# On the Foldability of Tryptophan-Containing Tetra- and Pentapeptides: An Exhaustive Molecular Dynamics Study

Panagiota S. Georgoulia and Nicholas M. Glykos*

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

S Supporting Information

ABSTRACT: Short peptides serve as minimal model systems to decipher the determinants of foldability due to their simplicity arising from their smaller size, their ability to echo protein-like structural characteristics, and their direct implication in force field validation. Here, we describe an effort to identify small peptides that can still form stable structures in aqueous solutions. We followed the *in silico* folding of a selected set of 8640 tryptophan-containing tetra- and pentapeptides through 15 210 molecular dynamics simulations amounting to a total of 272.46 μs using explicit representation of the solute and full treatment of the electrostatics. The evaluation and sorting of peptides is achieved through scoring functions, which include terms based on interatomic vector distances, atomic fluctuations, and rmsd matrices between successive frames of a trajectory. Highly scored peptides are studied further via successive simulation rounds of increasing simulation length and using different empirical force fields. Our method suggested only a handful of peptides with strong foldability prognosis. The discrepancies between the predictions of the various force fields for such short sequences are also extensively discussed. We conclude that the vast majority of such short peptides do not adopt significantly stable structures in water solutions, at least based on our computational predictions. The present work can be utilized in the rational design and engineering of bioactive peptides with desired molecular properties.
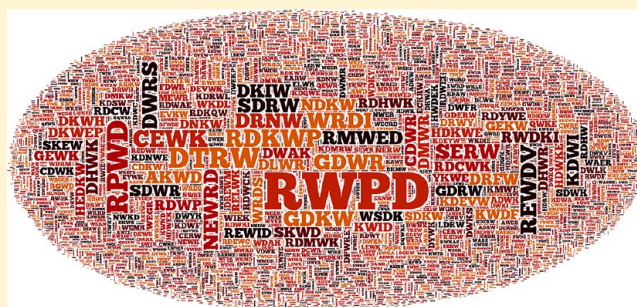
1. INTRODUCTION

Peptide folding simulations[1−8] have evolved to a highly active research field because their significance is threefold: peptides serve as model systems to optimize and validate force fields,[9−13] as model systems to study the protein folding problem,[14−16] and as potential candidates in drug design studies.[17−20] Fast-folding miniproteins are the new "play toys" because their folding can be approximated by both theory and experiment.[21−33]

The quest for fast-folding proteins reveals questions regarding the lower size limit of peptides that can adopt stable folds in aqueous solutions. To our knowledge, the smallest peptides of unique structure comprise approximately 10 residues, like chignolin[34] and CLN025,[35] thus termed as microproteins. Capped terminal ends and unnatural amino acids lower further the size limit for protein-like fold stability to 7−10 residues.[36] Smaller peptides of biological interest have been studied occasionally[3,37−42] but were found to be only marginally stable in water solutions. However, the usage of peptides in real-world applications, such as in molecular biology, nanomedicine, and nanotechnology,[43] depends on their size, foldability, and solubility.[17−19,44] De novo design of non-natural peptides often results in compounds with variable solubility, low bioavailability, and limited stability, which impedes their therapeutic application, whereas naturally occurring peptides have greater specificity and lower toxicity

but are large enough to be typically subject to proteolysis.[17−19] Therefore, there is a need to design low-molecular weight peptides that adopt stable folds in water environments and possess particular molecular properties that fit each specific need at a time. By using the ever-increasing power of computers, state-of-the-art empirical force fields, and algorithms to explore systematically a large number of peptides, we can narrow them down to a manageable number of putatively structurally stable ones.

To this end, we performed a systematic search for putatively foldable peptides that are four- and five-residue long. Tetrapeptides and pentapeptides have been implicated in the initiation of folding,[45−47] in structure prediction studies,[48] and in peptide design,[49,50] whereas larger peptides (greater than eight residues) are prone to aggregation due to low solubility.[51−53] In this work, a subset of 8640 peptides from the vast total of 3 360 000 possible sequence combinations is examined. We followed their *in silico* folding and structural characterization through an exhaustive set of 15 210 independent molecular dynamics simulations resulting in a total of 272.46 μs of simulation time in explicit solvent and with full treatment of the electrostatics. We present a combination of

scoring functions designed specifically to identify folding events and designate stably folded peptides.

In what follows, we show that, according to the general expectation, the vast majority of tetrapeptides and pentapeptides is quite disordered with their folded population occupying only a small fraction of the observed simulation time. Systematic analysis of these simulations only identified four peptides (two pentapeptides and two tetrapeptides) for which computational evidence is consistent with the presence of a stable structure. The present study revealed great discrepancies in the predictions of the various examined force fields regarding the folding behavior of this particular set of peptides. The same peptide can be predicted to be anything from fully disordered to folded to either one or more distinct structures, according to some or all force fields. The predicted structures are sometimes the same or differ even among force fields of the same family. We conclude with possible interpretations of our findings and their role in the validation and improvement of the current-generation theoretical tools.

2. METHODS

2.1. Peptide Selection. The initial peptide pool that we examined comprises 1440 tetrapeptides and 7200 pentapeptides (Table 1 and Figure 1). The reduction of the total number of unique sequence combinations (160 000 and 3 200 000, respectively) was accomplished through the incorporation of a set of restrictions. The first was the requirement for the presence of a tryptophan residue to favor the development of electrostatic interactions and hydrophobic packing[54,55] and also to be able to obtain near-UV CD spectra[56] for later experimental validation. The second restriction was the requirement for the presence of two charged residues of opposite charges, which would guarantee solubility in aqueous solutions and favor the development of stabilizing interactions[6,57] with the tryptophan residue. For the pentapeptide set, we used as a restriction the presence of three unique charged residues (two with a positive charge, one with negative, or the opposite).

2.2. System Preparation and Simulation Protocol. The system preparation procedure was performed automatically for all simulations reported here (a grand total of 15 210 independent runs). All simulations were performed with NAMD[58] using explicit solvent, full treatment of the electrostatics, and periodic boundary conditions. We examined the force fields CHARMM22,[59] CHARMM27-CMAP,[60] OPLS-AA,[61,62] AMBER99SB,[63,64] and AMBER99SB-ildn,[65] as noted in Table 1. The choice of force field in each simulation round was made based on the performances of the force fields at the time the simulations took place,[13,66] regarding the folding of small peptides.[12,67,102,107,108]

The starting structure of the peptides was in the fully extended form, as obtained from the program Ribosome (Shrinivasan, http://www.roselab.jhu.edu/~raj/Manuals/ribosome.html) with both terminal ends unprotected. For both the CHARMM and the OPLS force fields, missing hydrogen atoms were built with the program PSFGEN from the NAMD[58] distribution and solvation and ionization were performed with VMD.[68] In the case of AMBER, system preparation was performed with the program TLEAP from the AMBER tools distribution.[69] In all cases, peptides were embedded in a truncated octahedral water box, with initial cube dimensions of 28 × 28 × 28 Å³ and an initial solute–solute distance of 12 Å, filled with pre-equilibrated TIP3P[70]

water molecules along with sodium and chloride ions to a final concentration of ~30 mM. All final systems comprised approximately 900 atoms.

The simulation protocol was the following: At first, systems were energy minimized for 1000 conjugate gradient steps. Then the temperature was increased with a ΔT step of 20 K until a final temperature of 320 K (or to the final temperature noted in Table 1) over a period of 32 ps. Subsequently the systems were equilibrated for 1000 steps with the positions of the Cα atoms restrained followed by an equilibration period for the production of the NpT runs with the temperature (320 K) and the pressure (1 atm) controlled using the Nosé–Hoover Langevin Dynamics and Langevin piston barostat control method and the Verlet-I multiple-time step integration algorithm, as implemented by NAMD.[58] The Langevin damping coefficient was set to 1 ps^{-1}, and the piston's oscillation period was set to 200 fs, with a decay time of 100 fs. The production run was performed with the impulse Verlet-I[71] multiple time step integration algorithm as implemented by NAMD.[58] The inner time step was 2 fs, with nonbonded interactions calculated every two time steps and full electrostatics every four time steps using the particle mesh Ewald method[72] with a grid spacing of approximately 1 Å and a tolerance of 10^{-6}. A cutoff for the van der Waals interactions was applied through a switching function, acting between 7 and 9 Å. The SHAKE[73] algorithm (with a tolerance of 10^{-8}) was used to restrain all bonds including the hydrogen atoms. Trajectories were obtained by saving the atomic coordinates of the whole system every 0.8 ps.

