Title
Folding very short peptides using molecular dynamics.

Permalink
https://escholarship.org/uc/item/3m53q1xp

Journal
PLoS computational biology, 2(4)

ISSN
1553-734X

Authors
Ho, Bosco K
Dill, Ken A

Publication Date
2006-04-14

DOI
10.1371/journal.pcbi.0020027

Peer reviewed
Folding Very Short Peptides Using Molecular Dynamics

Bosco K. Ho*, Ken A. Dill

Department of Pharmaceutical Chemistry, University of California San Francisco, San Francisco, California, United States of America

Peptides often have conformational preferences. We simulated 133 peptide 8-mer fragments from six different proteins, sampled by replica-exchange molecular dynamics using Amber7 with a GB/SA (generalized-Born/solvent-accessible electrostatic approximation to water) implicit solvent. We found that 85 of the peptides have no preferred structure, while 48 of them converge to a preferred structure. In 85% of the converged cases (41 peptides), the structures found by the simulations bear some resemblance to their native structures, based on a coarse-grained backbone description. In particular, all seven of the β hairpins in the native structures contain a fragment in the turn that is highly structured. In the eight cases where the bioinformatics-based I-sites library picks out native-like structures, the present simulations are largely in agreement. Such physics-based modeling may be useful for identifying early nuclei in folding kinetics and for assisting in protein-structure prediction methods that utilize the assembly of peptide fragments.

Introduction

Peptide fragments of proteins often have intrinsic propensities for the formation of their native conformations. For example, NMR experiments [1] show that long peptide fragments have native-like conformations [2–7]. Some short peptides in solution have also been shown to adopt their native secondary structures: α helices [8,9] and β hairpins [10–14].

As a consequence, peptide conformational propensities that are taken from the protein database (PDB) [1–17] are now widely used in protein-structure prediction algorithms. A popular set of peptide fragment conformations is the I-sites library of David Baker and his co-workers [18,19]. Extensive libraries of peptide fragments have now been compiled [20–22] and have become essential elements in protein-prediction methods [23]. From the recent CASP protein-structure prediction competition, it was noted that most of the successful de novo methods use a fragment-based approach [23,24]. Typically, a candidate protein native structure is spliced together from fragments that are extracted from a database of conformations, and then treated to conformational scoring and optimization.

Can physical models capture these conformational propensities of peptides? There is good evidence that they can. First, simple physical models can reproduce the structural biases of certain peptide fragments [25–28]. To date, however, such studies have largely focused on selected peptides that are expected to fold. Our interest here is to know whether physical models can also discriminate peptides that fold from peptides that do not. Second, in molecular dynamics simulations of small peptides, the ensemble of conformers divides into well-defined clusters. This has been found for a penta-β peptide in explicit water [29,30], and for a small α-helical peptide [31]. Third, molecular dynamic simulations of small peptides reproduce the α-helical propensities of certain fragments from the I-sites sequence-structure library [32]. Many models of protein folding kinetics assume that peptide fragments of the chain that have preferred conformations are responsible for nucleating the folding process [33–35].

Here, we study 133 peptide 8-mer fragments from six different proteins of different folds, using replica-exchange molecular dynamics sampling [36] in Amber7, with the parm96 parameters and the GB/SA (generalized-Born/solvent-accessible electrostatic approximation to water) implicit solvent model of Tsui and Case [37]. We chose this force field as it is the only implicit-solvation model that can adequately reproduce the native state of the β hairpin of protein G [38].

We are interested in whether this physical model can identify native-like secondary structures in peptide fragments. If so, it indicates the importance of local interactions in those cases. Our study involves complete coverage of those proteins. For each protein, we systematically generate a series of 8-mer peptide fragments with overlapping sequences from the original protein sequence. Neighboring fragments have a five-residue overlap (and three-residue gap). We chose 8-mers because this length appears adequate to identify elements of structure in PDB studies [19] and because much longer fragments become too expensive for computer simulations. We simulate each peptide using 16 replicas for 5 ns/replica, and keep only the last 1 ns.

In each case, we determine whether the peptide has converged to its native conformation in the folded protein. We consider two measures of convergence. First, we monitor the RMSD between the simulated conformations and the experimental PDB structure of that peptide. However, for
To carry out specific biochemical reactions, proteins must adopt precise three-dimensional conformations. During the folding of a protein, the protein picks out the right conformation out of billions of other conformations. It is not yet possible to do this computationally. Picking out the native conformation using physics-based atomically detailed models, sampled by molecular dynamics, is presently beyond the reach of computer methods. How can we speed up computational protein-structure prediction? One idea is that proteins start folding at specific parts of a chain that kink up early in the folding process. If we can identify these kinks, we should be able to speed up protein-structure prediction. Previous studies have identified likely kinks through bioinformatic analysis of existing protein structures. The goal of the authors here is to identify these putative folding initiation sites with a physical model instead. In this study, Ho and Dill show that, by chopping a protein chain into peptide pieces, then simulating the pieces in molecular dynamics, they can identify those peptide fragments that have conformational biases. These peptides identify the kinks in the protein chain.

Structural Bias in the Peptide Conformation Ensemble

A mesostring is a one-dimensional list of the mesostates of each residue in a peptide. A mesostate refers to a discrete region of the φ-ψ angles of the backbone of a residue. Mesostate [a] corresponds to a helical conformation, including the α-helix, 310-helix, or π-helix. Mesostate [b] corresponds to an extended β-strand conformation. Mesostate [l] corresponds to a left-handed helical conformation.

We use the mesostrings to cluster conformations in our simulations. Based on the three mesostates described above, an 8-mer has 3^8 = 6,561 possible mesostrings. When each simulation is completed, each 8-mer peptide will have different populations for the 6,561 mesostrings, hence different free energies. The mesostring that represents the highest population (the lowest free energy) is called the ground mesostring. We use the properties of the ground mesostring to determine structural bias in a peptide. The ground mesostrings are classified in terms of either a reverse-turn or a helical-turn conformation (see Figure 1). We define a helical-turn as a mesostring that contains at least four [a] mesostates in a row, and a reverse-turn as a mesostring that contains either the [bab] or [bba] motifs.

