

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/228086840>

The Coarse-Grained OPEP Force Field for Non-Amyloid and Amyloid Proteins

ARTICLE *in* THE JOURNAL OF PHYSICAL CHEMISTRY B · JUNE 2012

Impact Factor: 3.3 · DOI: 10.1021/jp301665f · Source: PubMed

CITATIONS

30

READS

26

3 AUTHORS, INCLUDING:



[Yassmine Chebaro](#)

Institut de Génétique et de Biologie Moléculair...

15 PUBLICATIONS 324 CITATIONS

[SEE PROFILE](#)



[Philippe Derreumaux](#)

French National Centre for Scientific Research...

167 PUBLICATIONS 3,811 CITATIONS

[SEE PROFILE](#)

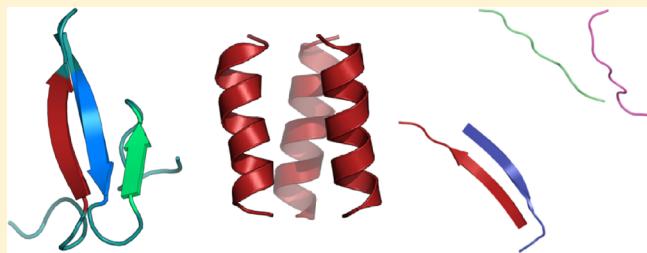
The Coarse-Grained OPEP Force Field for Non-Amyloid and Amyloid Proteins

Yassmine Chebaro,^{†,§} Samuela Pasquali,[†] and Philippe Derreumaux*,^{†,‡}

[†]Laboratoire de Biochimie Théorique, CNRS UPR 9080, Université Paris Diderot, Sorbonne Paris Cité, Institut de Biologie Physico-Chimique, 13 rue Pierre et Marie Curie, 75005 Paris

[‡]Institut Universitaire de France, 103 Bd Saint-Michel, Paris 75005, France

ABSTRACT: Coarse-grained protein models with various levels of granularity and degrees of freedom offer the possibility to explore many phenomena including folding, assembly, and recognition in terms of dynamics and thermodynamics that are inaccessible to all-atom representations in explicit aqueous solution. Here, we present a refined version of the coarse-grained optimized potential for efficient protein structure prediction (OPEP) based on a six-bead representation. The OPEP version 4.0 parameter set, which uses a new analytical formulation for the nonbonded interactions and adds specific side-chain–side-chain interactions for α -helix, is subjected to three tests. First, we show that molecular dynamics simulations at 300 K preserve the experimental rigid conformations of 17 proteins with 37–152 amino acids within a root-mean-square deviation (RMSD) of 3.1 Å after 30 ns. Extending the simulation time to 100 ns for five proteins does not change the RMSDs. Second, replica exchange molecular dynamics (REMD) simulations recover the NMR structures of three prototypical β -hairpin and α -helix peptides and the NMR three-stranded β -sheet topology of a 37-residue WW domain, starting from randomly chosen states. Third, REMD simulations on the $\text{cc}\beta$ peptide show a temperature transition from a three-stranded coiled coil to amyloid-like aggregates consistent with experiments, while simulations on low molecular weight aggregates of the prion protein helix 1 do not. Overall, these studies indicate the effectiveness of our OPEP4 coarse-grained model for protein folding and aggregation, and report two future directions for improvement.



INTRODUCTION

Folding proteins/RNA and biomolecular complexes to their native structures is very challenging using all-atom molecular mechanics force fields.¹ The surrounding water molecules and ions must be treated, the quality of the underlying force field must be high, and the sampling of the conformational space must be very efficient. Recently, Shaw et al. realized a tour de force by elucidating the atomic detail of how three proteins with 37–56 amino acids² and 12 structurally diverse proteins with 10–80 amino acids³ fold into their three-dimensional native structures using ultralong molecular dynamics (MD) simulations over periods ranging between 100 μs and 1 ms.² Such a landmark work made possible by using one special purpose, a very fast supercomputer (Anton) and the modified AMBER² or CHARMM³ force fields, represents one big step ahead in protein folding.⁴

However, the shortest time for protein and RNA folding is about tens of microseconds, i.e., 2 orders of magnitude larger than standard all-atom MD simulations in explicit solvent, and the identification of statistically folding pathways requires many unfolding/folding events or trajectories and sophisticated analysis methods as well.⁵ Even with the use of enhanced conformational search techniques,⁶ the atomistic simulation of a single protein converges very slowly toward equilibrium and repeating it at a genomic scale is currently out of reach. Finally,

because many phenomena of high biological/medical interest such as the formation of amyloid plaques and vesicles, and the recognition of binary or multicomponent proteins in a crowded environment, include too many degrees of freedom, scientists are developing and improving coarse-grained (CG) models for various systems, such as lipids and surfactants,⁷ membrane proteins,⁸ DNA⁹ and RNA molecules,^{10,11} and soluble proteins.

Despite the loss of some atomistic details in the side-chains and the main-chain of amino acids or in the backbone/sugar/base of nucleic acids, CG models have proven to be quite useful in the description of many long time processes bridging the gap between the atomistic and mesoscopic scales. Not only are these approaches needed to speed up the simulation times, but they are also crucial to determine the essential forces and relevant degrees of freedom that shape the structures, thermodynamics, and kinetics of biomolecular systems.