2.3. Design of Simulations and Sorting of Peptides. Given the known theoretical[4,37,39,74,75] and experimental[76–87] folding rates of such short peptides, we performed five independent simulation rounds for each of the two sets of peptides, as presented in Table 1. The initial pass comprises all peptide sequences and short trajectories (5 ns for the tetrapeptides and 20 ns for the pentapeptides) aiming toward identification of early folding events. These comprise mainly loop-closure events and formation of the first turn of a helix (helix nucleation) that occur in the 10–20[79–84] and 1–5 ns time scale,[85–87] respectively, for peptide sequences of comparable length.

During this first pass, short trajectories were evaluated through a scoring function designed to specifically identify a rapid transition that leads to a loop-closure event along with the formation of a stable structure. This target function is based on interatomic vector distances and will be referred to as TF1 (see section 2.4 for more details). Peptides were selected by systematic application of the TF1.

In the second pass, peptides were evaluated through another scoring function to identify stably folded peptides. This second target function is based on an rmsd matrix between structures from successive frames of the trajectory and will be referred to as TF2 (section 2.4). Prominent peptides were selected again by systematic application of the TF2 function. The efficacy of both scoring functions, TF1 and TF2, in sorting the peptides according to certain criteria (section 2.4), was ascertained after visually examining the majority of the produced graphs.

In the third pass, peptides were selected after applying the full battery of analysis, principal component analysis and corresponding cluster analysis, calculation of average atomic fluctuations, especially of the tryptophan's side-chain, along with the scoring function TF2. Two putatively stably folded tetrapeptides were identified after examining different force

fields and simulation temperatures (fourth and fifth pass in Table 1). Foldable pentapeptides were identified through 1 and 2 μs simulations (fourth and fifth pass in Table 1).

2.4. Target Functions. The first scoring function, TF1, is applied to short trajectories (first pass in Table 1) and its purpose is to capture folding events. For such short sequences a direct and computationally efficient way to follow structural changes is the evolution of interatomic distances (Figure S1, Supporting Information). Distance is a measure that allows also a direct comparison between theory and experiment. For example, FRET[88] studies on small peptides (9−10 residues)[89,90] and a series of polyproline polymers (6−40 residues)[91] showed an excellent agreement with the simulation-derived distances.

We follow the evolution of three distances between the three pairs of Cα atoms that are more distant (Figure S1, Supporting Information). The distances are evaluated based on the observation of transitions, their number and duration, and the correlation among all three possible pairs of the three calculated distances. High scores are attributed to peptides with fast and early transitions and as few and coordinated transitions as possible. The TF1 score is a summation over all unique (Cα_i−Cα_j) distances of the following function:

$$T = \sum_{i,j=0,i \neq j}^{2} t_i t_j \mathrm{Corr}(t_i t_j)$$

where

$$t_{i,j} = \frac{D \, \max_{\mathrm{rmsd}}}{\mathrm{aver} \, \min_{\mathrm{rmsd}} \mathrm{rmsd}}$$

where $\mathrm{Corr}(t_i t_j)$ is the Pearson correlation coefficient between the variation versus time of two (Cα_i−Cα_j) distances, D is the difference between the maximum and minimum observed distance value, aver is the mean value and rmsd is the corresponding root-mean-square deviation calculated over all observed distance values, and $\max_{\mathrm{rmsd}}$ and $\min_{\mathrm{rmsd}}$ are the maximun and minimum rmsd values calculated over a sliding window of a certain width.

The D parameter is used to evaluate the sharp transitions, the aver parameter is used to exclude peptides that remained in the extended conformation, and the rmsd parameter is used to exclude peptides with multiple folding/unfolding transitions; rmsd is calculated over a sliding window (400 fs in this case) to evaluate gradient transitions of the distance's values and local fluctuations between the transitions in order to rank peptides appropriately.

The second scoring function, TF2, is applied to all trajectories from the second round of simulations onward (Table 1) and its purpose is to identify stably folded peptides. We identify the formation of a cluster of closely related structures through the calculation of root-mean-square matrices (rmsd matrix, see Figure S1, Supporting Information). These matrices are square and symmetrical, with their axes representing structures from successive frames of the trajectory. The matrix entries are filled with the rmsd value between two structures after their superposition. The rmsd values are calculated using all heavy atoms so as to find folded structures where both the backbone and the side chains have stabilized. The goal is to identify the formation of a cluster of closely related (low rmsd values) structures. The color interpretation of the formation of a cluster of structures is a large dense blue box (rmsd < 2.0 Å) along the diagonal in the graphical

representation of a matrix (Figures 2 and 3, see also Supporting Information Figures S1, S3, S4, and S5).

The TF2 function (a graphical representation of which is shown in Figure S2, Supporting Information) is calculated through a custom computer program: (1) read the square matrix; (2) convert it to binary mode by assigning every rmsd value ≥2.0 Å to 0 and every other value to 1; (3) for every possible symmetric submatrix of dimensions $(N − m) \times (N − m)$, where N is the initial matrix dimensions and m ranges from 2 to $N − 2$, count the percent of matrix entries with value 1; (4) calculate a histogram of the distribution of the percent of matrix entries with value 1 per dimension of each submatrix of the previous step; (5) return the product of median and mode of the distribution.

2.5. Trajectory Analysis. The program CARMA[92] together with custom scripts were used for most of the analyses, including removal of overall rotations and translations, calculation of the evolution of the Cα distances, calculation of the rmsd matrices, principal component analysis in both Cartesian[93,94] and dihedral space[95,96] and corresponding cluster analysis, calculation of average structures and of the atomic root-mean-square fluctuations, and production of PDB files from the trajectories. All molecular graphics work and figure preparation was performed with the programs VMD[66] and CARMA.[92]

2.6. Sufficient Sampling. To demonstrate that the extreme systems we have chosen (approximately 900 atoms and an initial solute−solute distance of 12 Å, referred to as small PBC box) do not suffer from altered dynamics and periodicity artifacts, we performed two additional simulations. Two peptides, RWPD (Figure 1, see also Supporting Information Figures S2, S4, and S7) and ECKRW (see Supporting Information Figures S5 and S6), were studied through folding simulations until convergence was achieved (446 and 388 ns, respectively). We followed the same simulation protocol (section 2.2) except for the initial cube dimensions, which were $32 \times 32 \times 32$ Å^3 (large PBC box), and the initial solute−solute distance, which was 16 Å. The final systems comprised approximately 1700 atoms.

Convergence was examined by dividing the trajectories into two nonoverlapping halves and performing dihedral angle principal components analysis on each half, followed by the calculation of the eigenvector overlap (as a function of the number of eigenvectors) for the two halves. The overlaps in dihedral space using the top three and top six eigenvectors are 0.77 and 0.87 for the tetrapeptide system. The corresponding values for the pentapeptide system are 0.69 and 0.73 for the top three and top six eigenvectors, respectively.

To validate our extreme systems (small PBC box), we calculated the eigenvector overlap between the trajectories of the RWPD and ECKRW peptides that were produced by the two protocols with the small and large PBC boxes. As representative trajectories for the small PBC box, we used the 1 μs trajectories from the fifth pass and fourth pass (as noted in Table 1) for the two peptides. In the case of the tetrapeptide system, the overlap in dihedral space is 0.89 with three eigenvectors and 0.95 with six eigenvectors, thus indicating that the two trajectories are consistent. Similarly, for the pentapeptide system, the overlap in dihedral space is 0.81 with the three eigenvectors and 0.82 with the six eigenvectors.