How do we know when a simulation has converged? We calculate the backbone entropy using the Boltzmann formula: \( S = -k \sum_i p_i \ln p_i \) where \( p_i \) is the probability that the peptide is in mesostring \( i \). The backbone entropy is calculated over a certain window in a trajectory, where the sum is made over only the mesostrings that are observed in the window. The backbone entropy \( S \) is useful for two purposes. First, it measures for a given peptide the sharpness of the distribution of probabilities of the mesostrings. The more peaked the distribution is, and thus the more favored a mesostring is, the lower is the backbone entropy. In this way, the backbone entropy indicates whether any one conformation is substantially favored over the others, for the given peptide. Second, the backbone entropy should converge at equilibrium, approaching an asymptotic value with time in the simulation.

Even if a new mesostring emerges late within the sampling (as is often the case), it only changes the backbone entropy if it has a significant population. We use the convergence of the backbone entropy to indicate the convergence of the simulation.

We study peptide fragments extracted from a series of well-characterized proteins: protein G, protein L, protein A, and α-spectrin, and chymotrypsin inhibitor. For each peptide, we simulate the ensemble of states at equilibrium. We find that some of these peptides exhibit strong structural biases. We analyze the relationship of those structural biases to the topology of the native structure.

**Results/Discussion**

Structural Bias in the Peptide Conformation Ensemble

Do peptides have native-like conformations? Figure 2 shows the simulated free-energy profiles of RMSD for the peptides of protein G. We call the region of RMSD < 2 Å native-like. We find that some fragments spend a significant amount of time near their native structures (seq3, seq9, and seq10). Some peptides have a broad conformational distribution (seq14), while others have a narrow distribution (seq16). Narrow distributions indicate structural bias in the peptide. To investigate this structural bias further, we list in Table 1 the lowest free-energy mesostrings of several protein
G peptides. We show in Figure 1, a representative conformation of the ground mesostrings of these peptides.

Figure 3 shows the variation in backbone entropy for the peptides of protein G. To calculate the variation in Figure 3, we deliberately chose a smaller window (0.2 ns) than the window used for the analysis (1 ns in Tables 1–4) to emphasize the fluctuations. In most of the peptides, the backbone entropy equilibrates almost immediately, with the exception of seq16, which decreases to a near zero value at about 3.5 ns. Consequently, we carry out the main analysis of the structural bias over the last 1 ns of our 5-ns trajectories. The backbone entropy specifically measures the conformation freedom in the backbone. Backbone entropy is a useful measure only when the free-energy basins in phase space are dominated by the local conformation of the backbone, and not by nonlocal interactions. As these peptides are short, nonlocal interactions should be minimal, and the backbone entropy should be the dominant entropy.

We define the existence of structural bias in a peptide in terms of two properties of the ground mesostring. First, we use the probability $P_1$ in observing the ground mesostring, which is derived from the relative free energies. Second, we use the free-energy gap $\Delta F$ between the ground mesostring and the next mesostring to measure the relative probability of the ground mesostring from all the other mesostrings. Specifically, we consider a peptide to have structural bias if $P_1 > 45\%$ and $\Delta F > 0.6$ kcal/mol. Of the 133 peptides we studied, we found that 48 peptides have structural bias (bold in Tables 2–4). We refer to such peptides as structured peptides.

Comparison of the Peptide Conformations with Native Structures

What parts of the native structure are picked out by the structured peptides? In Table 5, we list the ground mesostrings of the peptides in simulation. We highlight (in bold) the sequences that are structured and compare these structured peptides to the native secondary structures. The structured peptides adopt either a helical-turn or reverse-turn. Figure 4 shows the location of the structured peptides within the native fold topology. Below we describe the relationship between the structured peptides, the native structure, and experimental studies of the folding of these proteins.

In the protein G fragments, we find eight structured peptides that adopt a stable helical-turn conformation (Table...
Table 1. Mesosstrings of Various Peptides from Protein G

| Peptide | Mesosstring | Free Energy of the Mesostring in kcal/mol | P in Percent |
|---------|-------------|------------------------------------------|--------------|
| seq1: 1-MTYKULN | bbbabbbb | -3.09 | 31 |
| | babaaaaa | -2.59 | 12 |
| | baaaabbb | -2.54 | 11 |
| | baaaabb | -2.51 | 11 |
| | babaabb | -2.31 | 7 |
| | babaabba | -2.26 | 7 |
| | bbbabbb | -3.29 | 44 |
| seq2: 1-NGKTLKG | bbb1aabbb | -2.77 | 16 |
| | abaaabbb | -2.50 | 10 |
| | bbbaabbb | -2.03 | 4 |
| | ablababb | -1.88 | 3 |
| seq3: 25-TAEKVFKQ | baaaaaaa | -3.16 | 39 |
| | baaaaaa | -3.00 | 26 |
| | abaaaaaa | -2.53 | 11 |
| | baaaaa | -2.09 | 5 |
| | aaaaaab | -1.97 | 4 |
| | bbabbaa | -1.87 | 3 |
| | bbaabbb | -1.69 | 92 |
| | abababa | -1.71 | 2 |
| | aaaaabb | -1.42 | 1 |
| | baaaaaa | -1.39 | 1 |

In summary, of the 48 structured peptides found in the simulations, only five differ significantly from the native structure. Given that there are 436 residues in our six
Conclusion

In this study, we have applied replica-exchange molecular dynamics, using the parm96 force field with a GB/SA solvent model, to the simulation of 133 peptide 8-mer fragments, extracted from six proteins with five different folds. We found that 48 of these peptides are strongly structured. The remaining 85 peptides have no preferred structure. Of the 48 that are structured, 41 of them fold into approximately their native conformations. In seven instances, the simulated structures are significantly inconsistent with their native structures.

Why are only 35% of the peptides structured? The reason is that by using very short peptides, we have eliminated most of the nonlocal interactions—hydrophobic clustering, cooperative helical hydrogen bonds. We thus attribute any structural bias to sidechain interactions, which will depend on specific sequence motifs.