In the context of amino acids, various levels of granularity have been developed since the pioneering work of Levitt and Warshel¹² ranging from one,^{13,14} two,^{15–18} three,^{19–22} four,^{23–25} five,²⁶ or six beads.^{27,28} More complex mappings also exist. For instance, Klein et al. employ a representation in

Received: February 20, 2012

Revised: June 6, 2012

Published: June 28, 2012

which roughly three to four heavy atoms (non-hydrogen) and adjacent hydrogens are mapped to a single CG. Most side-chains are represented by one CG, except Lys and Arg including a hydrophobic and a hydrophilic site and Tyr, Phe, and Trp which use two, two, and three particles, respectively.²⁹ Conversely, the MARTINI force field employs a four-to-one mapping; that is, on average, four heavy atoms are represented by a single interaction center, with an exception for ring-like molecules.³⁰ Finally, the PRIMO model combines one to several heavy atoms into CG sites that are chosen to allow an analytical, high-resolution reconstruction of all-atom models based on molecular bonding geometry constraints.³¹

In addition to the level of coarsening and the positions of the selected beads, there are several approaches to derive the potential and essentially the nonbonded interactions and test its robustness. The first one, thermodynamic-based, consists of fitting and predicting free energies such water/oil partitioning coefficients of the amino acid side-chain analogues (MARTINI),³⁰ or surface tension and density (Klein's force field).²⁹ The second approach, more structure-based in character, allows one to recover physical potentials from a model protein database and PDB distributions^{16,32} or the variance–covariance fluctuation matrix calculated from all-atom MD simulations.^{17,33} The resulting potential is then validated by testing the stability of native structures using MD simulations.^{16,19} A variant is to use minimized structures at $T = 0$ K by attempting to discriminate the native structures from decoy structures.^{23,28} Using models at room temperature is certainly more correct but requires running extensive MD simulations. Other methods for force field derivation include the force matching method³⁴ and the factor-expansion method.³⁵ The final class of force fields attempts to predict the structures of lowest free energy and thermodynamics properties such as the melting temperatures consistent with experiment.^{23,28,36,37}

The optimized potential for efficient protein structure prediction, OPEP, belongs to the latter category, though its first set of parameters used native and decoys at 0 K.^{38,39} OPEP, which is basically an all-atom backbone with CG side-chains, allows a good trade-off between chemical details, structural resolution, energetic accuracy, and CPU speed. Coupled to Monte Carlo,^{38,40–42} metadynamics,⁴³ the activation–relaxation technique (ART),^{44,45} and a greedy approach,^{46–48} OPEP has been used for protein folding and structure prediction. Combined with ART,^{49–53} MD,^{54–58} and REMD,^{59–62} OPEP simulations have also provided interesting insights into amyloid peptide oligomers.

Clearly, coarse graining cannot offer the structural and energetic precision of all-atom models in explicit solvent,² nor the vibrational accuracy of spectroscopic force fields.^{63,64} Despite these limitations, the OPEP3.2 parameter set was found rather accurate in a number of test cases. For instance, OPEP recognized satisfactorily the native structures from decoy structures using the energy (but not the free energy) criterion.³⁹ REMD-OPEP provided a good fit between predicted and experimental structures of six peptides (<28 amino acids),⁵⁹ and MD-OPEP kept the 50-residue three-helix bundle domain of staphylococcal protein A and the 56-residue B1 domain of protein G near their NMR structures at 300 K.⁵⁴ The performances of OPEP3.2 remained, however, to be evaluated on a larger set of proteins, which is the goal of the present study.

Using a training set of four proteins, we found it necessary to modify the analytical form and parameters for the side-chain–

side-chain interactions. The new parameter set, OPEP4, was then subjected to three tests. We first examined the stability of 17 proteins with 37–152 amino acids by using 30 ns MD simulations at 300 K, and checked that the RMSDs of five proteins do not increase using 100 ns time scales. We then explored the free energy surfaces of four monomeric peptides with 12, 16, and 37 amino acids using REMD simulations starting from random states. Third, to determine whether OPEP4 can discriminate between amyloidogenic and non-amyloidogenic peptides, we examined the $\text{cc}\beta$ peptide showing experimentally the hallmark of coiled coil and amyloid at low and high temperatures, respectively, and the peptide spanning helix H1 of the prion protein for which experiments have failed thus far to demonstrate amyloid formation.

MATERIALS AND METHODS

Protein Data Set for MD and REMD. Our set for MD consists of 17 proteins: 6 α , 7 $\alpha\beta$, and 4 β proteins with 37–152 amino acids. The six α proteins include the bacterial ribosomal protein L20 (PDB entry 1GYZ),⁶⁵ the first UBA domain in the human ubiquitin (2DA1), the DNA-binding domain of the 434 repressor (1PRA),⁶⁶ the C-terminal domain of nucleophosmin (2VXD),⁶⁷ the protein B DNA binding domain RP1 (1BW6),⁶⁸ and the N-terminal J domain of murine polyomavirus T antigens (1FAF).⁶⁹ The $\alpha\beta$ proteins include the N-terminal domain of *S. cerevisiae* RNase HI (1QHK),⁷⁰ the LysM domain from *E. coli* membrane-bound lytic murein transglycosylase D(MltD) (1E0G),⁷¹ the core variant of the B1 domain of Streptococcal protein G (1FCL),⁷² the apo calbindin D9K (1CLB),⁷³ the B1 immunoglobulin-binding domain of Streptococcal protein G (1PGB),⁷⁴ the *Bacillus subtilis* yndB START domain of 152 amino acids (2KTE),⁷⁵ and the *Escherichia coli* ribosomal protein L25 of 94 residues (1B75).⁷⁶ The four β proteins include the FBP28WW domain from *M. musculus* (1EOL),⁷⁷ the first Src homology 3 domain of Nck2 (2B86),⁷⁸ the antifungal protein from *A. giganteus* (1AFP),⁷⁹ and the Src-homology 3 domain (1SHG).⁸⁰

Our set for REMD simulation consists of six systems. The first four monomeric proteins include the 12-residue tryptophan zipper 1 (trpzip1) of sequence SWTWEGNKWTWK (PDB entry 1LE0),⁸¹ trpzip2 differing from trpzip1 by a transposition of an asparagine and a glycine in the hairpin turn (1LE1),⁸¹ the 16-residue p16-31 peptide of sequence NYHLENEVARLKKLVGE,⁸² and the 37-residue WW domain of sequence SMGLPPGWDEYKTHNGK-TYYYNHNTKTSTWTDPRMSS (1E0M).⁷⁷ The fifth and sixth systems include the 17-residue $\text{cc}\beta$ peptide SIRELEAR-IRELELRIG (1S9Z)⁸³ in its trimeric form and the peptide spanning helix H1 of the murine prion protein of sequence DWEDRYYREN in its monomeric, dimeric, and tetrameric forms.