The choice of a slightly smaller PBC box that still produces equivalent trajectories allows, however, an increase in the observed throughput by 30%, which has a significant impact on
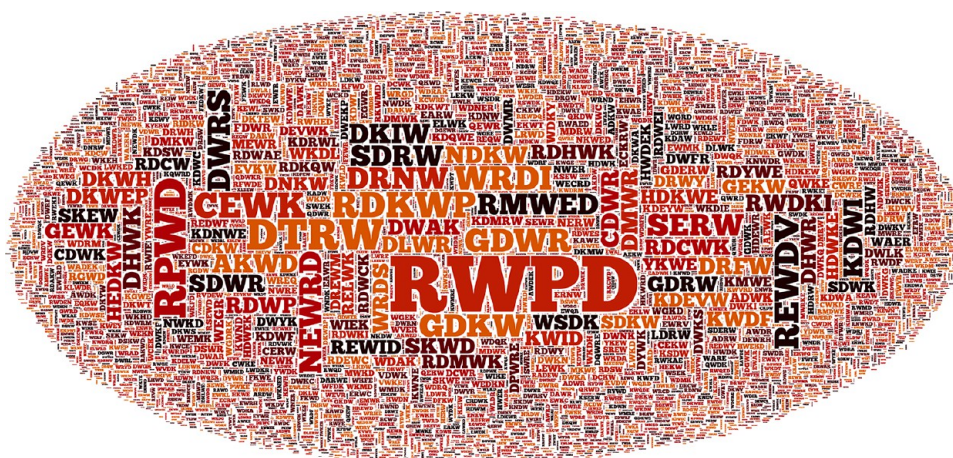
Figure 1. Word cloud of 8640 tetrapeptides and pentapeptides of their respective foldability. Greater prominence is given to the putatively foldable peptides according to a cumulative score from all five independent simulation rounds as presented in Table 1 (TF1 score from first pass and TF2 score from second through fifth pass).

Table 1. Summary of All Simulations Performed in the Two Sets of Tetrapeptides and Pentapeptides[a]

	number of peptides	number of runs	simulation time per run (ns)	notes
			Tetrapeptides	
1st pass	1440	4	5	four independent repeats performed with CHARMM22 at 320 K
2nd pass	130	1	30	one repeat performed with CHARMM22 at 320 K
3rd pass	36	1	100	one repeat performed with CHARMM22 at 320 K
4th pass	4	4	300	four independent repeats performed with CHARMM22 at 283, 298, 320, and 340 K
5th pass	2	3	1000	three independent repeats performed with CHARMM-CMAP, AMBER99SB, OPLS-AA, at 320 K
			Pentapeptides	
1st pass	7200	1	20	one repeat performed with CHARMM22 at 320 K
2nd pass	480	1	100	one repeat performed with CHARMM22 at 320 K
3rd pass	32	4	120	four independent repeats performed with CHARMM-CMAP, AMBER99SB, AMBER99SB-ildn, OPLS-AA at 320 K
4th pass	8	1	1000	one repeat performed with AMBER99SB-ildn at 320 K
5th pass	2	3	2000	three independent repeats performed with AMBER99SB-ildn at 298, 340, and 360 K

[a]For each successive round of simulations, we indicate the number of peptides under study and the length and number of independent runs of simulations, along with the temperature and the force field used in each case. The total simulation time amounts to 47.1 μs (37.680 core hours) and 225.36 μs (135.216 core hours) for the tetrapeptide and pentapeptide set, respectively.

the physical time needed to complete the grand total of 15 210 simulations reported here (see section 2.7).

2.7. Computational Means and Requirements. All simulations reported here (a grand total of 15 210 independent runs) were performed on a stateless Beowulf cluster (http://norma.mbg.duth.gr) based on quad core processors on a gigabit interconnect using an SMP-enabled version of NAMD (v.2.6, v.2.7b, or v.2.8). Due to the small size of the systems, the simulations scaled to either one or four cores, giving a maximum observed throughput of 30 ns/day or 100 ns/day, respectively, for a 900 atom system, requiring 461 days of physical time on 32 dedicated cores for a total of 272.46 μs of simulation time.

3. RESULTS

3.1. Most Tetrapeptides and Pentapeptides Are Disordered. Figures 1 and 2 (and Figure S3 in the Supporting Information) show the order of peptides after the application of the scoring functions TF1 and TF2, during the successive simulation rounds, as presented in Table 1.

In the word cloud of Figure 1, we see all of the 8640 peptides that participated in the simulations presented here (Table 1).

Greater prominence, in the sense of larger font, is given according to their foldability, as evaluated by both scoring functions TF1 and TF2 and from all simulations. The distribution of these cumulative scores reveals only a small fraction (<1%) of outlier peptides that receive high score by the target functions TF1 and TF2. Most peptides are disordered and are characterized by multiple folding/unfolding events, which leads to their underscoring by both target functions. This great difference in the scores that separates the fraction of folder peptides from the large population of nonfolders enables the distinction and consequent selection of the cluster of peptides that are likely to be foldable based on the evolution of the interatomic vector distances and the frame-to-frame rmsd matrices.

Less than 10% of the peptides under study where designated as putatively foldable by the target function TF1 during the first pass and qualified for the next simulation round. As the simulation time increases at the second pass (Table 1), we can view with greater ease the formation of a folded structure, which is evaluated by the scoring function TF2. Figure 2 and Figure S3 (in the Supporting Information) show the order of the pentapeptides and tetrapeptides, respectively, after the
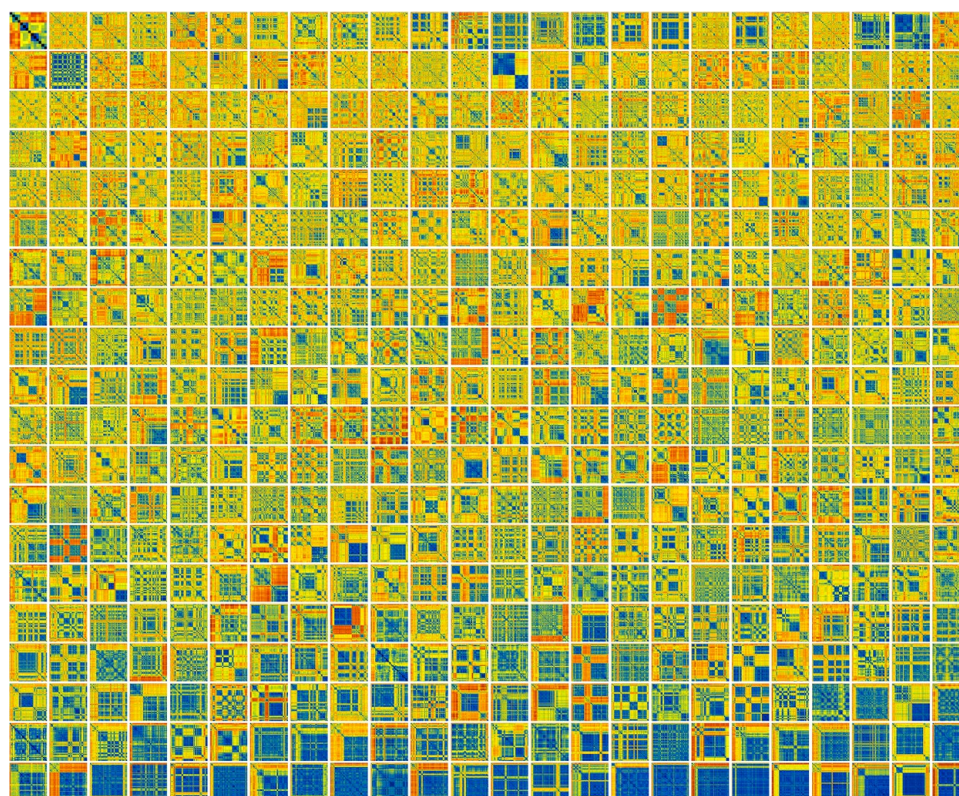
Figure 2. The rmsd matrices of the 480 pentapeptides (second pass) calculated from 100 ns trajectories. The order from left to right and top to bottom follows the increase in the score received by the scoring function TF2. Each image corresponds to an individual peptide and is a color representation of the square matrix of the rmsds among all possible structures from successive frames of the trajectory (see also Figure S1, Supporting Information). For each matrix, the origin ($t = 0$) is at the top left-hand corner. All matrices were calculated using all heavy atoms ($\sim$40 atoms), and the linear color scale is kept the same, ranging from dark blue (0 Å) through yellow ($\sim$3.5 Å) to dark red (7.3 Å).

application of the target function TF2 at the second round of simulations. Only 10% of the pentapeptides (Figure 2) folds to a cluster of closely related structures occupying most of the 100 ns trajectory. Most peptides present a large and dispersed cluster of structures, characteristic of unstable behavior with multiple folding−unfolding events. This is in agreement with the results of Figure S2, Supporting Information, which shows the results from the application of the scoring function TF2 on the tetrapeptide set.