As with all molecular dynamics simulations, the results will

Table 2. Ground Mesostrings of Protein G and Protein L

| Protein | Peptide | Sequence | RMSD in Å | Mesosring | P_1 in Percent | ΔF in kcal/mol | TS in kcal/mol | Native Structure |
|---------|---------|----------|-----------|-----------|----------------|----------------|----------------|-----------------|
|         |         |          | Native    | Ground    |                |                |                |                 |
| Protein G | seq1 | 1-WTYKLILN | 5.8 | bbbbaaaab | bbbbaaabb | 31% | 0.59 | 1.27 |                 |
|          | seq2 | 4-KLLNGKT  | 5.0 | bbbbaaaab | bbbbaaaab | 41% | 0.53 | 1.37 |                 |
|          | seq3 | 7-LNGLTEKL | 3.0 | baaaabbb | baaaabbb | 60% | 0.97 | 1.20 | hairpin-turn |
|          | seq4 | 10-KTLKGTT | 4.5 | aaaaabaa | aaaaabaa | 70% | 1.28 | 1.03 | turn-strand |
|          | seq5 | 13-KGETTTEA | 6.0 | bbbbaaaa | bbbbaaaa | 18% | 0.35 | 1.57 |                 |
|          | seq6 | 16-TTTTEAVD | 4.1 | bbbbaaab | aaaaaaa | 48% | 0.67 | 1.33 | turn |
|          | seq7 | 19-EAVDAATA | 3.8 | baabaaa | ababaaa | 37% | 0.15 | 1.19 |                 |
|          | seq8 | 22-DAATAEKV | 3.6 | baaaabaa | baaaabaa | 64% | 0.79 | 1.00 | helix |
|          | seq9 | 25-TAEKVFQ | 1.4 | aaaaabaa | aaaaabaa | 61% | 0.93 | 1.11 | helix |
|          | seq10 | 28-KVFKQ | 0.5 | aaaaaaab | aaaaaaab | 40% | 0.76 | 1.39 |                 |
|          | seq11 | 31-KQVANDNG | 2.5 | aaaaaa | baaaabaa | 58% | 1.02 | 1.19 | helix-cap |
|          | seq12 | 34-ANDNGVDG | 2.4 | aaasab | baaaab | 19% | 0.65 | 2.08 |                 |
|          | seq13 | 37-NVGDGWT | 5.1 | albbabbb | baabaa | 37% | 1.08 | 1.74 |                 |
|          | seq14 | 40-GVTDYDD | 4.2 | abababbb | abababbb | 10% | 0.11 | 1.93 |                 |
|          | seq15 | 43-WTTDDKAT | 3.1 | baabaa | ababab | 64% | 1.19 | 1.01 | strand-turn |
|          | seq16 | 46-DATKTFTE | 3.8 | baabaa | ababaa | 94% | 2.28 | 0.24 | hairpin-turn |
|          | seq17 | 49-TKTPYTFF | 6.0 | abbbbaa | abbbbaa | 23% | 0.01 | 1.25 |                 |
| Protein L | seq1 | 1-KANLIFAN | 4.3 | baababbb | babaaab | 42% | 0.20 | 1.01 |                 |
|          | seq2 | 4-LIFANGST | 2.5 | bbaaabb | baabaab | 55% | 0.83 | 1.19 | hairpin-turn |
|          | seq3 | 7-ANGSTQTYA | 5.0 | albbbab | babaaab | 31% | 0.44 | 1.46 |                 |
|          | seq4 | 10-TEEQAFFX | 6.7 | bbbbaaab | bbaaab | 45% | 0.51 | 1.11 |                 |
|          | seq5 | 13-TAEFKTTF | 5.7 | bbbbaabb | abaaab | 37% | 0.99 | 1.73 |                 |
|          | seq6 | 16-FKGFTEKA | 2.5 | bbbbaa | bbaabab | 14% | 0.01 | 1.54 |                 |
|          | seq7 | 19-TFKEKSAE | 3.8 | babaabaa | babaaab | 39% | 0.23 | 1.06 |                 |
|          | seq8 | 22-KATSEAYA | 4.6 | aaaaaaab | aaaaaaab | 51% | 0.67 | 1.03 | cap-helix |
|          | seq9 | 25-SEAYAYAD | 4.2 | aaaaaaab | aaaaaaab | 22% | 0.13 | 1.44 |                 |
|          | seq10 | 28-FAYADTFX | 4.2 | aaaaaaab | bbaaabbb | 19% | 0.11 | 1.42 |                 |
|          | seq11 | 31-ADTLKEDN | 2.4 | aababa | bbaaabaa | 82% | 1.16 | 0.56 | helix-cap |
|          | seq12 | 34-LKNDGEY | 2.4 | abaaab | bbaabb | 68% | 1.28 | 0.96 | turn |
|          | seq13 | 37-DNGEYTVD | 5.5 | albabbb | baababab | 30% | 0.23 | 1.53 |                 |
|          | seq14 | 40-EYTYVDAD | 4.4 | lbbbbb | bbaaaba | 44% | 0.78 | 1.23 |                 |
|          | seq15 | 43-DVDAKEGY | 3.8 | bbbbaa | bbaaabb | 51% | 0.97 | 1.25 | hairpin-turn |
|          | seq16 | 46-DAVYTLN | 3.2 | blllaabb | bbbbaabb | 31% | 0.61 | 1.59 |                 |
|          | seq17 | 49-GYTLNIKFGA | 6.9 | lbbabbb | bababbb | 15% | 0.18 | 2.03 |                 |

RMSD is the most likely value of RMSD extracted from the free-energy profile of RMSD. The ground mesostring is sometimes nearly identical to less-populated mesostrings. If the most populated mesostrings differ by only one mesostate, we group them into a consensus mesostring, which contains one indefinite mesostate signified by [—].

P_1 is the probability of the ground mesostring.

ΔF is the free-energy difference between the ground mesostring and the next mesostring.

TS is the entropy of the mesostrings.

Native Structure is the description of the structure of the peptide in the native structure.

Bolded lines highlight structured peptides: P_1 > 45%, and ΔF > 0.6 kcal/mol.

DOI: 10.1371/journal.pcbi.0020027.t002

Proteins, there is, on average, a kink (secondary structural indicator) approximately every nine residues along the chain.