Overall, the PDB entries 1PGB, 1SHG, and 1S9Z have been obtained by X-ray diffraction, there is no experimental structure for the H1 dimer and tetramer, and the atomic structures of all other systems have been determined by NMR spectroscopy in aqueous solution free of any cations and ligands.

Simulations and Analysis. All systems are studied in their zwitterion forms, except the amyloid peptides which are blocked by acetyl and NH₂ groups and the trpzip1 and trpzip2 peptides which are in their zwitterion forms at the N-terminus and blocked by NH₂ at the C-terminus, as done by experiments.

OPEP-MD simulations⁵⁴ are started from the experimental conformations and performed with a time step of 1.5 fs using the RATTLE algorithm.⁸⁴ They are carried out in spheres with reflecting boundary conditions. If the simulations are repeated, they use a different seed for the initial velocity distribution. The temperature is fixed by the Berendsen's thermostat⁸⁵ with a weak coupling constant of 0.1 ps. Note that OPEP-REMD simulations show little variations in the equilibrium structures and heat capacity curves of two model peptides (monomer and trimer) using both the Berendsen and Langevin thermostats.⁸⁶

OPEP-REMD simulations are carried out with a number of replicas and a temperature range dependent on the system size.⁵⁹ In all cases, a logarithmic-like temperature distribution is used and exchanges between two consecutive replicas are attempted every 7.5 ps, leading to an acceptance ratio of 30–40%.

Different properties are monitored for analysis. We use PTWHAM, the weighted-histogram method for parallel tempering,⁸⁷ to compute the heat capacity as a function of temperature. Secondary structures are assigned using the program STRIDE.⁸⁸ Error bars for the secondary structure propensities show the interval of confidence on the mean value given by the bootstrap statistical analysis method. Cluster analysis is done using the g-cluster tool in the GROMACS 3.3.1⁸⁹ package with a C_α RMSD cutoff varying between 1.5 and 4.0 Å according to the system size. C_α RMSD calculations are performed by using the McLachlan algorithm⁹⁰ in the program PROFIT (www.bioinf.org.uk/software/profit). Two intramolecular side-chains i, j are considered in contact if their distance is less than the sum of their van der Waals radii + 1 Å. Hydrogen bonds are assigned using the definition of Kabsch and Sander based on the calculation of the energy.⁹¹ Finally, we use two critical data for comparing the OPEP and experimental structures. The first RMSD uses the full structures and amino acids. The second RMSD uses the rigid cores by excluding the amino acids identified as flexible experimentally. Since NMR PDB entries can contain several models, the NMR reference state used here is always the model 1.

From OPEP3 to OPEP4 CG Models. OPEP uses a reduced representation of the amino acids consisting of all main-chain atoms (N, C_α , C, O, and H) and one bead for each side-chain (Sc), except the proline amino acid which is represented by all heavy atoms.^{39,92} Each Sc bead is described by appropriate van der Waals radius and geometry with respect to the backbone atoms N, C_α and C. The OPEP force field, which includes aqueous solution effects implicitly, has evolved over the years, and version 3 is expressed as a sum of bonded or local terms (bond lengths, bond angles, improper torsions of the side-chains and the amide bonds, backbone torsions) and nonbonded terms including van der Waals and hydrogen-bonding (H-bond) interactions. The analytical expression of all energy terms can be found in ref 39. Overall, OPEP3 is expressed as a function of 261 weighted energy terms: 213 for Enonbonded with 210 for Sc–Sc interactions, 4 for H-bond interactions; 4 local terms and 40 for the propensities of the residues to be in α or β states. Note the latter 40 terms are set to zero in version 3.2 and version 4.³⁹

In OPEP3, the hydrogen-bonding potential consists of two-body (EHB1) and four-body (EHB2) terms. The EHB1 term is summed over all residues i and j separated by $j = i + 4$ and $j > i + 4$, and is defined by two coupling constants (see eqs 7–9 in ref 39). The EHB2 term has also two energy components, one for α -helices and the other for parallel or antiparallel β -sheets

(see eq 10 in ref 39). In OPEP3, the van der Waals energy function between two side-chains i and j is defined by

$$E_{\text{vdW}} = E_{\text{AR3}} - \epsilon_{i,j} \left(\frac{r_{ij}^0}{r_{ij}} \right)^6 H(-\epsilon_{i,j}) \quad (1)$$

with E_{AR3} , the potential for the attractive–repulsive interactions described by

$$E_{\text{AR3}} = \epsilon_{i,j} \left(\left(\frac{r_{ij}^0}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{ij}^0}{r_{ij}} \right)^6 \right) H(\epsilon_{i,j}) \quad (2)$$

where $H(x)$ is the Heavyside side function, and equals 1 if $x \geq 0$ and 0 if $x \leq 0$, r_{ij} is the distance between particles i and j , $r_{ij}^0 = (r_i^0 + r_j^0)/2$ with r_i^0 the van der Waals radius of particle i , and $\epsilon_{i,j}$ is the coupling contant. $H(x)$ is set to 1 and the 12–6 potential is used when the interaction is hydrophobic or results from oppositely charged residues; otherwise, it is set to 0 and the –6 term is used. This form of the potential is also used in the UNRES force field.⁹³

As a first step toward enlarging our perspective on proteins, we used MD-OPEP 3.2 on four training proteins (1FAF: 79 amino acids, all α ; 2DA1: 79 amino acids, all α ; 1CLB: 75 amino acids, $\alpha\beta$ and 1BW6: 56 amino acids, all α) and found that, after 30 ns at 300 K, 1FAF and 2DA1 remain within 3.5 and 3.0 Å C_α RMSD from their NMR rigid cores, but the proteins 1CLB and 1BW6 deviate by 5.3 and 5.1 Å C_α RMSD. Looking at the final 1CLB and 1BW6 structures using OPEP3, we found a partial destabilization of α -helices.