These results taken together show that the vast majority of the peptides under study are highly flexible and disordered for most of the observed simulation time. The second scoring function has the advantage of being independent of the peptide length, unlike the scoring function TF1: the interatomic vector distances in the tetrapeptide set are calculated between Cα atoms that are separated by one or two peptide bonds, whereas in the pentapeptide set they are separated by two or three, resulting in higher distance values, which in turn means higher scoring by the target function TF1. However, the distribution of the scores in the tetrapeptide set inclusively shares the same characteristics with the pentapeptide set, revealing only a small fraction of folder tetrapeptides. The apparent conclusion from the first two simulation rounds is that the number of putatively foldable peptides is very small compared with the number of possible sequence combinations.

A secondary conclusion from the examination of the rmsd matrices of the folding trajectories (Figure 2 and Figures S3 and S4, Supporting Information) is that apart from the great population of highly flexible and disordered peptides, we also

observe peptides with a unique well-defined structure (last row in Figure 2 and Figure S3, Supporting Information). What is even more interesting is the observation of peptides with two well-defined structures (Figure 2 and Figures S3 and S4, Supporting Information). Non-two-state behavior has been observed experimentally in octapeptides.[54] Complex, non-two-state length-dependent folding has also been reported for polyalanine peptides, both theoretically[97] and experimentally.[98] Recently, a model of the conformational dynamics was presented, suggesting a lower length limit for a two-state folder of nine residues.[99]

The small number of tetrapeptides and pentapeptides with strong foldability prognosis were studied further during the third simulation round (third pass in Table 1). Figure 3 and Figures S4 and S5, Supporting Information, show the rmsd matrices obtained from the folding trajectories of the 36 tetrapeptides (Figure S4, Supporting Information) and 32 pentapeptides (Figure 3 and Figure S5, Supporting Information). These results are in agreement with the conclusions discussed above and show that only a small number of peptides adopt a well-defined and stable structure. A secondary characteristic is that in the pentapeptide set the stabilized structure comprises both the backbone and the side chains (compare the two halves of the rmsd matrices in Figure 3 and Figure S5, Supporting Information). On the other hand, in the tetrapeptide set, there are more profound differences in the stability of the backbone compared with the behavior of the side chains, with the latter being considerably more flexible
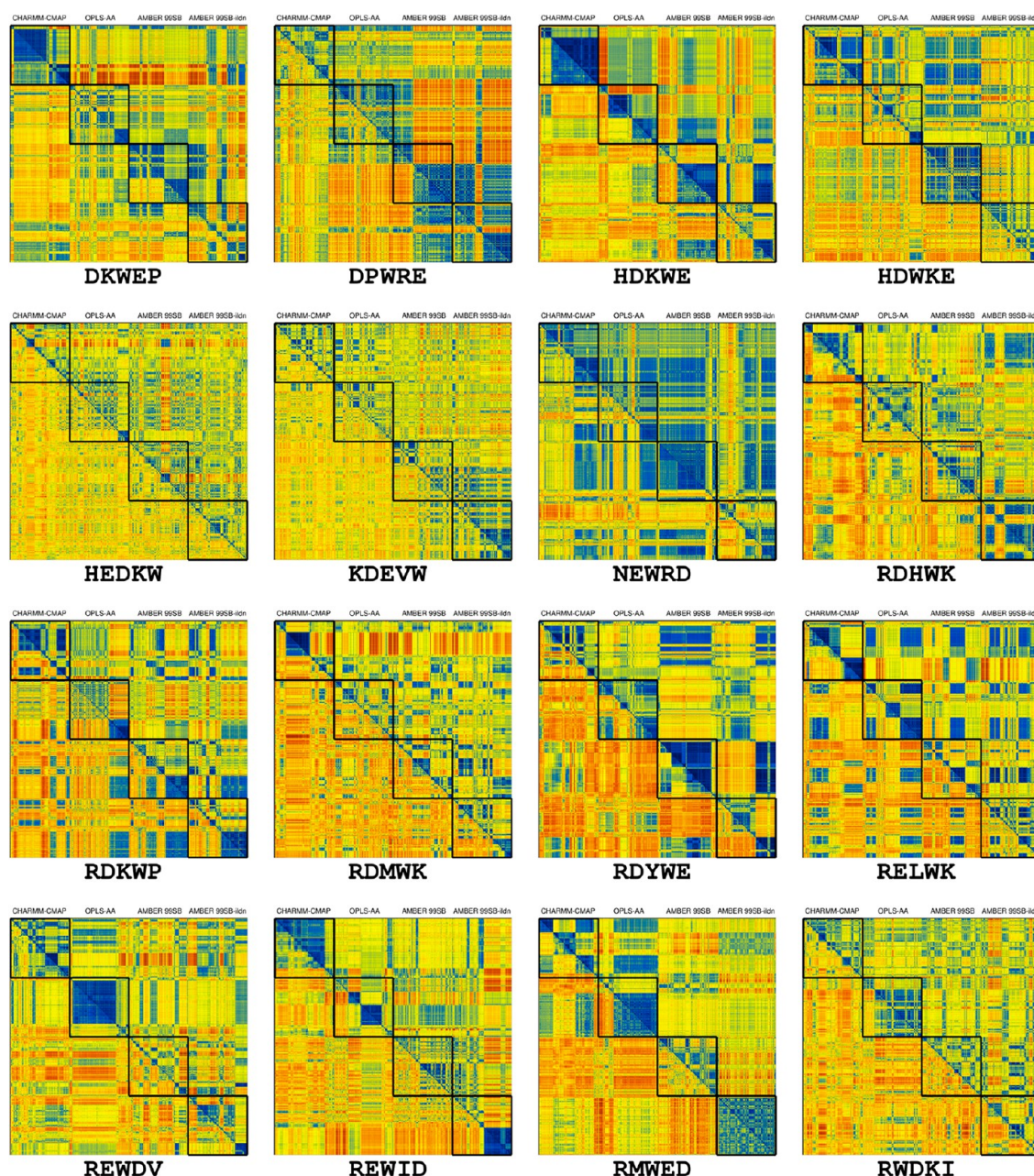
Figure 3. Inter-rmsd matrices for 16 pentapeptides (from the third pass) calculated among four independent trajectories with four different force fields. Each image corresponds to an individual peptide and is a color representation of the square matrix of the (inter) rmsd's among all possible structures from successive frames of an artificial trajectory (480 ns total simulation time for each peptide) produced by concatenating four independent trajectories produced by four force fields (third pass in Table 1). Each matrix was calculated using only backbone atoms (above the diagonal) and all heavy atoms (below the diagonal). The limits of each trajectory are indicated with square black boxes, with the corresponding force field, CHARMM-CMAP, OPLS-AA, AMBER99SB, and AMBER99SB-ildn, noted on the top. The linear color scale is kept the same, ranging from dark blue (0 Å) through yellow (~3.9 Å) to dark red (7.8 Å).

(compare the two halves of the rmsd matrices in Figure S4, Supporting Information).

3.2. Inconsistency between the Force Fields. Figure 3 and Figure S5, Supporting Information, show a color representation of the rmsd matrices for the total of 32 pentapeptides from the third round of simulations (third pass in Table 1). Each rmsd matrix is calculated for the concatenated artificial trajectory of all four force fields under study at this stage (Table 1). As can been seen from those two figures, there is great variance in the predictions among the various force field versions. The systematic preference of the various force fields

toward certain regions of the Ramachandran plot is well documented in the literature.[11,13,63,100−109] However, for the short and almost random sequences that we study, there is significant variance in the force fields' predictions regarding their foldability.