Comparison with the I-Sites Library

Do the structural biases that are found in our simulations correlate with those in the PDB? We focus on the I-sites server (http://www.bioinfo.rpi.edu/~bystrc/hmmstr/server.php), a fragment database that predicts the structures of short sequence motifs. In that database, predictions that have a high confidence score (>0.8) are found to predict a structure that is <1.4 Å from the native structure with a 74% probability. I-sites make eight such high-confidence predictions over four of the proteins in our dataset. Table 3 shows those successes of I-sites. Our structured peptides overlap with the I-sites predictions in six of the eight I-sites predictions. This suggests that the I-sites sequence-structure correlations are at least partly encoded in the local structural biases found in the structured peptides.
| Protein | Peptide | Sequence | RMSD in Å | ΔF in kcal/mol | TS in kcal/mol | Native Structure |
|---------|---------|----------|-----------|----------------|----------------|-----------------|
| Protein A | seq1 | 1-QOQAFYVEI | 3.7 | 0.16 | 1.08 | helix |
| | seq2 | 4-AYFIELHL | 3.9 | 0.15 | 1.11 | helix |
| | seq3 | 7-ELHLPNL | 3.3 | 0.34 | 1.10 | helix |
| | seq4 | 10-HLPNNEE | 3.9 | 0.33 | 1.21 | helix |
| | seq5 | 13-NNLEQQRN | 2.9 | 0.28 | 1.14 | helix |
| | seq6 | 16-EQRQRNGF | 4.0 | 0.66 | 0.96 | helix |
| | seq7 | 19-HQFELQSL | 3.5 | 0.53 | 1.83 | helix |
| | seq8 | 22-FQGLXCD | 3.7 | 0.24 | 1.24 | helix |
| | seq9 | 25-ZRDPQQSQ | 3.7 | 0.18 | 1.62 | helix |
| | seq10 | 28-DFPSQAN | 1.6 | 0.35 | 0.92 | helix |
| | seq11 | 31-SQANLILA | 3.8 | 0.06 | 1.15 | helix |
| | seq12 | 34-ANLLEAARK | 3.5 | 0.33 | 1.63 | helix |
| | seq13 | 37-LAEEKKLMDA | 1.9 | 0.28 | 0.60 | helix |
| | seq1 | 1-VLSFQEGW | 3.9 | 0.98 | 1.53 | helix |
| | seq2 | 4-SKGEQWLV | 3.9 | 0.88 | 1.41 | helix |
| | seq3 | 7-EWQLVH | 1.7 | 0.58 | 0.76 | helix |
| | seq4 | 10-LVHWAQ | 0.6 | 0.78 | 1.22 | helix |
| | seq5 | 13-HWAKVEA | 4.1 | 0.13 | 1.37 | helix |
| | seq6 | 16-AVXAIQA | 4.0 | 0.05 | 1.60 | helix |
| | seq7 | 130-ADAVADGE | 4.1 | 0.31 | 1.71 | helix |
| | seq8 | 22-VAGHQJHOQ | 3.4 | 0.57 | 1.62 | helix |
| | seq9 | 25-HQODILIR | 3.4 | 1.31 | 1.17 | helix |
| | seq10 | 28-DILIRFLK | 4.5 | 0.51 | 0.88 | helix |
| | seq11 | 31-IRLFKSPH | 4.1 | 1.06 | 0.81 | helix |
| | seq12 | 34-FKSHPETL | 2.0 | 0.12 | 1.23 | helix |
| | seq13 | 37-HPETLELF | 1.4 | 1.11 | 0.76 | helix |
| | seq14 | 40-PEEPORPF | 3.1 | 1.66 | 0.93 | helix |
| | seq15 | 43-KDRFQKHL | 3.5 | 0.13 | 0.96 | helix |
| | seq16 | 46-RFKHKLTE | 4.3 | 0.28 | 1.06 | helix |
| | seq17 | 49-HLKTEOEN | 3.1 | 1.07 | 1.18 | turn |
| | seq18 | 52-TEAKEMAS | 3.0 | 0.16 | 1.41 | helix |
| | seq19 | 55-GKEASDL | 3.0 | 0.49 | 1.55 | helix |
| | seq20 | 58-AEDELKKA | 1.9 | 0.28 | 1.10 | helix |
| | seq21 | 61-DLKAQGVT | 3.0 | 1.00 | 1.30 | helix |
| | seq22 | 64-EGTVTLT | 4.3 | 0.11 | 1.47 | helix |
| | seq23 | 67-VTVVTLAG | 3.7 | 0.75 | 1.48 | helix |
| | seq24 | 70-ILALGIR | 3.0 | 0.87 | 1.80 | helix |
| | seq25 | 73-LGAILKKE | 3.5 | 0.99 | 1.80 | helix |
| | seq26 | 76-LKGGKCHH | 3.9 | 0.57 | 1.44 | helix |
| | seq27 | 79-KGGHREA | 3.5 | 0.54 | 1.66 | helix |
| | seq28 | 82-HEAEALKP | 4.7 | 0.04 | 0.91 | helix |
| | seq29 | 85-AELKLPLQ | 3.8 | 0.15 | 1.76 | helix |
| | seq30 | 88-XLATQSHA | 3.4 | 0.25 | 1.29 | helix |
| | seq31 | 91-AQSHATKH | 3.3 | 0.44 | 1.40 | helix |
| | seq32 | 94-HATKHHKIP | 2.5 | 0.30 | 1.34 | helix |
| | seq33 | 97-KHIFIPIYK | 3.5 | 0.61 | 1.39 | helix |
| | seq34 | 100-PIKYLEF | 3.4 | 0.90 | 0.95 | helix |
| | seq35 | 103-KYLEFISE | 4.2 | 0.53 | 1.25 | helix |
| | seq36 | 106-EIFSRAII | 4.6 | 0.13 | 0.99 | helix |
| | seq37 | 109-SEAISHVL | 0.6 | 7.12 | 0.90 | helix |
| | seq38 | 112-ILQDKSGP | 4.2 | 0.72 | 0.94 | helix |
| | seq39 | 115-VLHSHRP | 3.7 | 0.81 | 1.14 | turn |
| | seq40 | 118-SRHPSGNGF | 3.1 | 0.98 | 1.48 | turn |
| | seq41 | 121-PGNFGADA | 3.9 | 0.43 | 2.31 | helix |
| | seq42 | 124-FGADAKQG | 3.6 | 0.72 | 2.14 | helix |
| | seq43 | 127-DACGANKK | 4.3 | 0.80 | 1.66 | helix |
| | seq44 | 130-AMNIVKH | 3.1 | 0.26 | 1.50 | helix |
| | seq45 | 133-KAHELFR | 3.7 | 0.28 | 1.13 | helix |
| | seq46 | 136-LELFREDI | 0.4 | 0.88 | 0.73 | helix |
| | seq47 | 139-FKDIJAK | 0.4 | 1.28 | 0.95 | helix |
| | seq48 | 142-DIAKYEK | 3.9 | 1.68 | 0.68 | helix |
| | seq49 | 145-ARYKELIQG | 3.7 | 0.81 | 1.66 | helix |

RMSD is the most likely value of RMSD extracted from the free-energy profile of RMSD. The ground mesostring is sometimes nearly identical to less-populated mesostrings. If the most populated mesostrings differ by only one mesostate, we group them into a consensus mesostring, which contains one indefinite mesostate signified by [—].

P<sub>i</sub> is the probability of the ground mesostate.