Our strategy was thus to change the form of the nonbonded interactions using the function we developed for a coarse-grained RNA,¹¹ with the r^{-6} term replaced by r^{-8} , and a steeper form quickly vanishing with the distance. Following ref 11, the OPEP4 vdW expression is described by

$$E_{\text{vdW}} = E_{\text{AR4}} - \epsilon_{i,j} \left(\frac{r_{ij}^0}{r_{ij}} \right)^8 H(-\epsilon_{i,j}) \quad (3)$$

All pairwise side-chain interactions considered as attractive–repulsive are now described by

$$E_{\text{AR4}} = \epsilon_{i,j} \left(\left(\frac{G(r_{ij}^0)}{r_{ij}} \right)^6 e^{-2r_{ij}} + 0.6563701 \tanh[2(r_{ij} - r_{ij}^0 - 0.5) - 1] \right) H(\epsilon_{i,j}) \quad (4)$$

where $G(r_{ij}^0)$ is determined by imposing the value of the potential at a specific location r_{ij}^0 .¹¹ The behavior when r_{ij} is greater than r_{ij}^0 is controlled by the second part of eq 4, where the parameters are chosen to ensure a steeper profile in comparison with the standard Lennard-Jones potential, as shown in Figure 3 of ref 11.

Overall, by limiting the energy values between the side-chains at longer distances, eq 3 improves the MD simulations of 1BW6 and 1CLB, but this is not sufficient, however. We then decided to change the $\epsilon_{i,j}$ coupling constants of some interaction pairs present in α -helices. Indeed, using a statistical analysis on 150 PDB structures, we identified several $(i, i + 3)$ and $(i, i + 4)$ Sc–Sc contacts frequently observed in α -helices

Table 1. Summary of the 17 Systems Used for MD Simulations^a

PDB	<i>L</i>	α %	β %	C_α RMSD	rigid core
1GYZ ^b	60	65		3.5 ± 0.2	60–86, 92–115
2DA1	70	57		3.2 ± 0.4	16–63
1FAF	79	60.8		3.3 ± 0.1	5–41, 50–69
1PRA	69	63.8		3.4 ± 0.4	1–63
2VXD	54	76		3.4 ± 0.4	1–54
1BW6 ^c	56	62.5		3.4 ± 0.3	10–48
				3.6 ± 0.3	10–54
1QHK ^d	47	31.9	27.7	3.6 ± 0.2	6–13, 18–52
1E0G ^e	48	31.2	16.7	3.6 ± 0.3	2–47
1CLB	75	91.7	8	2.9 ± 0.2	3–14, 20–39, 47–53, 60–73
1FCL	56	26.8	28.6	2.8 ± 0.3	3–55
1PGB	56	26.8	42.8	3.3 ± 0.2	1–56
1B75	94	15	38.3	2.9 ± 0.2	1–8, 26–31, 38–78, 88–92
2KTE	152	26.6	24	3.6 ± 0.4	1–36, 41–52, 62–105, 118–152
1E0L	37		40.5	2.1 ± 0.5	7–33
2B86	59		40.7	2.8 ± 0.3	3–14, 20–51, 57–58
1AFP	51		25.5	2.9 ± 0.2	1–16, 24–27, 37–51
1SHG	57		47.4	2.4 ± 0.3	8–11, 29–61

^aFor each system, we give the number of amino acids (*L*), the percentage of α -helix and β -strand, the definition of the rigid core, and the mean C_α RMSD using 20–30 ns based on a single MD run for all systems, except the six α proteins (1GYZ–1BW6), where two runs are used. ^bNumbering in PDB starts at position 59 and ends at 118. The N-terminal half 1–58 is unstructured experimentally. ^c10–48 excludes helix H4 (residues 49–54), which is more disordered than the other three helices based on the backbone 1H–15N NOEs. ^dNo long range structural information on residues 1–5 and 53–63 was observed by NMR. PDB amino acid numbering is from 6 to 52. ^eResidues 389–452 were expressed, but non-random-coil chemical shifts and long-range NOEs were only observed between residues 389 and 445 and numbering in PDB covers this region.

that were not distinguished in OPEP3, which used the same coupling constant for each side-chain pair independently of the *i, j* separation. We have identified 11 interactions which have now a different coupling term if they are of (*i, i + 3*) and (*i, i + 4*) types or (*i, i + 2*) and (*i, i > 4*) types. The new (*i, i + 3*) and (*i, i + 4*) interactions include Lys-Glu, Lys-Asp, Glu-Arg, and Asp-Arg. For (*i, i + 4*), we also identified the following Lys-Gln, Lys-Leu, Ala-Arg, Ala-Gln, Ala-Glu, Leu-Glu, and Ile-Lys interactions. All these 11 interactions are no longer considered repulsive as in OPEP3 but attractive–repulsive.

Since each protein native state is associated with the lowest free energy and not the lowest internal energy, and thus recognizing native structures from non-native states based on the effective potential energy is not sufficient, we use as a first approximation that all new 11 (*i, i + 3*) and (*i, i + 4*) ϵ_{ij} 's are identical and only the four ϵ_{ij} 's associated with the two-body and four-body H-bond terms are varied, with the other 257 parameters remaining fixed to their OPEP v3.2 values. Next, using a genetic algorithm,³⁹ we checked that the newly v4 parameter set recognizes the native minimized structure from a minimized decoy data set for 1PGB, 2CI2, 1SHG, and 1ABZ.³⁹ Finally, we verified that OPEP4 improves the MD simulations on our four training proteins—we found that 1FAF, 2DA1, 1CLB, and 1BW6 remain within ~3.0 Å C_α RMSD (vs 3.5, 3.0, 5.3, and 5.1 Å using MD-OPEP3)—and works on our ensemble of 21 test proteins using either MD or REMD simulations as described now.