From peptide sequences for which there is a consensus among the force fields (NEWRD, RDKWP, and RELWK in Figure 3) to peptide sequences for which there is complete disagreement (REWID and RMWED in Figure 3 and RWEDA in Figure S5, Supporting Information), we observe almost every possible combination in their predictions even within the small
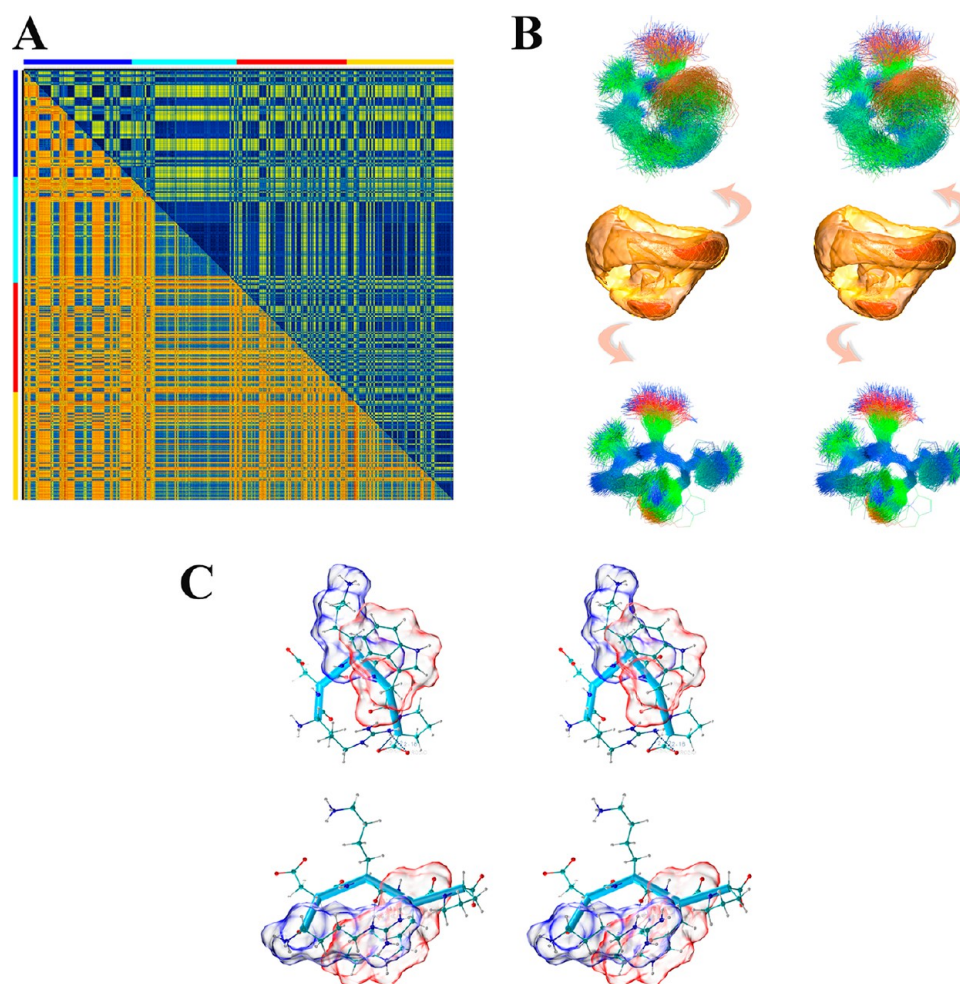
Figure 4. The pentapeptide RDKWP. In panel A, we show the rmsd matrix of the concatenated trajectory corresponding to the simulations in the four temperatures (fourth and fifth pass in Table 1), using the AMBER99SB-ildn force field (7 μs total simulation time). The independent trajectories are indicated with the horizontal and vertical bars, where blue stands for 298 K, cyan for 320 K, red for 340 K, and yellow for 360 K. Above the diagonal, the rmsd calculation is performed only for the backbone atoms and below the diagonal for all-heavy atoms. The linear color scale is kept the same, ranging from dark blue (0 Å) through yellow (~3.5 Å) to dark red (7.1 Å). In panel B, we show the three-dimensional folding landscape from the trajectory obtained at 298 K with the AMBER99SB-ildn force field. The middle is a wall-eyed stereodiagram of the projection of the trajectory on the space defined by the top three principal components as obtained from a principal component analysis in Cartesian space using all-heavy atoms. Three isosurfaces are drawn at the mean density (orange transparent isosurface), 1σ above mean (yellow wireframe), and 3σ above mean (red solid surface) of the map distribution. The arrows point to the two clusters of structures (with 30% and 25% occupancy of the total simulation time) observed during the trajectory. For each cluster, we show (in wall-eyed stereodiagram) 500 snapshot structures, obtained directly from the trajectory. The structures are colored according to their atomic rms fluctuation from the calculated (for the cluster) average structure using a color gradient from blue to red (maximum rmsf, 3.43 Å). In panel C, we show in stereodiagram the representative structures (the trajectory's structure with the smallest rms deviation from the cluster's average one) for the two distinct clusters. All atoms are depicted in ball-and-stick, and the backbone direction is indicated with cyan tube. The hydrogen bond is shown with blue dotted line and the packing between tryptophan−lysine and tryptophan−arginine is represented with transparent surface.

set of 32 pentapeptides. For most cases, a peptide is folded by only one force field or force fields of the same family (RDHWK and ECKRW by the AMBER99SB-ildn, DPWRE by the two AMBER force fields, RDMWK, RIDKW by the CHARMM force field and KDWTR, REWDV, RWHED by the OPLS force field). There are also peptides that are completely disordered according to all four force fields (EKIRW, HEDKW, KDEVW, KREGW, KWLED, RDEIW, REKWM, RLEWD, RWDKI, and WEHKR) even though they were designated as stable folders by the CHARMM22 force field during the previous simulation rounds (Table 1). Peptides with two distinct conformations are observed here more clearly (DKWEP, HDKWE, HDWKE, and RDYWE), a behavior that is predicted by all force fields, regardless of the extent to which they agree in the predicted

structures. In some cases the AMBER force fields agree better with the OPLS, whereas in other cases there is better agreement between the AMBER and the CHARMM force field.

The great variance that we observe in the predictions of the force fields is reason for skepticism, especially in the absence of experimental evidence. This diversity of the results is rather force field dependent and not an artifact of the limited simulation time as shown in Figure S6 (Supporting Information) for which 1 μs simulations were used. The 1 μs simulations of the eight pentapeptides with the AMBER99SB-ildn force field (fourth pass in Table 1) give the same results (compare Figure 3 and Figure S5, Supporting Information, with Figure S6, Supporting Information), regarding the foldability of the peptides with the 0.12 μs simulations with the same force

field (third pass in Table 1). This is also supported by experimental measurements in such short peptides, which give a folding rate in the 10−100 ns regime.[77] This behavior is persistent in the 1 μs simulations of the two tetrapeptides as well (fifth pass in Table 1). In Figure S7, Supporting Information, we see representative results obtained from the folding trajectories of a tetrapeptide, RWPD, using various force fields as well as simulation temperatures. In the case of this tetrapeptide, the force fields predict similar backbone conformations and vary on the conformation of the side chains. Further analysis showed that the ensemble of conformations visited during the various force field trajectories is the same, but their relative occupancy differs, at least for the observed simulation time (Figure 4 and Figure S7, Supporting Information).

3.3. Two Putatively Foldable Peptides. Even a cursory examination of all the rmsd matrices calculated for the folding trajectories presented here (a representative small sample from the approximately 800 calculated matrices overall) allows us to conclude that such small peptides are characterized by multiple folding−unfolding events and remain folded for no more than 20−30% of the observed simulation time even in the prolonged 1 μs trajectories (Figure 4 and Figures S6 and S7 in the Supporting Information). However, a tiny percentage (<1%) of the peptides studied appear to be stably folded. In Figure 4 and Figures S7 and S8 (Supporting Information), we present results for the two best peptides, a pentapeptide (RDKWP) and a tetrapeptide (RWPD).

According to our simulations, the pentapeptide RDKWP remains folded for 50−60% of the trajectory, and that is supported by all force fields and for all temperatures examined here (Figures 2, 3, and 4 and Figure S6, Supporting Information). In Figure 4A,B, we see that two distinct clusters of structures are formed at low temperatures, with the second one being progressively more populated as temperature increases. This cluster of structures (the one shown above the energy landscape in panel B) persists for almost 38% of the simulation time even at the elevated temperature of 360 K. The predicted structures from the trajectories at the various temperatures are closely related with a Cα rmsd of 0.4−0.7 Å and an all-heavy rmsd of 1.1−1.6 Å for the most populated one (the top one in panel C) and a Cα rmsd of 0.1−0.2 Å and an all-heavy rmsd of 0.6−1.2 Å for the less populated structure (bottom one in panel C). This is also evident by the rmsd matrix of the concatenated trajectories of the various temperatures in panel A. Panel C shows the interactions that stabilize each set of structures. In the less populated structure (top one in panel C), we observe the hydrophobic packing of tryptophan's side chain against the lysine's aliphatic chain and a hydrogen bond between arginine's NH group and the free C-terminus of the peptide. In the more populated structure (bottom one in panel C), we observe the hydrophobic packing of tryptophan's side chain against arginine, a hydrogen bond between arginine's NE group and the backbone O of lysine, and a salt bridge between the lysine and the aspartic acid. The well-defined folds adopted by RDKWP are also evident in the Ramachandran plot (Figure S8, Supporting Information) calculated for one of the folding trajectories (2 μs with the AMBER99SB-ildn). Each residue shows a clear preference toward a certain Ramachandran region and mainly in the β and PP$_{II}$ (polyproline) regions.