RMSD is the most likely value of RMSD extracted from the free-energy profile of RMSD. The ground mesostring is sometimes nearly identical to less-populated mesostrings. If the most populated mesostrings differ by only one mesostate, we group them into a consensus mesostring, which contains one indefinite mesostate signified by [—].

P<sub>i</sub> is the probability of the ground mesostate.
make eight high-confidence predictions in four of the six proteins. In those instances, our simulations are largely consistent with theirs, indicating that the intrinsic physical preferences contribute to the PDB structures. However, the present simulations are also more informative, giving 48 structures (with 85% reliability) among the 133 peptides we tested, in contrast to the eight (having 74% reliability) found by I-sites.

Current structure-prediction systems rely on a pragmatic mix of bio-informatics and physical modeling [23,24]. A key component of these systems is the use of fragment libraries to

### Table 4. Ground Mesostrings of β-Sheet Proteins

| Protein               | Peptide     | Sequence    | RMSD in Å | Native Mesostring | Ground Mesostring | P₁ in Percent | ΔF in kcal/mol | TS in kcal/mol | Native Structure |
|-----------------------|-------------|-------------|-----------|-------------------|-------------------|--------------|----------------|----------------|------------------|
| Chymotrypsin inhibitor| seq1        | 1-NLXTWFE   | 5.2       | bbabbbabbbab     | bbaabb-          | 65           | 0.89           | 0.90           | loop             |
| seq2                  | 4-TPWELV    | 4.2         | abbaabbab   | bbaabb-ab        | 15               | 0.24         | 1.65           | 0.60           | 3₁0 helix        |
| seq4                  | 10-PELVK     | 2.9         | abbaabbab   | bbaabb-ab        | 34               | 0.51         | 1.39           | 0.60           | helix            |
| seq5                  | 13-SVEAAKVK | 0.5         | aaabbbabb   | bbaabb-ab        | 45               | 0.17         | 0.75           | 0.60           | helix-cap         |
| seq6                  | 16-EEKVIL    | 4.3         | abbaabbab   | bbaabb-ab        | 40               | 0.64         | 1.18           | 0.60           | strand           |
| seq7                  | 19-KVILQDX   | 3.9         | abbaabbab   | bbaabb-ab        | 37               | 0.07         | 0.83           | 0.60           | strand           |
| seq8                  | 22-LQKPEAQ  | 2.7         | abbaabbab   | bbaabb-ab        | 37               | 1.00         | 1.09           | 0.60           | hairpin-turn     |
| seq9                  | 25-KPEAQIUV  | 4.9         | bbabbbbab   | bbaabb-ab        | 41               | 0.31         | 1.25           | 0.60           | strand           |
| seq10                 | 28-AQITLVP   | 5.7         | bbabbbbab   | bbaabb-ab        | 46               | 0.68         | 0.91           | 0.60           | strand           |
| seq11                 | 31-TIVLPV    | 3.1         | bbabbbbab   | bbaabb-ab        | 20               | 0.14         | 1.54           | 0.60           | strand           |
| seq12                 | 34-PVGlTV    | 4.2         | bbabbbbab   | bbaabb-ab        | 12               | 0.06         | 1.87           | 0.60           | strand           |
| seq13                 | 37-TIVTM     | 4.0         | bbabbbbab   | bbaabb-ab        | 30               | 0.44         | 1.23           | 0.60           | strand           |
| seq14                 | 40-TMEVRL    | 3.7         | bbabbbbab   | bbaabb-ab        | 64               | 0.07         | 1.02           | 0.60           | strand           |
| seq15                 | 43-VRDIRV    | 3.2         | bbabbbbab   | bbaabb-ab        | 48               | 0.17         | 0.75           | 0.60           | strand           |
| seq16                 | 46-DSVLFLV   | 6.4         | abbbbabbb   | bbaabb-ab        | 40               | 0.64         | 1.18           | 0.60           | strand           |
| seq17                 | 49-RFLVXD    | 4.2         | bbabbbbab   | bbaabb-ab        | 37               | 0.07         | 0.83           | 0.60           | strand           |
| seq18                 | 52-VXDLN     | 4.1         | bbabbbbab   | bbaabb-ab        | 64               | 1.00         | 1.09           | 0.60           | helix            |
| seq19                 | 55-LDIAEVP   | 3.3         | bbabbbbab   | bbaabb-ab        | 22               | 0.15         | 1.36           | 0.60           | bulge            |
| seq20                 | 58-IAEVP     | 3.7         | bbabbbbab   | bbaabb-ab        | 66               | 0.97         | 0.99           | 0.60           | bulge            |
| α Spectrin            | seq1        | 1-KTVLAL    | 4.3         | bbabbbbab       | bbaabb-ab        | 44               | 0.53         | 0.90           | 0.60           | loop             |
| seq2                  | 4-VLALYD    | 3.7         | bbabbbbab   | bbaabb-ab        | 34               | 0.31         | 1.23           | 0.60           | loop             |
| seq3                  | 7-LVQYE     | 4.0         | abbbbabbb   | bbaabb-ab        | 55               | 0.78         | 0.88           | 0.60           | loop             |
| seq4                  | 10-QQKESP    | 3.6         | bbabbbbab   | bbaabb-ab        | 44               | 0.53         | 0.90           | 0.60           | loop             |
| seq5                  | 13-KSPREV    | 3.8         | abbbbabbb   | bbaababbb        | 58               | 1.01         | 1.06           | 0.60           | diverging-turn   |
| seq6                  | 16-REYTM     | 4.5         | abbbbabbb   | bbaababbb        | 48               | 0.86         | 1.18           | 0.60           | diverging-turn   |
| seq7                  | 19-TMKPDIL   | 2.7         | bbabbbbab   | bbaababbb        | 23               | 0.18         | 1.45           | 0.60           | bulge            |
| seq8                  | 22-KDILTL    | 4.4         | bbabbbbab   | bbaababbb        | 78               | 1.73         | 0.97           | 0.60           | bulge            |
| seq9                  | 25-LTLNLST   | 3.9         | bbabbbbab   | bbaababbb        | 75               | 1.42         | 0.93           | 0.60           | bulge            |
| seq10                 | 28-LGTSKX    | 4.0         | babaabaabb   | bbaababbb        | 53               | 1.06         | 1.25           | 0.60           | hairpin-turn     |
| seq11                 | 31-STKDFWK   | 3.2         | ababbbabb   | bbaababbb        | 36               | 0.22         | 1.15           | 0.60           | hairpin-turn     |
| seq12                 | 34-KWDDKEV   | 5.8         | abbbbabbb   | bbaababbb        | 43               | 0.70         | 1.32           | 0.60           | hairpin-turn     |
| seq13                 | 37-WKXVND    | 3.8         | bbabbbbab   | bbaababbb        | 46               | 0.60         | 1.01           | 0.60           | hairpin-turn     |
| seq14                 | 40-EYNQDQG   | 3.7         | abbbbabbb   | bbaababbb        | 25               | 0.07         | 1.33           | 0.60           | hairpin-turn     |
| seq15                 | 43-DGQGFVPA  | 5.6         | abbbbabbb   | bbaababbb        | 12               | 0.07         | 1.62           | 0.60           | hairpin-turn     |
| seq16                 | 46-GFVPAY    | 3.2         | bbabbbbab   | bbaababbb        | 36               | 0.78         | 1.42           | 0.60           | hairpin-turn     |
| seq17                 | 49-PAAVTK    | 3.3         | bbabbbbab   | bbaababbb        | 41               | 0.14         | 0.93           | 0.60           | hairpin-turn     |