RESULTS AND DISCUSSION

MD-OPEP4 Simulations. Table 1 reports for each protein its amino acid length and secondary structure composition, the mean RMSD of the conformations generated within 20–30 ns with respect to the experimental rigid core at 300 K. Note that two runs are performed for the six α proteins, and one run is used for the β -sheet and mixed $\alpha\beta$ proteins. Figure 1 shows the

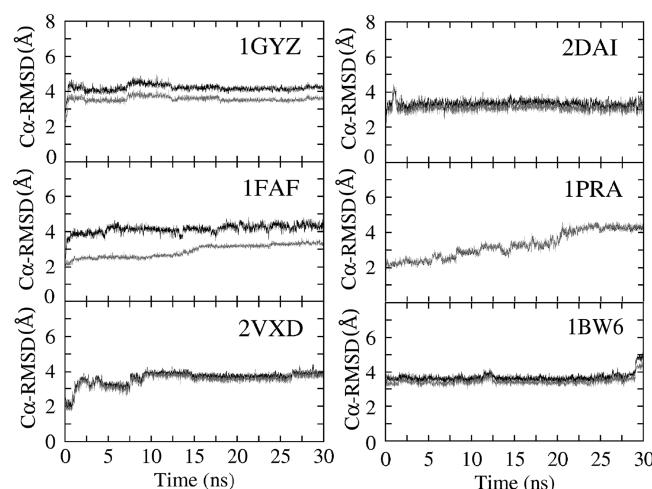


Figure 1. All α proteins: C_α RMSD as a function of simulation time. Two runs are used. The black (gray) lines represent the RMSD calculated using the full sequences (rigid cores).

C_α RMSD of the six α proteins as a function of time with respect to both their experimental rigid cores and full structures, and Figure 2 superposes the structures generated at 30 ns on the experimental conformations. Figures 3 and 4 show the same data for the (1QHK, 1E0G, 1FCL, and 1CLB) and (2B86, 1AFP, and 1B75) proteins, respectively. Because some of the results in Figure 1 (especially for 2VXD, 1PRA, and 1BW6) may indicate that the MD sampling has not reached convergence within 30 ns, these three α systems and the $\alpha\beta$ (1FAF) and β (1E0G) proteins were simulated on longer time scales. Figure 5 shows the RMSDs do not increase using 100 ns trajectories.

Among 17 targets, 7 targets display a C_α RMSD between 2.1 and 2.9 Å from their experimental rigid cores. These include

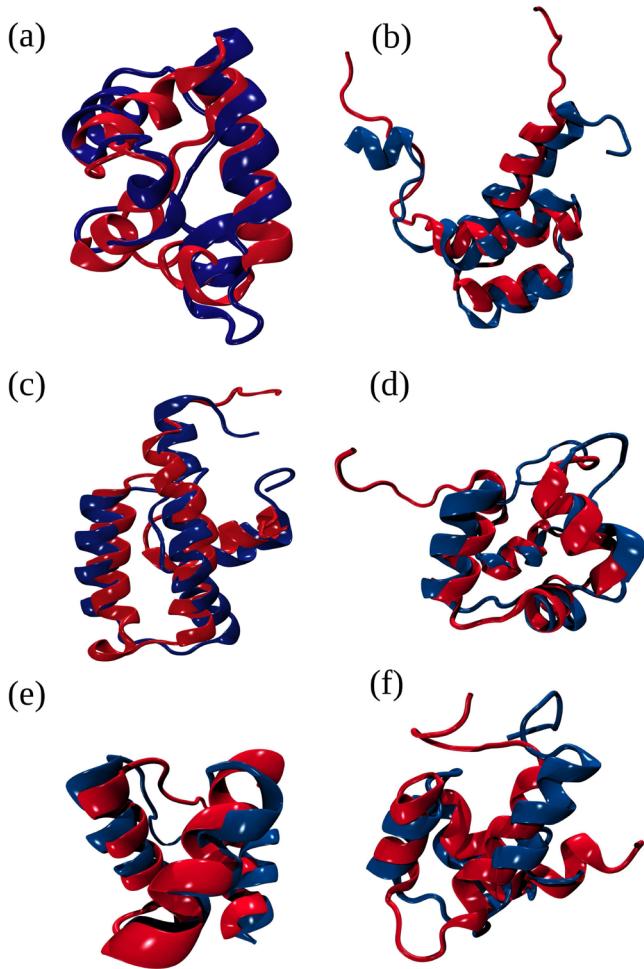


Figure 2. Conformations of all α proteins. Superposition of the conformation generated at 30 ns (in blue) on the NMR starting structure (in red). The systems are (a) 1GYZ, (b) 2DA1, (c) 1FAF, (d) 1PRA, (e) 2VXD, and (f) 1BW6.

the proteins 1FAF, 1CLB, 1FCL, 1E0L, 2B86, 1B75, and 1SHG with amino acid lengths ranging from 37 to 99. These targets with α , β , or mixed $\alpha\beta$ character conserve well their native secondary structures within 30 ns (see, for instance, panel 3G for 1CLB, panel 4B for 2B86, and panel 4C for 1AFP). The remaining 10 proteins with a C_α RMSD between 3.2 and 3.6 Å can be separated into two groups.