The shorter tetrapeptides seem to have stable backbone structures and more flexible side chains. This is clearly seen in

their representative, RWPD, in Figure S7 (Supporting Information). RWPD remains folded for approximately 40−50% of the simulation time, depending the temperature of the simulation (Figure S7, Supporting Information, panel A). The observed structures are rather similar, with unique backbone conformation and alternating conformations of the tryptophan and arginine side chains. This behavior persists with the other force fields as well (Figure S7, Supporting Information, panel B). The structures are stabilized mainly by electrostatic interactions between the free terminal ends, resulting in loop backbone structures, and occasionally we see the packing of the tryptophan side chain against the side chain of arginine or proline (Figure S7, Supporting Information, panel C). In most peptides in the tetrapeptide set, the stabilizing interaction comprises mainly the backbone atoms (Figure S4, Supporting Information), which fold into turn-like structures (data not shown). This is also supported by the Ramachandran plots (Figure S8, Supporting Information), wherein the φ/ψ angles remain in the β/PP$_{II}$ region.

4. DISCUSSION

In this work, we presented an attempt to computationally identify foldable peptide sequences. Development of new and efficient bioactive molecules and peptides of pharmacological interest remains, mainly, a matter of serendipity. The present work could be further exploited as a novel method toward the rational design and engineering of small bioactive peptides with stable structures and desired molecular properties. Through an exhaustive set of more than 15 000 independent simulations, putatively foldable peptides were designated by a scoring function based on frame-to-frame rmsd matrices calculated for the folding trajectories. The scoring function was extensively tested and proved successful by revealing stable folders even with the sequence restrictions imposed by us. This scoring function (TF2), because it is solely based on a widely accepted metric measure, could function properly in larger peptide sequences and regardless of the amino acid content.

Our major finding is in full agreement with the literature: most peptides appear to be disordered. The extensive computational work presented here involves only a fraction of sequences from the vast sequence space. Within the perspective of this particular set of 8640 sequences, we conclude that such short peptides are mostly disordered: they are characterized by multiple folding/unfolding events, and their folded population does not exceed 20−30% of the observed simulation time. However, even with the imposition for the presence of bulky residues like tryptophan, arginine, and lysine, some sequences were designated as putatively foldable.

A second conclusion from this work concerns the discrepancies in the force fields' predictions regarding the foldability of such sort sequences. Even within such limited amino acid variance and small number of peptide sequences we see almost every possible scheme from full consistency to full disagreement even among the same family of force fields. For putatively foldable peptides (like RDKWP and RWPD), the various force fields converge on their prediction of a well-defined and stable structure. It is the disordered peptides for which they disagree the most. Even in the absence of actual experimental data, the divergence in the predictions of the force fields regarding the foldability of such short sequences shows the difficulty in the description of the disordered state by current empirical nonpolarizable force fields, at least for small nonpolymer peptides.

■ ASSOCIATED CONTENT

S **Supporting Information**

An example of an rmsd matrix, the evolution of interatomic distances and snapshot structures of a representative tetrapeptide, a graphical representation of the TF2 function, the rmsd matrices of 130 tetrapeptides (second pass in Table 1) calculated from 30 ns trajectories, the rmsd matrices of 36 tetrapeptides (third pass in Table 1) calculated from 100 ns trajectories, the inter-rmsd matrices for 16 pentapeptides and four different force fields (complementary to Figure 3), the rmsd matrices for eight pentapeptides (fourth pass in Table 1) and the AMBER99SB-ildn force field, and results from the simulations of the RWPD tetrapeptide using four temperatures and three force fields. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Tel +30-25510-30620, fax +30-25510-30620, http://utopia.duth.gr/~glykos/, e-mail glykos@mbg.duth.gr.

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E. Peptide Folding: When Simulation Meets Experiment. *Angew. Chem., Int. Ed.* **1999**, *38*, 236−240.

(2) Ferrara, P.; Apostolakis, J.; Caflisch, A. Thermodynamics and Kinetics of Folding of Two Model Peptides Investigated by Molecular Dynamics Simulations. *J. Phys. Chem. B* **2000**, *104*, 5000−5010.

(3) Wu, X.; Wang, S. Folding Studies of a Linear Pentamer Peptide Adopting a Reverse Turn Conformation in Aqueous Solution through Molecular Dynamics Simulation. *J. Phys. Chem. B* **2000**, *104*, 8023−8034.

(4) Yeh, I.-C.; Hummer, G. Peptide Loop-Closure Kinetics from a Microsecond Molecular Dynamics Simulation in Explicit Solvent. *J. Am. Chem. Soc.* **2002**, *124*, 6563−6568.

(5) Gnanakaran, S.; Nymeyer, H.; Portman, J.; Sanbonmatsu, K. Y.; García, A. E. Peptide Folding Simulations. *Curr. Opin. Struct. Biol.* **2003**, *13*, 168−174.

(6) Wei, C.-C.; Ho, M.-H.; Wang, W.-H.; Sun, Y.-C. Molecular Dynamics Simulation of Folding of a Short Helical Peptide with Many Charged Residues. *J. Phys. Chem. B* **2005**, *109*, 19980−19986.

(7) Ho, B. K.; Dill, K. A. Folding Very Short Peptides Using Molecular Dynamics. *PLOS Comput. Biol.* **2006**, *2*, No. e27.

(8) 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.

(9) Snow, C. D.; Nguyen, N.; Pande, V. S.; Gruebele, M. Absolute Comparison of Simulated and Experimental Protein-Folding Dynamics. *Nature* **2002**, *42*, 102−106.

(10) Gnanakaran, S.; Garcia, A. E. Validation of an All-Atom Protein Force Field: From Dipeptides to Larger Peptides. *J. Phys. Chem. B* **2003**, *107*, 12555−12557.

(11) 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.

(12) Aliev, A. E.; Courtier-Murias, D. Experimental Verification of Force Fields for Molecular Dynamics Simulations Using Gly-Pro-Gly-Gly. *J. Phys. Chem. B* **2010**, *114*, 12358−12375.

(13) 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*, No. e32131.

(14) Best, R. B. Atomistic Molecular Simulations of Protein Folding. *Curr. Opin. Struct. Biol.* **2012**, *22*, 52−61.

(15) Bowman, G. R.; Voelz, V. A.; Pande, V. S. Taming the Complexity of Protein Folding. *Curr. Opin. Struct. Biol.* **2011**, *21*, 4−11.

(16) Caflisch, A. Complexity in Protein Folding: Simulation Meets Experiment. *Curr. Phys. Chem.* **2012**, *2*, 4−11.

(17) Edwards, C. M. B.; Cohen, M. A.; Bloom, S. R. Peptides as Drugs. *QJM* **1999**, *92*, 1−4.

(18) Keller, T. H.; Pichota, A.; Yin, Z. A Practical View of 'Druggability'. *Curr. Opin. Chem. Biol.* **2006**, *10*, 357−361.

(19) Kliger, Y. Computational Approaches to Therapeutic Peptide Discovery. *Biopolymers* **2010**, *94*, 701−710.

(20) Borhani, D. W.; Shaw, D. E. The Future of Molecular Dynamics Simulations in Drug Discovery. *J. Comput.-Aided Mol. Des.* **2012**, *26*, 15−26.

(21) Duan, Y.; Kollman, P. A. Pathways to a Protein Folding Intermediate Observed in a 1-Microsecond Simulation in Aqueous Solution. *Science* **1998**, *282*, 740−744.

(22) Shirts, M.; Pande, V. S. Computing: Screen Savers of the World Unite! *Science* **2000**, *290*, 1903−1904.

(23) Zagrovic, B.; Snow, C. D.; Shirts, M. R.; Pande, V. S. Simulation of Folding of a Small Alpha-Helical Protein in Atomistic Detail Using Worldwide-Distributed Computing. *J. Mol. Biol.* **2002**, *323*, 927−937.

(24) Simmerling, C.; Strockbine, B.; Roitberg, A. E. All-Atom Structure Prediction and Folding Simulations of a Stable Protein. *J. Am. Chem. Soc.* **2002**, *124*, 11258−11259.