RMSD is the most likely value of RMSD extracted from the free-energy profile of RMSD. The ground mesostring is sometimes nearly identical to less-populated mesostrings. If the most populated mesostrings differ by only one mesostate, we group them into a consensus mesostring, which contains one indefinite mesostate signified by ‘-‘.

P₁ is the probability of the ground mesostring.

ΔF is the free-energy difference between the ground mesostring and the next mesostring.

TS is the entropy of the mesostrings.

Native Structure is the description of the structure of the peptide in the native structure.

DOI: 10.1371/journal.pcbi.0020027.t004
identify folding initiation sites. Here we have identified the physical origin of the sequence-structure relations identified in the fragment libraries—local structural bias in short peptide sequences. The calculations are not exorbitant, as each peptide takes \( \sim 160 \) CPU node hours, and, in many cases, our results go beyond the fragment libraries. By replacing fragment libraries with peptide simulations to identify folding initiation sites, we move closer to the goal of predicting protein structures using only physical models.

**Materials and Methods**

Replica-exchange simulations of the peptides. Replica-exchange simulations were conducted using a PERL wrapper (http://www.dillgroup.ucsf.edu/\~jchodera/code/relx) around the SANDER molecular dynamics program for the Amber7 molecular-modeling package [49]. We used 16 replicas exponentially spaced between 270K and 360K, achieving an exchange-acceptance probability of approximately 50%. Exchanges were attempted every 1 ps, with constant-energy dynamics conducted between exchanges. After each exchange attempt, the velocities were redrawn from the appropriate Maxwell-Boltzmann distribution to ensure proper thermostating. A 2-fs time step was used, and bonds to hydrogens were constrained with SHAKE [50]. Configurations were stored every 1 ps for analysis. Simulations were run for 5 ns per replica and the first 4 ns were used for equilibration. The peptides were capped with ACE and NME blocking groups, and initialized in the extended state. Systems were set up using the LEAP program. Peptide parameters were taken from the Amber Parm96 force field, and the GB/SA model of Tsui and Case was used [51], along with a surface area penalty term of 5 cal mol\(^{-1}\) A\(^{-2}\).

**Calculating thermodynamic observables.** We use replica exchange [30] to simulate the equilibrium ensemble. It samples \( k \) parallel replicas, each of which is at a different temperature. Hence, to extract thermodynamic observables for a given temperature, say \( T = 300K \), we must reweigh the configurations taken from the \( k \) different temperatures \( T\) in order to combine them into a representative ensemble. We do this reweighing of the replicas with an implementation [51] of the Weighted Histogram Analysis Method [52].

We first calculate the dimensionless free-energy \( \Omega \) for each replica \( k \). Starting with a crude estimate of \( \delta E \) the weight of states with energy \( E \) in replica \( k \):

\[
\Omega_k = \frac{N_k}{N^k \exp(\delta E - \beta k)}
\]

\( N_k \) is the number of states that energy \( E \) in replica \( k \) and \( N \) is the number of states in all replicas. The temperature \( T \) is given by the equation:

\[
\beta = \frac{1}{kT}
\]

**Calculating free-energy differences.** To calculate free-energy differences, we can use the equilibrium ensemble:

\[
\delta G = \sum_{k=1}^{K} \frac{N_k}{N^k} \delta E_k
\]

**Table 5. Comparison of the Structural Bias with the Native Structure**

| Protein | Lines | Structure per Residue |
|---------|-------|-----------------------|
| Protein G | 1 | MTYKLLNGXKLGETTCTEAVDAATAEKVFEQYANDNGVDGRTYDATKTFTVTE |
| | 2 | 1 | QQAFAELHPLNEEQNGUPFGQSLKDPXPSNLAWAEKXKND |
| | 3 | 1 | KANLIFANGSSQAETKPGTEATSEAYADTLKPEKQTEL |
where $N^k_x$ is the number of snapshots in replica $k$ with energy $E$. From the distribution of $\Omega^k_x$, we calculate a new estimate of $f^k$ by

$$f^k = -\log \left[ \sum_x \Omega^k_x \exp(\beta E) \right]$$  \hspace{1cm} (2)

We iterate the above two steps until $f^k$ converges. Then we use these dimensionless free energies $f^k$ to reweigh the relative free-energy profile $F$ of observable $x$ to the target temperature $\beta_{tar}$:

$$F_x(\beta_{tar}) = -\frac{1}{\beta_{tar}} \log \left[ \sum_k \sum_x N^k_x \exp(\beta_{tar} E) \right]$$  \hspace{1cm} (3)

After using the Weighted Histogram Analysis Method to calculate the relative free energies $F_i$ of a mesostring $i$, we calculate the probabilities $P_i$ by

$$P_i = \frac{\exp(-\beta_{tar} F_i)}{\sum_j \exp(-\beta_{tar} F_j)}$$  \hspace{1cm} (4)

When we merge similar mesostrings into a consensus mesostring, we calculate the free-energy difference to another mesostring $j$ by

$$\Delta F = -\frac{1}{\beta_{tar}} \log \left( \frac{P_{\text{consensus}}}{P_j} \right)$$  \hspace{1cm} (5)

**Defining the backbone mesostates.** A key part of our analysis is the discretizing of the backbone degrees of freedom. This is based on the original analysis of the protein backbone [33]. In that analysis, Ramachandran and coworkers showed that the stereochemistry of the protein backbone breaks up the backbone $\phi$-$\psi$ angles into three distinct regions, each separated by significant energy barriers. We can thus describe the conformation of a peptide as a string of discrete mesostates—we call this the mesostring. A given mesostring is separated in energy from other mesostrings. Each mesostring corresponds to a low-energy basin in the conformation space of the peptide backbone. It is then straightforward to extract the local structure from the lowest free-energy basin. This partitioning in terms of discrete regions in the backbone angles has been observed in a molecular dynamics simulation of an $\alpha$-helical peptide [31].