First, we observe that the enhanced C_α RMSD of the 2VXD, 1FAF, 2DA1, 1PRA, 1BW6, 1PGB, and 2KTE proteins results from fluctuations in α -helices and loops. For instance, the three-helix bundle 2VXD displays a rigid core RMSD of 3.4 Å due either to the unwinding of the C-terminal helix H3 by four residues in run 2 or the unfolding of the C-terminal ends of helices H1 and H3 by three residues in run 1 (panel 2E). The three-helix bundle 2DA1 displays a very stable helix H2 and some fluctuations in helices H1 and H3, leading to a rigid core RMSD of 3.2 Å (panel 2B). The mean C_α RMSD using the full sequence and rigid core of 1PRA is 3.6 and 3.4 Å. There are five helices spanning 2–13, 17–24, 28–36, 45–52, and 56–60. In run 1, the first four helices are well conserved and the short helix H5 is lost (panel 2D). In run 2, the helices H2–H5 are well conserved, while helix H1 only spans residues 2–7. The NMR structure of 1BW6 is characterized by four helices spanning 10–23, 28–35, 40–48, and 49–54, but helix H4 is

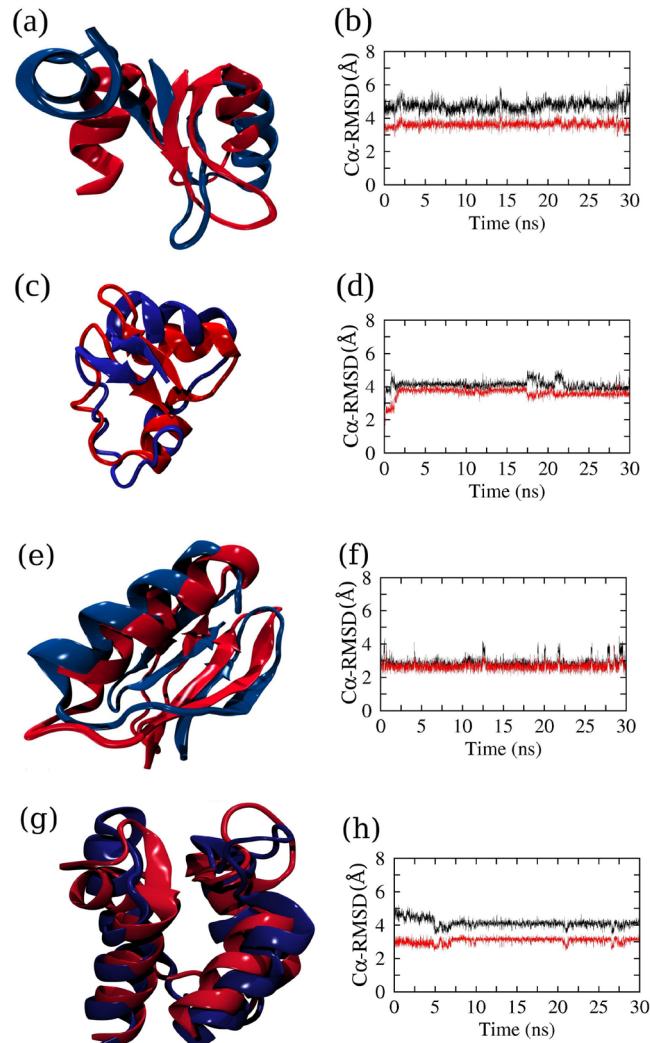


Figure 3. Conformations of α/β proteins. Superposition of the MD conformation at 30 ns (in blue) on the starting experimental structure (in red) and evolution of the C_α RMSD as a function of time. The systems are (a, b) 1QHK, (c, d) 1E0G, (e, f) 1FCL, and (g, h) 1CLB. The black (red) lines show the RMSD calculated on the full sequences (rigid cores).

more flexible than the others.⁶⁸ The C_α RMSD of 1BW6 is 3.4 and 3.6 Å using the residues 10–48 and 10–54, respectively. In both runs, helix H4 is lost, and helix H1 fluctuates between the amino acids 10–20 (run 1) and 10–18 (run 2) (panel 2F). The X-ray structure of 1PGB consists of a helix spanning residues 23–36 packed on a four-stranded β -sheet formed by two symmetrically opposed β -hairpins. The C_α RMSD using the full sequence is 3.3 Å due to extension of the helix at the C- and N-termini by two residues and the motion of the loops between the helix and the β -hairpins. Finally, the protein 2KTE of 152 amino acids consists of a β -sheet formed by five β -strands and three helices spanning 22–28, 32–36, and 120–143. The rigid core C_α RMSD is 3.6 Å. The sheet and helix H1 are well conserved, but the short helix H2 is lost and helix H3 is extended at its C-terminus spanning 124–151.

Second, we observe that the enhanced RMSD for the 1GYZ, 1QHK, and 1E0G proteins can also result from the difference in the amino acid sequences used by simulations and experiments. 1E0G is characterized by a $\beta\alpha\alpha\beta$ motif spanning residues 4–7, 13–19, 24–30, and 41–44 in the NMR model.

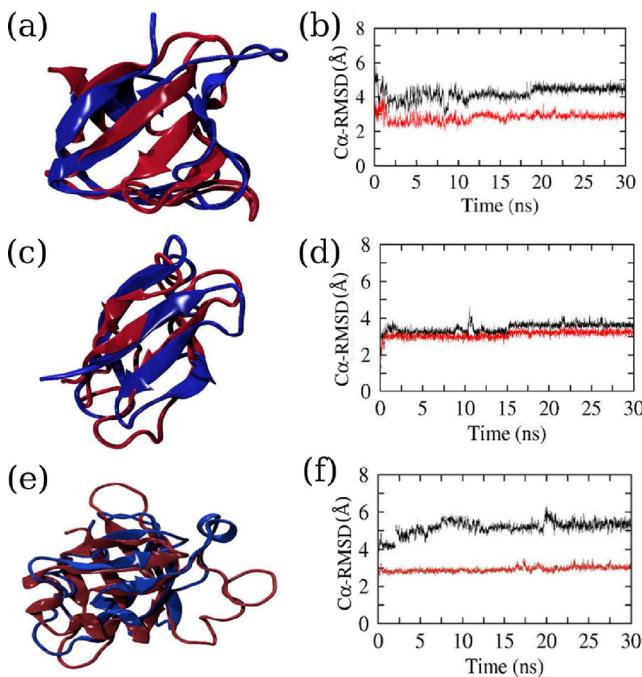


Figure 4. Conformations of 2β and $1\alpha\beta$ proteins. Superposition of the MD conformation at 30 ns (in blue) on the starting NMR structure (in red) and time evolution of the C_α RMSD. The systems are (a, b) 2B86, (c, d) 1AFP, and (e, f) 1B75. The black (and red) lines represent the RMSD calculated on the full sequences (and rigid cores).