(25) Pitera, J. W.; Swope, W. Understanding Folding and Design: Replica-Exchange Simulations of "Trp-Cage" Miniproteins. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 7587−7592.

(26) Chowdhury, S.; Lee, M. C.; Xiong, G.; Duan, Y. Ab Initio Folding Simulation of the Trp-Cage Mini-protein Approaches NMR Resolution. *J. Mol. Biol.* **2003**, *327*, 711−717.

(27) Yang, W. Y.; Gruebele, M. Folding at the Speed Limit. *Nature* **2003**, *423*, 193−197.

(28) Ensign, D. L.; Kasson, P. M.; Pande, V. S. Heterogeneity even at the Speed Limit of Folding: Large-Scale Molecular Dynamics Study of a Fast-Folding Variant of the Villin Headpiece. *J. Mol. Biol.* **2007**, *374*, 806−816.

(29) Freddolino, P. L.; Liu, F.; Gruebele, M.; Schulten, K. Ten-Microsecond MD Simulation of a Fast-Folding WW Domain. *Biophys. J.* **2008**, *94*, L75−L77.

(30) Kubelka, J.; Henry, E. R.; Cellmer, T.; Hofrichter, J.; Eaton, W. A. Chemical, Physical and Theoretical Kinetics of an Ultrafast Folding Protein. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 18655−18662.

(31) 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.

(32) Cellmer, T.; Buscaglia, M.; Henry, E. R.; Hofrichter, J.; Eaton, W. A. Making Connections between Ultrafast Protein Folding Kinetics and Molecular Dynamics Simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *108*, 6103−6108.

(33) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, E. How Fast-Folding Proteins Fold. *Science* **2011**, *334*, 517−520.

(34) Honda, S.; Yamasaki, K.; Sawada, Y.; Morii, H. 10 Residue Folded Peptide Designed by Segment Statistics. *Structure* **2004**, *12*, 1507−1518.

(35) Hatfield, M. P. D.; Murphy, R. S.; Lovas, S. VCD Spectroscopic Properties of the β-Hairpin Forming Miniprotein CLN025 in Various Solvents. *Biopolymers* **2010**, *93*, 442−450.

(36) Kier, B. L.; Andersen, N. H. Probing the Lower Size-Limit for Protein-Like Fold Stability: Ten-Residue Microproteins with Specific, Rigid Structures in Water. *J. Am. Chem. Soc.* **2008**, *130*, 14675−14683.

(37) Simmerling, C. L.; Elber, R. Computer Determination of Peptide Conformations in Water: Different Roads to Structure. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92*, 3190−3193.

(38) Demchuck, E.; Bashford, D.; Case, D. A. Dynamics of a Type VI Reverse Turn in a Linear Peptide in Aqueous Solution. *Folding Des.* **1997**, *2*, 35–46.

(39) Fuchs, P. F. J.; Bonvin, A. M. J. J.; Bochicchio, B.; Pepe, A.; Alix, A. J. P.; Tamburro, A. M. Kinetics and Thermodynamics of Type VIII β-Turn Formation: A CD, NMR, and Microsecond Explicit Molecular Dynamics Study of the GDNP Tetrapeptide. *Biophys. J.* **2006**, *90*, 2745–2759.

(40) Flöck, D.; Rossetti, G.; Daidone, I.; Amadei, A.; Di Nola, A. Aggregation of Small Peptides Studied by Molecular Dynamics Simulations. *Proteins* **2006**, *65*, 914–921.

(41) Zanuy, D.; Flores-Ortega, A.; Casanovas, J.; Curcó, D.; Nussinov, R.; Alemán, C. The Energy Landscape of a Selective Tumor-Homing Pentapeptide. *J. Phys. Chem. B* **2008**, *112*, 8692–8700.

(42) Khavinson, V.Kh.; Malinin, V. V.; Grigoriev, E. I.; Ryzhak, G. A. Tetrapeptide Regulating Blood Glucose Level in Diabetes Mellitus. U.S. Patent 7,491,703 B2, Feb. 17, 2009.

(43) Adhikari, B.; Banerjee, A. Self-Assembling Peptides: From Molecules to Nanobiomaterials. *J. Indian Inst. Sci.* **2011**, *91*, 471–483.

(44) Kirshenbaum, K.; Zuckermann, R. N.; Dill, K. A. Designing Polymers That Mimic Biomolecules. *Curr. Opin. Struct. Biol.* **1999**, *9*, 530–535.

(45) Lewis, P. N.; Momany, F. A.; Scheraga, H. A. Folding of Polypeptide Chains in Proteins: A Proposed Mechanism of Folding. *Proc Natl. Acad. Sci. USA* **1971**, *68*, 2293–2297.

(46) Zimmerman, S. S.; Scheraga, H. A. Local Interactions in Bends of Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **1977**, *74*, 4126–4129.

(47) Rackovsky, S. On the Nature of the Protein Folding Code. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90*, 644–648.

(48) Feng, Y.; Luo, L. Use of Tetrapeptide Signals for Protein Secondary-Structure Prediction. *Amino Acids* **2008**, *35*, 607–641.

(49) Meus, J.; Brylinski, M.; Piwowar, M.; Piwowar, P.; Wisniowski, Z.; Stefaniak, J.; Konieczny, L.; Surowka, G.; Roterman, I. A Tabular Approach to the Sequence-to-Structure Relation in Proteins (Tetrapeptide Representation) for *de Novo* Protein Design. *Med. Sci. Monit.* **2006**, *12*, BR208–214.

(50) Dallüge, R.; Oschmann, J.; Birkenmeier, O.; Lücke, C.; Lilie, H.; Rudolph, R.; Lange, C. A Tetrapeptide Fragment-Based Design Method Results in Highly Stable Artificial Proteins. *Proteins* **2007**, *68*, 839–849.

(51) Chiti, F.; Calamai, M.; Taddei, N.; Stefani, M.; Ramponi, G.; Dobson, C. M. Studies of the Aggregation of Mutant Proteins *in Vitro* Provide Insights into the Genetics of Amyloid Diseases. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 16419–16426.

(52) Schwartz, R.; King, J. Frequencies of Hydrophobic and Hydrophilic Runs and Alterations in Proteins of Known Structure. *Protein Sci.* **2006**, *15*, 102–112.

(53) Zbilut, J. P.; Chua, G. H.; Krishman, A.; Bossa, C.; Colafranceschi, M.; Giuliani, A. Entropic Criteria for Protein Folding Derived from Recurrences: Six Residues Patch as the Basic Protein Word. *FEBS Lett.* **2006**, *580*, 4861–4864.

(54) Mahalakshmi, R.; Sengupta, A.; Raghothama, S.; Shamala, N.; Balaram, P. Tryptophan Rich Peptides: Influence of Indole Rings on Backbone Conformation. *Biopolymers* **2006**, *88*, 36–54.

(55) Eidenschink, L.; Kier, B. L.; Huggins, K. N. L.; Andersen, N. H. Very Short Peptides with Stable Folds: Building on the Interrelationship of Trp/Trp, Trp/Cation, and Trp/Backbone-Amide Interaction Geometries. *Proteins* **2009**, *75*, 308–322.

(56) Greenfield, N. J. Using Circular Dichroism Spectra to Estimate Protein Secondary Structure. *Nat. Protoc.* **2006**, *1*, 2876–2890.

(57) Glättli, A.; Daura, X.; Bindschädler, P.; Jaun, B.; Mahajan, Y. R.; Mathad, R. I.; Rueping, M.; Seebach, D.; vanGunsteren, W. F. On the Influence of Charged Side Chains on the Folding-Unfolding Equilibrium of B-Peptides: A Molecular Dynamics Simulation Study. *Chem.—Eur. J.* **2005**, *11*, 7276–7293.

(58) Kale, L.; Skeel, R.; Bhandarkar, M.; Brunner, R.; Gursoy, A.; Krawetz, N.; Phillips, J.; Shinozaki, A.; Varadarajan, K.; Schulten, K.

(59) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(60) MacKerell, A. D., Jr.; Feig, M.; Brooks, C. L., III. Extending the Treatment of Backbone Energetics in Protein Force Fields: Limitations of Gas-Phase Quantum Mechanics in Reproducing Protein Conformational Distributions in Molecular Dynamics Simulations. *J. Comput. Chem.* **2004**, *25*, 1400–1415.

(61) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.

(62) Kaminski, G.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. *J. Phys. Chem. B* **2001**, *105*, 6474–6487.

