitative Agreement. Yun et al.[9] published—in printed form only—an NMR study of the ftz peptide in both water and 50% (v/v) TFE/water mixture. Unfortunately, neither the NMR data nor any structural models have been deposited with either the PDB or BMRB databases. We have, however, been able to pick up the NMR-derived NOE upper bounds directly from the printed form of that paper (Yun et al.,[9] Figure 2) and to compare the simulation derived $\langle r^{-6}\rangle$-based averages with the experimental upper bounds. The detailed results are shown in the Supporting Information Tables S1 and S2 for the water and TFE/water simulations, respectively. In the following paragraphs, we compare the experimental results with the simulation-derived averages in terms of their overall agreement but also by focusing on the specific peptide regions that a systematic deviation appears to be present.

Overall, the agreement between experiment and simulation is excellent. For the water simulation, the average upper bound NOE violation[53] is only 0.050 Å with 5 violations out of a total of 46 NOEs. The very small number of experimentally recorded NOEs (considering that this a 19-residue peptide) is indicative of a highly flexible system, in good qualitative agreement with the simulation results. The only systematic deviation between experiment and simulation concerns the $d_{\alpha N}(i, i+2)$ NOEs between the pairs of residues (+2, +4) and (+3, +5) which are systematically predicted weaker by the simulation. To visualize the type of peptide structures that are under-represented in the simulation, we have selected and examined all simulation-derived structures for which the $d_{\alpha N}(i, i+2)$ distances between these two pairs of residues was less than 3.3 Å. The resulting peptide structures were structurally similar, and they all showed an α-helical turn for residues −2 to +2 followed by a reverse turn which forced the C-terminus of the peptide to fold onto itself (rather similar to the W3 structure shown in Figure 4).

For the TFE/water simulation, the average upper bound NOE violation is 0.113 Å with 9 violations out of a total of 56 NOEs. The larger number of observed NOEs is in agreement with the stabilization of the peptide's secondary structure. The most significant contribution to the increase of the average violation is again due to $d_{\alpha N}(i, i+2)$ NOEs, this time between the (+7, +9), (+8, +10), and (+9, +11) residue pairs. Indeed, if these three NOEs are excluded from the calculation, the average upper bound violation becomes only 0.054 Å. We have again attempted to place this deviation in a structural context by selecting peptide structures with short $d_{\alpha N}(i, i+2)$ distances for these residue pairs. As with the water simulation, it appears that what these structures share is the presence of a reverse turn for the residues examined. Taken together, the observations concerning the $d_{\alpha N}(i, i+2)$ distances may indicate a force-field-dependent disfavoring of reverse turns, although this statement is difficult to support on the basis of such limited experimental evidence.

Finally, and although no structural models have been deposited by Yun et al.,[9] images of the NMR-derived peptide structures have been published in printed form for both the water and TFE/water solvents (Yun et al.,[9] Figure 3). Both images show structures for the ftz peptide that are in excellent agreement with the results shown in Figure 4 of this paper: the water structure is highly flexible with a relatively short helical segment, whereas in the presence of TFE the helical structure is significantly extended and stabilized.

Before closing this section, we should note that Yun et al.[9] also reported results from circular dischroism (CD) studies of the ftz peptide in water and TFE/water mixtures (Yun et al.,[9] Figure 1a). A direct comparison between the complete CD spectra and the results obtained from the simulations is probably meaningless given the incomplete sampling of the unfolded state, as was discussed in section 2.4. A qualitative assessment of the agreement between the experimental CD results and the simulations can, however, be performed by

DOI: 10.1021/acs.jpcb.7b10292
J. Phys. Chem. B 2018, 122, 106−116

focusing on specific wavelengths that are associated with the peptide's anticipated secondary structure features. Using the procedure described by Ding et al.[54] and taking into account only the experimental ellipticities recorded at 222 nm, we obtain an estimated helical content of ∼7% for the ftz peptide in water and ∼29% for the peptide in 50% TFE/water mixture. Although both of these estimates are on the low side of the simulation-derived percentages, their relative proportion (a difference of approximately a factor of 4 between water and TFE/water mixture) is nearly identical with the results obtained from simulation (∼10% versus ∼40%). In an attempt to make the comparison with the CD data more meaningful, we show in the Supporting Information Figure S3 the results obtained from a more extensive calculation performed with the DichroCalc server.[55] The procedure is as follows. The ellipticities at 222 nm of approximately 6000 equally spaced structures from each of the water and TFE/water simulations were calculated using the empirical parameter set of the DichroCalc server.[55] The distributions of these ellipticities are shown in the form of solid curves in Figure S3. Superimposed on these two curves are shown the experimentally obtained ellipticities at 222 nm (black and red circles in Figure S3). Comparison between the simulation-derived distributions and the experimental data shows that the agreement appears to be quite convincing, with the only consistent difference between them being an apparent overstabilization of the helical structures in the case of the water simulation.