The original analysis of the backbone identified three distinct regions in the $\phi$-$\psi$ angles [53]. Recent studies of the protein database found that these three regions can be further divided up into five clusters of density [54,55]. Some of the barriers between these five regions are small, which leaves three regions separated by large barriers. However we cannot use the database analysis to define the boundaries of the backbone mesostates because current force fields cannot replicate the database distribution of $\phi$-$\psi$ angles. We must define the boundaries the backbone mesostates in terms of the force field in our molecular dynamics: we ran replica-exchange simulations of the alanine dipeptide and the glycine dipeptide for 10 ns and calculated the free-energy profile of the $\phi$-$\psi$ angles in bins of 5°. Based on the resulting free-energy profile, we break up the Ramachandran plot in terms of the following mesostates:

- $[\beta] : (-180° < \phi < 0°, 45° < \psi < 180°)$
- $U(-180° < \phi < 0°, -180° < \psi < -135°)$
- $U(120° < \phi < 180°, 45° < \psi < 180°)$
- $U(120° < \phi < 180°, -180° < \psi < -135°)$
- $[a] : (-180° < \phi < 0°, -135° < \psi < 45°)$
- $U(120° < \phi < 180°, -135° < \psi < 45°)$
- $[l] : (0° < \phi < 120°, -180° < \psi < 180°)$
- $U(120° < \phi < 180°, -135° < \psi < 45°)$

And for glycines:

- $[\beta] : (-180° < \phi < 0°, 45° < \psi < 180°)$
- $U(-180° < \phi < 0°, -180° < \psi < -135°)$
- $U(0° < \phi < 180°, 135° < \psi < 180°)$
- $U(0° < \phi < 180°, -180° < \psi < -45°)$
- $[a] : (-180° < \phi < 0°, -135° < \psi < 45°)$
- $[l] : (0° < \phi < 120°, -45° < \psi < 135°)$

**Acknowledgments**

Thanks to John Chodera for the replica-exchange wrapper for the molecular dynamics package. Thanks to Banu Ozkan, Vince Voelz and Albert Wu for many invaluable discussions.

**Author contributions.** BKH and KAD conceived and designed the experiments. BKH performed the experiments. BKH analyzed the data. BKH wrote the paper.

**Funding.** We appreciate the support of NIH grant GM49993.

**Competing interests.** The authors have declared that no competing interests exist.

**References**

1. Dyson HJ, Wright PE (1998) Equilibrium NMR studies of unfolded and partially folded proteins. Nat Struct Biol 5 (Supplement): 499–505.
2. Dyson HJ, Merukut G, Waldo JP, Lerner RA, Wright PE (1992) Folding of peptide fragments comprising the complete sequence of proteins. Models for initiation of protein folding. I. Myohemerythrin. J Mol Biol 226: 795–817.
3. Shin HC, Merukut G, Waldo JP, Tennant LL, Dyson HJ, Wright PE (1995) Peptide models of protein folding initiation sites. 3. The G-H helical hairpin of myoglobin. Biochemistry 32: 6356–6364.
4. Waldo JP, Fehrs VA, Merukut G, Dyson HJ, Wright PE (1993) Peptide models of protein folding initiation sites. 2. Secondary structure formation by peptides corresponding to the G- and H-helices of myoglobin. Biochemistry 32: 6337–6347.
5. Ramirez-Alvarado M, Serrano L, Blanco FJ (1997) Conformational analysis of peptides corresponding to all the secondary structural elements of protein L BI domain: Secondary structure propensities are not conserved in proteins with the same fold. Protein Sci 6: 162–174.

6. Eliezer D, Chung J, Dyson HJ, Wright PE (2000) Native and non-native secondary structure and dynamics in the pH 4 intermediate of apomyoglobin. Biochemistry 39: 2949–2901.

7. Mohana-Borges R, Goto NK, Kroon GJ, Dyson HJ, Wright PE (2004) Structural characterization of unfolded states of apomyoglobin using residual dipolar couplings. J Mol Biol 340: 1131–1142.

8. Marqusee S, Robbins VH, Baldwin RL (1989) Unusually stable helix formation in short alanine-based peptides. Proc Natl Acad Sci U S A 86: 5296–5290.

9. Munoz V, Serrano L (1994) Elucidating the folding problem of helical peptides using empirical parameters. Nat Struct Biol 1: 399–409.

10. Blanco FJ, Rivas G, Serrano L (1994) A short linear peptide that folds into a native stable beta-hairpin in aqueous solution. Nat Struct Biol 1: 584–590.

11. Searle MS, Williams DH, Packman LC (1995) A short linear peptide derived from the N-terminal sequence of ubiquitin folds into a water-stable non-native beta-hairpin. Nat Struct Biol 2: 799–806.

12. Zerella R, Evans PA, Ionides JM, Packman LC, Trotter BW, Mackas JP, Williams DH (1999) Autonomous folding of a peptide corresponding to the N-terminal beta-hairpin from ubiquitin. Protein Sci 8: 1320–1331.

13. Espinosa JP, Munoz V, Gellman SH (2001) Interplay between hydrophobic cluster and loop propensity in beta-hairpin formation. J Mol Biol 306: 397–402.

14. Rotondi KS, Gierszach LM (2003) Role of local sequence in the folding of cellular retinoic acid binding protein 1: Structural propensities of reverse turns. Biochemistry 42: 7797–7795.

15. Eisenberg D, Weiss RM, Terwilliger TC (1984) The hydrophobic moment parameter. Current Opinion in Structure Biology 4: 436–456.

16. Kamtekar S, Schiffer JM, Xiong H, Babik JM, Hecht MH (1993) Protein turns: Detection of periodicity in protein hydrophobicity. Proc Natl Acad Sci U S A 90: 140–144.

17. Kanetkar S, Schiffer JM, Xiong H, Babik JM, Hecht MH (1993) Protein design by binary patterning of polar and nonpolar amino acids. Science 262: 1680–1685.

18. Han RK, Baker D (1996) Global properties of the mapping between local amino acid sequence and local structure in proteins. Proc Natl Acad Sci U S A 93: 5814–5818.