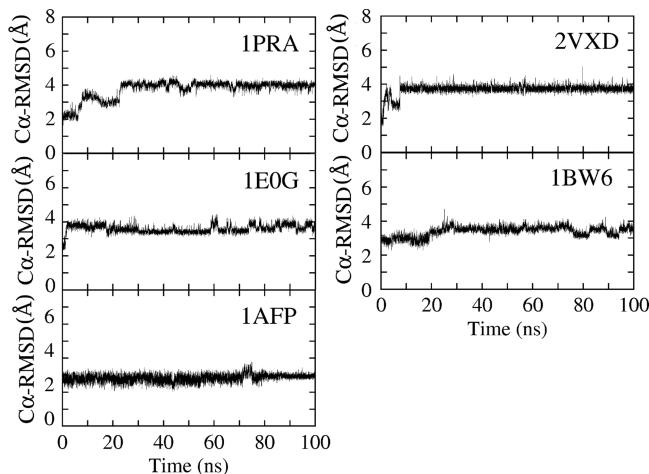


Figure 5. 100 ns MD for 1PRA, 1E0G, 1AFP, 2VXD, and 1BW6. The RMSD is calculated on the rigid core.

The C_α RMSD is 3.9 Å using the full sequence and 3.6 Å using the rigid core (panel 3D). While the two β -strands remain fully formed (panel 3C), helix H1 is extended by four residues (spanning 9–19) and helix H2 is shortened by three residues (spanning 24–27). NMR experiment was however performed on a sequence including eight additional C-terminal amino acids as reported in Table 1. 1QHK is a 47-residue protein with a $\beta\beta\alpha\beta\alpha$ topology, and each of the two helices packs against one face of the β -sheet. The average C_α RMSD over the whole sequence is 4.8 and 3.6 Å over the rigid core. The two helices remain formed, but helix H2 moves to a perpendicular orientation with respect to the native structure (Figure 3A). The sequence used for NMR includes, however, 5 and 11

additional N-terminal and C-terminal residues, respectively. Finally, the protein 1GYZ with four helices at positions 61–67, 75–84, 94–99, and 101–114 has a mean C_α RMSD of 4.2 and 3.5 Å using the full sequence and the rigid core, respectively. In both runs, we find partial unfolding of helix H1 and unfolding of helix H3, leading to unwinding and reorientation of helix H4. Again, the sequence used for NMR includes 58 additional N-terminal residues which are fully unstructured, and H1 and H3 interact in the NMR structure.

Overall, the difference in the C_α RMSDs calculated on the full sequences and the rigid cores is less than 1.2 Å for 15 proteins and reaches 2 Å for two proteins, due to the flexibility of a loop of 17 residues in 1B75 (panel 4E), and two loops of 5 residues in 2B86 (panel 4A) consistent with NMR experiments. Averaged over the 17 proteins, the percentage of native side-chain contacts is 70% and the rigid-core RMSD is 3.1 Å using the time interval 20–30 ns, while the mean RMSD at 20 ns is 4.1 Å using the full sequences.

It is instructive to compare with two recent MD studies using other CG force fields. Using a set of 956 proteins and a two-bead model, Majek found that 58% of the proteins stayed within 5 Å C_α RMSD from their native structures during 20 ns at 300 K and the mean RMSD of the final structures is 4.95 Å.¹⁶ Tested on eight proteins with 17–98 amino acids, the three-bead model developed by Ha-Duong yielded C_α RMSDs varying between 3 and 8 Å from the experimental structures during 200 ns at 300 K.¹⁹

REMD-OPEP4 Simulations on Non-Amyloid and Amyloid Peptides. Table 2 reports a summary of all REMD simulations. For each system, we give its oligomer state ranging between monomer, dimer, and tetramer, the number of replicas, the simulation time per replica, the temperature range, the time interval used for analysis, the nature of the initial structure, and the RMSD cutoff for clustering the conformations. In what follows, our results are compared with experiments and previous computational studies.

Trpzip1 and trpzip2: Two Monomeric β -Hairpin Models. Starting from a random state and using 12 replicas, each of 50 ns, the center of the most populated cluster of trpzip1 at 304 K is found to deviate by 1.7 Å RMSD from the NMR structure.⁹⁴ The NMR structure is characterized by two β -strands at positions 2–5 and 8–11,⁹⁴ and trpzip1 is also considered natively ordered when Trp9 is closer to Trp4 than Trp2, as suggested by unfolding all-atom MD simulations.⁹⁵ In our folded conformation, the native β -strands of trpzip1 are conserved and Trp9 is in contact with Trp4 and not Trp2. Using the same simulation protocol, the center of the most populated cluster of trpzip2 deviates by 1.9 Å RMSD from the NMR structure with the loop found at the right position.

On the basis of CD spectroscopy, the melting temperatures of trpzip1 and trpzip2 were observed at 323 and 345 K, respectively.⁸¹ Here, the T_m , identified by a peak in the heat capacity profile, is calculated at 374 and 360 K for trpzip1 and trpzip2. Using a homemade implicit solvent model coupled to replica exchange Monte Carlo (MC) simulations, Irback et al. found T_m 's of 303 and 305 K by following the network of native hydrogen bonds as a function of temperature.⁹⁶ Using the ABSINTH Hamiltonian based on an implicit model and an all-atom representation coupled to Monte Carlo sampling, Vitalis and Pappu could only fold trpzip1 to within 3.5 Å RMSD from the NMR structure.⁹⁷ Using the PACE force field consisting of a united-protein model and a CG water, Han et al. found a T_m

Table 2. Details of the REMD-OPEP Simulations^a

name	species	time (ns)	number of replicas	T range (K)	time for analysis	initial structure	RMSD cutoff (Å)
trpzip1	monomer	50	12	250–500	10–50	R	2
trpzip2	monomer	50	12	250–500	10–50	R	2
p16-31	monomer	50	14	250–500	10–50	R	1.5
WW domain	monomer	150	16	250–615	50–150	NMR	2
		150	16	250–615	50–150	R	2
<i>ccβ</i>	trimer	200	16	250–615	50–200	NMR	2.5
H1 PrP	monomer	50	14	250–500	10–50	R	2
H1 PrP	dimer	600	12	250–600	100–600	2α	2
H1 PrP	tetramer	800	12	250–600	300–800	4α	4

^aFor each system, we give the oligomer size or species ranging monomer from tetramer, the simulation time per replica, the number of replicas and the temperature range, the time interval used for analysis, the initial configuration (R stands for random state, and α stands for α-helix), and the RMSD cutoff for clustering using all amino acids, unless specified in the text.