All in all, comparison of the simulations with the experimental results shows a quantitative agreement not only at the level of the similarity of the peptide structures with both the X-ray and NMR-derived conformers but also with respect to recorded NOE measurements for the ftz peptide in two different solvents. Although some notable deviations have been identified concerning the $d_{\alpha N}(i, i+2)$ distances mentioned above, we believe that the level of agreement observed suffices for validating both the simulation protocol and the major conclusions drawn from the molecular dynamics trajectories.

4. CONCLUSIONS

Analysis of the folding simulations and comparison with the experimental data established that the nuclear-receptor-box-containing peptide derived from Drosophila's *fushi tarazu* segmentation gene product demonstrates a persistent helical bias for the LxxLL motif even in the absence of its cognate receptor. The picture painted from the water simulation is that of a highly flexible peptide with a mostly disordered N-terminus, a detectable helical preference for residues −2 to +6, and a flexible—but on average more compact—C-terminal region. The presence of TFE significantly stabilizes both the helical character of the peptide and the extent of the observed helicity, in good agreement with the experimental findings. Clearly, and at least for the case examined here, the structure of the peptide in complex with its receptor is fully consistent with the structural preferences demonstrated by the free peptide in solution. Although we would not like to enter a discussion concerning the effect that pre-existing structural preferences of peptides can have on protein structure, we must note that at least for such relatively long peptides the presence of such structurally persistent preferences appears to make sense in a thermodynamic context, especially when considering that the given peptide is involved in forming a protein−peptide complex.

The comparison between the water and TFE/water simulations together with the results from their validation against experimental data shows that parametrization of mixed solvents may have reached the stage of maturity needed for performing physically relevant biomolecular folding simulations. The differences observed in the folding behavior of the ftz peptide in the presence of TFE are fully consistent with the experimental NMR data and with the known effects of TFE on secondary structure stability (noting here the interesting simulation-derived prediction that the mixed solvent preferably stabilizes α-helical but not 3_{10}-helical structures). Having said that, and as discussed in section 2.4 and Figure 1, the presence of mixed solvents adds very significantly to the amount of simulation time needed to obtain meaningful statistics.

We will close this section by revisiting two notable findings that we have been unable to fully analyze or validate due to shortage of experimental data. The first concerns the mutual exclusivity of the 3_{10}- and α-helical regions shown in Figure 3 for the water simulation. Although it is possible that this might be a force-field-dependent artifact, the presence of mixed α- and 3_{10}-helical peptide structures in the TFE/water simulation probably excludes this scenario and points to a structural reason. We are currently studying a much shorter 3_{10}-helical peptide with the aim of establishing whether this behavior is indeed idiosyncratic for the ftz peptide, or whether it represents a more general secondary structure preference.

The second issue concerns the deviation between the experimental and simulation-derived $d_{\alpha N}(i, i+2)$ distances discussed in section 3.5. Although this appears to be an issue arising from the force field parametrization, it is very difficult to support such a statement with any confidence based on evidence obtained from such a flexible system and with such relatively scarce experimental data. To tackle the issue, we are currently revisiting other peptide simulations for which the corresponding NMR data contain such NOEs.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcb.7b10292.

Extent of sampling for secondary structure preferences; superposition of peptide structures from a dPCA-derived cluster; comparison of simulation-derived NOEs with the experimentally determined values; comparison between experiment and simulation for the circular dichroism data (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*Phone: +30-25510-30620. Fax: +30-25510-30620. Web page: https://utopia.duth.gr/glykos/. E-mail: glykos@mbg.duth.gr.

ORCID

Nicholas M. Glykos: 0000-0003-3782-206X

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Plevin, M. J.; Mills, M. M.; Ikura, M. The LxxLL motif: a multifunctional binding sequence in transcriptional regulation. *Trends Biochem. Sci.* **2005**, *30*, 66−69.

(2) Savkur, R. S.; Burris, T. P. The coactivator LXXLL nuclear receptor recognition motif. *J. Pept. Res.* **2004**, *63*, 207−212.