(63) 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.

(64) 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.

(65) 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.

(66) Guvench, O.; MacKerell, A. D., Jr. Comparison of Protein Force Fields for Molecular Dynamics Simulations. *Methods Mol. Biol.* **2008**, *443*, 63–88.

(67) Lange, O. F.; van der Spoel, D.; de Groot, B. L. Scrutinizing Molecular Mechanics Force Fields on the Submicrosecond Timescale with NMR Data. *Biophys. J.* **2010**, *99*, 647–655.

(68) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33–38.

(69) 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.

(70) 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.

(71) Izaguirre, J. A.; Reich, S.; Skeel, R. D. Longer Time Steps for Molecular Dynamics. *J. Chem. Phys.* **1999**, *110*, 9853–9864.

(72) Darden, T.; York, D.; Pedersen, L. An N log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(73) Ryckaert, J. P.; Ciccoti, G.; Berendsen, H. J. C. Numerical Integration of Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.

(74) Bashford, D.; Case, D. A.; Choi, C.; Gippert, G. P. A Computational Study of the Role of Solvation Effects in Reverse Turn Formation in Tetrapeptides APGD and APGN. *J. Am. Chem. Soc.* **1997**, *119*, 4964–4971.

(75) Feige, M. J.; Paci, E. Rate of Loop Formation in Peptides: A Simulation Study. *J. Mol. Biol.* **2008**, *382*, 556–565.

(76) Eaton, W. A.; Munoz, V.; Thompson, P. A.; Henry, E. R.; Hofrichter, J. Kinetics and Dynamics of Loops, α-Helices, β-Hairpins and Fast-Folding Proteins. *Acc. Chem. Res.* **1998**, *31*, 745–753.

(77) Bieri, O.; Wirz, J.; Hellrung, B.; Schutkowski, M.; Drewello, M.; Kiefhaber, T. The Speed Limit for Protein Folding Measured by

NAMD2: Greater Scalability for Parallel Molecular Dynamics. *J. Comput. Phys.* **1999**, *151*, 283–312.

Triplet-Triplet Energy Transfer. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 9597−9601.

(78) Kubelka, J.; Hofrichter, J.; Eaton, W. A. The Protein Folding 'Speed Limit'. *Curr. Opin. Struct. Biol.* **2004**, *14*, 76−88.

(79) Plaxco, K. W.; Dobson, C. M. Time-Resolved Biophysical Methods in the Study of Protein Folding. *Curr. Opin. Struct. Biol.* **1996**, *6*, 630−636.

(80) Lapidus, L. J.; Eaton, W. A.; Hofrichter, J. Measuring the Rate of Intramolecular Contact Formation in Polypeptides. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 7998−8002.

(81) Portman, J. J. Non-Gaussian Dynamics from a Simulation of a Short Peptide: Loop Closure Rates and Effective Diffusion Coefficients. *J. Chem. Phys.* **2003**, *118*, 2381−2391.

(82) Krieger, F.; Fierz, B.; Bieri, O.; Drewello, M.; Kiefhaber, T. Dynamics of Unfolded Polypeptide Chains As Model for the Earliest Steps in Protein Folding. *J. Mol. Biol.* **2003**, *332*, 265−274.

(83) Yeh, I.-C.; Wallqvist, A. Structure and Dynamics of End-to-End Loop Formation of the Penta-peptide Cys-Ala-Gly-Gln-Trp in Implicit Solvents. *J. Phys. Chem. B* **2009**, *113*, 12382−12390.

(84) Daidone, I.; Neuweiler, H.; Doose, S.; Sauer, M.; Smith, J. C. Hydrogen-Bond Driven Loop-Closure Kinetics in Unfolded Polypeptide Chains. *PLOS Comput. Biol.* **2010**, *6*, No. e1000645.

(85) Lin, M. M.; Mohammed, O. F.; Jas, G. S.; Zewail, A. H. Speed Limit of Protein Folding Evidenced in Secondary Structure Dynamics. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 16622−16627.

(86) De Sancho, D.; Best, R. B. What Is the Time Scale for α-Helix Nucleation? *J. Am. Chem. Soc.* **2011**, *133*, 6809−6816.

(87) Jas, G. S.; Hegefeld, W. A.; Májek, P.; Kuczera, K.; Elber, R. Experiments and Comprehensive Simulations of the Formation of a Helical Turn. *J. Phys. Chem. B* **2012**, *116*, 6598−6610.

(88) Schuetz, P.; Wuttke, R.; Schuler, B.; Caflisch, A. Free Energy Surfaces from Single-Distance Information. *J. Phys. Chem. B* **2010**, *114*, 15227−15235.

(89) Gustiananda, M.; Liggins, J. R.; Cummins, P. L.; Gready, J. E. Conformation of Prion Protein Repeat Peptides Probed by FRET. *Biophys. J.* **2004**, *86*, 2467−2483.

(90) Stultz, C. M.; Levin, A. D.; Edelman, E. R. Phosporylation-Induced Conformational Changes in a Mitogen-Activated Protein Kinase Substrate. Implications for Tyrosine Hydroxylase Activation. *J. Biol. Chem.* **2002**, *277*, 47653−47661.

(91) Schuler, B.; Lipman, E. A.; Steinbach, P. J.; Kumke, M.; Eaton, W. A. Polyproline and the "Spectroscopic Ruler" Revisited with Single-Molecule Fluorescence. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 2754−2759.

(92) Glykos, N. M. CARMA: A Molecular Dynamics Analysis Program. *J. Comput. Chem.* **2006**, *27*, 1765−1768.

(93) Ichiye, T.; Karplus, M. Collective Motions in Proteins: A Covariance Analysis of Atomic Fluctuations in Molecular Dynamics and Normal Mode Simulations. *Proteins* **1991**, *11*, 205−217.

(94) Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. Essential Dynamics of Proteins. *Proteins* **1993**, *17*, 412−425.

(95) 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.

(96) Altis, A.; Nquyen, P. H.; Hegger, R.; Stock, G. Dihedral Angle Principal Component Analysis of Molecular Dynamics Simulations. *J. Chem. Phys.* **2007**, *126*, No. 244111.

(97) Buchete, N. V.; Hummer, G. Coarse Master Equations for Peptide Folding Dynamics. *J. Phys. Chem. B* **2008**, *112*, 6057−6069.

(98) Wang, T.; Zhu, Y.; Getahun, Z.; Du, D.; Huang, C.-Y.; DeGrado, W. F.; Gai, F. Length Dependent Helix-Coil Transition Kinetics of Nine Alanine-Based Peptides. *J. Phys. Chem. B* **2004**, *108*, 15301−15310.

(99) Berezhkovskii, A. M.; Tofoleanu, F.; Buchete, N. V. Are Peptides Good Two-State Folders? *J. Chem. Theory Comput.* **2011**, *7*, 2370−2375.

(100) Yoda, T.; Sugita, Y.; Okamoto, Y. Secondary-Structure Preferences of Force Fields for Proteins Evaluated by Generalized-Ensemble Simulations. *Chem. Phys.* **2004**, *307*, 269−283.

(101) Steinbach, P.-J. Exploring Peptide Energy Landscapes: A Test of Force Fields and Implicit Solvent Models. *Proteins* **2004**, *57*, 665−677.

(102) Best, R. B.; Buchete, N. V.; Hummer, G. Are Current Molecular Dynamics Force Fields Too Helical? *Biophys. J.* **2008**, *95*, L07−09.

(103) 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.

(104) Freddolino, P. L.; Park, S.; Roux, B.; Schulten, K. Force Field Bias in Protein Folding Simulations. *Biophys. J.* **2009**, *96*, 3772−3780.

(105) Best, R. B.; Mittal, J. Balance between Alpha and Beta Structures in *ab Initio* Protein Folding. *J. Phys. Chem. B* **2010**, *114*, 8790−8798.

(106) Mittal, J.; Best, R. B. Tackling Force Field Bias in Protein Folding Simulations: Folding of Villin HP35 and Pin WW Domains in Explicit Water. *Biophys. J.* **2010**, *99*, L26−L28.

(107) 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.

(108) Patapati, K. K.; Glykos, N. M. Three Force Fields' Views of the 3_{10} Helix. *Biophys. J.* **2011**, *101*, 1766−1771.

(109) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. How Robust Are Protein Folding Simulations with Respect to Force Field Parameterization? *Biophys. J.* **2011**, *100*, L47−L49.