19. Bystroff C, Simon GT, Han RK, Baker D (1996) Local sequence–structure correlations in proteins.Curr Opin Biotechnol 7: 417–421.

20. Bystroff C, Baker D (1998) Prediction of local structure in proteins using a library of sequence–structure motifs. J Mol Biol 281: 565–577.

21. Tsai CJ, Maizel JV Jr, Nussinov R (2000) Anatomy of protein structural fragments reveals modular building block approach of protein structural fragments model native protein structures accurately. J Mol Biol 323: 297–307.

22. Tendulkar AV, Joshi AA, Sohoni MA, Wangkjar P (2004) Clustering of protein structural fragments reveals modular building block approach of nature. J Mol Biol 338: 611–629.

23. Aloy P, Stark A, Hadley C, Russell RB (2003) Predictions without templates: The role of a side chain blocking motif in achieving native specificities of designed beta-hairpins. J Mol Biol 338: 611–629.

24. Moellt J, Moul J (1995) A decade of CASP: Progress, bottlenecks and prognosis in protein structure prediction. Curr Opin Struct Biol 15: 285–289.

25. Avbelj F, Moult J (1995) Determination of the conformation of folding initiation sites in proteins by computer simulation. Proteins 23: 129–141.

26. Sujitvarasan R, Rose GD (1995) LNETS: A hierarchical procedure to predict the fold of a protein. Proteins 22: 81–99.

27. Gibbes N, Clarke AR, Sessions RB (2001) Ab initio protein structure prediction using physicochemical potentials and a simplified off-lattice model. Proteins 43: 186–202.

28. Klepeis JL, Floudas CA (2002) Ab initio prediction of helical segments in polyamides. J Comput Chem 23: 245–266.

29. Daura X, van Gunsteren WF, Mark AE (1999) Folding–unfolding thermodynamics of a beta-haepptide from equilibrium simulations. Proteins 34: 269–280.

30. de Groot BL, Daura X, Mark AE, Grubmüller H (2001) Essential dynamics of reversible peptide folding: Memory-free conformational dynamics governed by internal hydrogen bonds. J Mol Biol 309: 299–313.

31. Yu M, Nguyen PH, Stock G (2005) Energy landscape of a small peptide revealed by dihedral angle principal component analysis. Proteins 58: 45–52.

32. Bystroff C, Garde S (2003) Helix propensities of short peptides: Molecular dynamics versus bioinformatics. Proteins 50: 552–562.

33. Kim PS, Baldwin RL (1992) Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. Annu Rev Biochem 51: 459–489.

34. Dill KA, Fiebig KM, Chan HS (1995) Cooperativity in protein-folding kinetics. Proc Natl Acad Sci U S A 92: 1942–1946.

35. Baldwin RL, Rose GD (1999) Is protein folding hierarchical? I. Local structure and peptide folding. Trends Biochem Sci 24: 26–33.

36. Sugita Y, Okamoto Y (1999) Replica-exchange molecular dynamics method for protein folding. Chemical Physics Letters 314: 141–151.

37. Tsui V, Case DA (2000) Theory and applications of the generalized Born solvation model in macromolecular simulations. Biopolymers 56: 275–291.

38. Zhou R (2003) Free energy landscape of protein folding in water: Explicit vs. implicit solvent. Proteins 53: 148–161.

39. Munoz DL, Jr, Kim PS (1996) Context-dependent secondary structure formation of a designed protein sequence. Nature 380: 730–734.

40. Bai Y, Englander SW (1994). Bai Y, Englander SW (1994) Hydrogen bond strength and beta-sheet propensities: The role of a side chain blocking effect. Proteins 19: 282–296.

41. Jennings PA, Wright PE (1993) Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. Science 262: 892–898.

42. Izhaki LS, Neira JL, Ruiz-Sanz J, de Prat Gay G, Fersht AR (1995) Search for nucleation sites in smaller fragments of chymotrypsin inhibitor 2. J Mol Biol 254: 289–304.

43. Viguera AR, Jimenez MA, Rico M, Serrano L (1996) Conformational analysis of peptides corresponding to beta-hairpins and a beta-sheet that represent the entire sequence of the alpha-spectrin SH3 domain. J Mol Biol 255: 507–521.

44. Sundaralingam M, Sekaranudiy UC (1989) Water-inserted alpha-helical segments implicate reverse turns as folding intermediates. Science 244: 1333–1337.

45. Soman KV, Karimi A, Case DA (1991) Unfolding of an alpha-helix in water. Biopolymers 31: 1531–1536.

46. Du D, Zhu Y, Huang CY, Gai F (2004) Understanding the key factors that control the rate of beta-hairpin folding. Proc Natl Acad Sci U S A 101: 15915–15920.

47. Khimov BK, Thirumalai D (2000) Mechanisms and kinetics of beta-hairpin formation. Proc Natl Acad Sci U S A 97: 2544–2549.

48. Munoz V, Thompson PA, Hofrichter J, Eaton WA (1997) Folding dynamics and mechanism of beta-hairpin formation. Nature 390: 196–199.

49. Pearlman DA, Case DA, Caldwell JW, Ross WS, Cheatham TE, et al. (1995) Amber, a package of computer programs for applying molecular mechanics, normal-mode analysis, molecular-dynamics and free-energy calculations to simulate the structural and energetic properties of molecules. Comput Physics Commun 91: 1–41.

50. Rooker JP, Ciccotti G, Berendsen HJC (1977) Numerical-integration of Cartesian equations of motion of a system with constraints—molecular-dynamics of N-alkanes. Journal of Computational Physics 23: 327–341.

51. Chodera JD, Swope WC, Pitera JW, Case D, Dill KA (2006) Use of the Weighted Histogram Analysis Method for the analysis of simulated and parallel tempering simulations. J Chem Theory Comput: In press.

52. Kumar S, Bouzida D, Swendsen RH, Kollman PA, Rosenberg JM (1992) The Weighted Histogram Analysis Method for free-energy calculations on biomolecules. 1. The method. J Comput Chem 13: 1011–1021.

53. Ramachandran GN, Ramakrishnan C, Sasisekharan V (1963) Stereochemistry of polypeptide chain configurations. J Mol Biol 7: 95–99.

54. Karplus PA (1996) Experimentally observed conformation-dependent geometry and hidden strain in proteins. Protein Sci 5: 1406–1429.

55. Ho BK, Thomas A, Brasseur R (2003) Revisiting the Ramachandran plot: Hard-sphere repulsion, electrostatics, and H-bonding in the alpha-helix. Protein Sci 12: 2508-2522.