DOI: 10.1021/acs.jpcb.7b10292
J. Phys. Chem. B 2018, 122, 106−116

(3) Jin, L.; Li, Y. Structural and functional insights into nuclear receptor signaling. *Adv. Drug Delivery Rev.* **2010**, *62*, 1218−26.
(4) White, J. H.; Fernandes, I.; Mader, S.; Yang, X. J. Corepressor recruitment by agonist-bound nuclear receptors. *Vitam. Horm.* **2004**, *68*, 123−143.
(5) Yu, Y.; Li, W.; Su, K.; Yussa, M.; et al. The nuclear hormone receptor Ftz-F1 is a cofactor for the Drosophila homeodomain protein Ftz. *Nature* **1997**, *385*, 552−555.
(6) Blumberg, B.; Evans, R. M. Orphan nuclear receptors − − new ligands and new possibilities. *Genes Dev.* **1998**, *12*, 3149−3155.
(7) Thummel, C. S. From embryogenesis to metamorphosis: the regulation and function of drosophila nuclear receptor superfamily members. *Cell* **1995**, *83*, 871−877.
(8) Yoo, J.; Ko, S.; Kim, H.; Sampson, H.; Yun, J.-H.; Choe, K.-M.; Chang, I.; Arrowsmith, C. H.; Krause, H. M.; Cho, H.-S.; Lee, W. Crystal structure of fushi tarazu factor 1 ligand binding domain/fushi tarazu peptide complex identifies new class of nuclear receptors. *J. Biol. Chem.* **2011**, *286*, 31225−31231.
(9) Yun, J.-H.; Lee, C.-J.; Jung, J.-W.; Lee, W. Solution Structure of LXXLL-related cofactor peptide of orphan nuclear receptor FTZ-F1. *Bull. Korean Chem. Soc.* **2012**, *33*, 583−588.
(10) Case, D. A.; Cheatham, T. E.; 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.
(11) 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.
(12) 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.
(13) Chitra, R.; Smith, P. E. Properties of 2,2,2-trifluoroethanol and water mixtures. *J. Chem. Phys.* **2001**, *114*, 426−435.
(14) Yu, Y.; Wang, J.; Shao, Q.; Shi, J.; Zhu, W. The effects of organic solvents on the folding pathway and associated thermodynamics of proteins: a microscopic view. *Sci. Rep.* **2016**, *6*, 1−12.
(15) Dupradeau, F. Y.; Pigache, A.; Zaffran, T.; Savineau, C.; Lelong, R.; Grivel, N.; Lelong, D.; Rosanski, W.; Cieplak, P. The R.E.D. Tools: Advances in RESP and ESP charge derivation and force field library building. *Phys. Chem. Chem. Phys.* **2010**, *12*, 7821−7839.
(16) 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: Struct., Funct., Genet.* **2006**, *65*, 712−725.
(17) 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.
(18) 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: Struct., Funct., Genet.* **2010**, *78*, 1950−1958.
(19) 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.
(20) 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.
(21) Patapati, K. K.; Glykos, N. M. Three force fields' views of the 3_{10} helix. *Biophys. J.* **2011**, *101*, 1766−1771.
(22) 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.
(23) Patapati, K. K.; Glykos, N. M. Order through disorder: hypermobile C-terminal residues stabilize the folded state of a helical peptide. A molecular dynamics study. *PLoS One* **2010**, *5*, e15290.
(24) 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.
(25) Razavi, A. M.; Voelz, V. A. Kinetic network models of tryptophan mutations in β-hairpins reveal the importance of non-native interactions. *J. Chem. Theory Comput.* **2015**, *11*, 2801−2812.
(26) Best, R. B.; Zheng, W.; Mittal, J. Balanced protein−water interactions improve properties of disordered proteins and non-specific protein association. *J. Chem. Theory Comput.* **2014**, *10*, 5113−5124.
(27) Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E. Water dispersion interactions strongly influence simulated structural properties of disordered protein states. *J. Phys. Chem. B* **2015**, *119*, 5113−5123.
(28) Mercadante, D.; Milles, S.; Fuertes, G.; Svergun, D. I.; Lemke, E. A.; Gräter, F. Kirkwood−Buff approach rescues overcollapse of a disordered protein in canonical protein force fields. *J. Phys. Chem. B* **2015**, *119*, 7975−7984.
(29) 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.
(30) Serafeim, A.-P.; Salamanos, G.; Patapati, K. K.; Glykos, N. M. Sensitivity of folding molecular dynamics simulations to even minor force field changes. *J. Chem. Inf. Model.* **2016**, *56*, 2035−2041.
(31) Koukos, P. I.; Glykos, N. M. Folding molecular dynamics simulations accurately predict the effect of mutations on the stability and structure of a Vammin-derived peptide. *J. Phys. Chem. B* **2014**, *118*, 10076−10084.
(32) Baltzis, A. S.; Glykos, N. M. Characterizing a partially ordered miniprotein through folding molecular dynamics simulations: comparison with the experimental data. *Protein Sci.* **2016**, *25*, 587−596.
(33) Zhang, C.; Ma, J. Enhanced sampling and applications in protein folding in explicit solvent. *J. Chem. Phys.* **2010**, *132*, No. 244101.
(34) Glykos, N. M. CARMA: a molecular dynamics analysis program. *J. Comput. Chem.* **2006**, *27*, 1765−1768.
(35) Koukos, P. I.; Glykos, N. M. grcarma: a fully automated task-oriented interface for the analysis of molecular dynamics trajectories. *J. Comput. Chem.* **2013**, *34*, 2310−2312.
(36) Baltzis, A. S.; Koukos, P. I.; Glykos, N. M. Clustering of molecular dynamics trajectories via peak-picking in multidimensional PCA-derived distributions. *arXiv.* **2015**, 1512.04024 [q-bio.BM].
(37) Shen, Y.; Bax, A. SPARTA+: a modest improvement in empirical nmr chemical shift prediction by means of an artificial neural network. *J. Biomol. NMR* **2010**, *48*, 13−22.
(38) Frishman, D.; Argos, P. Knowledge-based protein secondary structure assignment. *Proteins: Struct., Funct., Genet.* **1995**, *23*, 566−579.
(39) Humphrey, W.; Dalke, A.; Schulten, K. VMD—visual molecular dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.
(40) Merritt, E. A.; Bacon, D. J. Raster3D photorealistic molecular graphics. *Methods Enzymol.* **1997**, *277*, 505−524.
(41) *The PyMOL molecular graphics system*, version 1.8; Schrödinger LLC: New York, 2015.
(42) 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.
(43) Hamming, R. W. Error detecting and error correcting codes. *Bell Syst. Tech. J.* **1950**, *29*, 147−160.
(44) Crooks, G. E.; Hon, G.; Chandonia, J. M.; Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **2004**, *14*, 1188−1190.
(45) Lacroix, E.; Viguera, A. R.; Serrano, L. Elucidating the folding problem of alpha-helices: local motifs, long-range electrostatics, ionic-strength dependence and prediction of NMR parameters. *J. Mol. Biol.* **1998**, *284*, 173−191.

DOI: 10.1021/acs.jpcb.7b10292
J. Phys. Chem. B 2018, 122, 106−116

(46) Buchan, D. W. A.; Minneci, F.; Nugent, T. C. O.; Bryson, K.; Jones, D. T. Scalable web services for the PSIPRED protein analysis workbench. *Nucleic Acids Res.* **2013**, *41*, W349−W357.
(47) Xiong, K.; Asher, S. A. Circular dichroism and UV resonance raman study of the impact of alcohols on the Gibbs free energy landscape of an α-helical peptide. *Biochemistry* **2010**, *49*, 3336−42.
(48) Santiveri, C. M.; Pantoja-Uceda, D.; Rico, M.; Jiménez, M. A. Beta-hairpin formation in aqueous solution and in the presence of trifluoroethanol: a (1)H and (13)C nuclear magnetic resonance conformational study of designed peptides. *Biopolymers* **2005**, *79*, 150−162.
(49) Buck, M. Trifluoroethanol and colleagues: cosolvents come of age. Recent studies with peptides and proteins. *Q. Rev. Biophys.* **1998**, *31*, 297−355.
(50) Mu, Y.; Nguyen, P. H.; Stock, G. Energy landscape of a small peptide revealed by dihedral angle principal component analysis. *Proteins: Struct., Funct., Genet.* **2005**, *58*, 45−52.
(51) Altis, A.; Nguyen, P. H.; Hegger, R.; Stock, G. Dihedral angle principal component analysis of molecular dynamics simulations. *J. Chem. Phys.* **2007**, *126*, 244111.
(52) Altis, A.; Nguyen, P. H.; Hegger, R.; Stock, G. Construction of the free energy landscape of biomolecules via dihedral angle principal component analysis. *J. Chem. Phys.* **2008**, *128*, 245102.
(53) Zagrovic, B.; van Gunsteren, W. F. Comparing atomistic simulation data with the NMR experiment: how much can NOEs actually tell us? *Proteins: Struct., Funct., Genet.* **2006**, *63*, 210−218.
(54) Ding, F. X.; Schreiber, D.; VerBerkmoes, N. C.; Becker, J. M.; Naider, F. The chain length dependence of helix formation of the second transmembrane domain of a G protein-coupled receptor of Saccharomyces cerevisiae. *J. Biol. Chem.* **2002**, *277*, 14483−14492.
(55) Bulheller, B. M.; Hirst, J. D. DichroCalc - circular and linear dichroism online. *Bioinformatics* **2009**, *25*, 539−540.

DOI: 10.1021/acs.jpcb.7b10292
J. Phys. Chem. B 2018, 122, 106−116