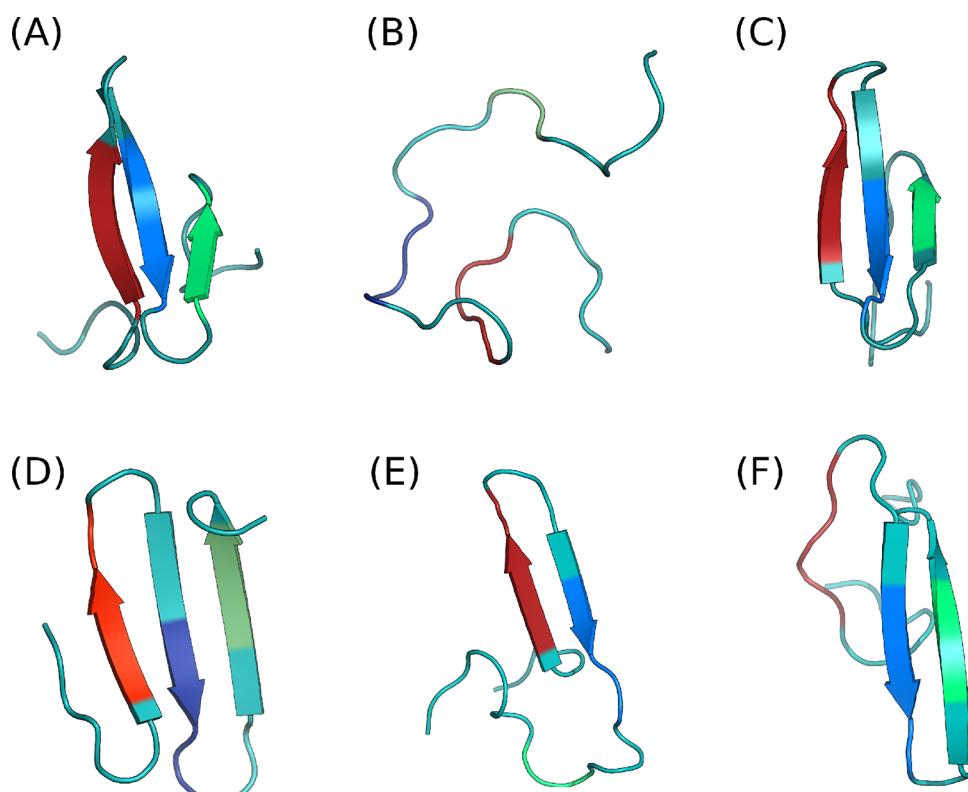


Figure 6. Structures of 1E0M. (A) NMR starting structure with strands spanning 8–12, 17–22, and 27–29. (B) Random starting structure, (C) Predicted structure at 3.8 Å RMSD with strands spanning 7–11, 16–21, and 26–30. (D) Predicted structure at 5.4 Å RMSD with strands spanning 7–11, 15–20, and 25–30. (E) Predicted β-hairpin with strands spanning 7–11 and 16–20. (F) Predicted β-hairpin with strands spanning 16–21 and 26–31.

value of 300–305 K for trpzip2.⁹⁸ Finally, using the CHARMM-CMAP force field with the GB solvent model, Chen et al. found that their REMD simulations were very difficult to converge.⁹⁹

Monomeric p16-31. The p16-31 peptide was found as an independent and autonomous helical folding unit of the gcn4 coiled coil. Its NMR structure at 278 K shows an α helix spanning 18–28.⁸² Starting from a random state with 14 replicas, each of 50 ns, we find that the first cluster at 265 K, with a population of 95%, is characterized by an α-helix spanning 17–25. It has been found by NMR that Arg25 plays a critical role in the helix formation by interacting with the Asn21 and Glu22 side-chains.⁸² These contacts are also formed in our folded conformation.

Monomeric WW Domain Prototype. The NMR structure of the 37-residue WW domain is characterized by three β-strands at positions 8–12, 17–22, and 27–29 with flexibility elsewhere.⁷⁷ We performed two REMD simulations with 16 replicas, each of 150 ns, starting from the NMR (Figure 5A) and one random (Figure 5B) structure. The analysis is based on the 50–150 ns time interval. Figure 6 shows the most populated clusters at 300 K using a C_α RMSD cutoff of 2 Å on residues 7–30. All clusters are found in both simulations. We find that the energy landscape is characterized by two three-stranded β-sheet structures (RMSDs of 3.8 and 5.4 Å in Figure 5C and D), each with a population of 35%, and two β-hairpin structures (Figure 5E and F) with a population of 10 and 5%. The first β-sheet structure displays β-strands spanning residues 7–11, 16–21, and 26–30; the second at positions 7–11, 15–

20, and 25–30. The first β -hairpin displays strands at positions 7–11 and 16–20, i.e., strands 1 and 2 but with a non-native register of H-bonds; the second hairpin spans residues 16–21 and 26–31, i.e., strands 2 and 3 with again a non-native register of H-bonds. Note that β -hairpins were discussed by all-atom MD simulations of the Beta3s three-stranded β -sheet protein.¹⁰⁰

The WW domain was described as a challenge in protein-folding simulations,¹⁰¹ since the 10 μ s all-atom MD trajectory using the CHARMM27 force field results in only α -helical structures.¹⁰² Recently, two extensive all-atom simulations in explicit solvent folded this system from very high to medium resolution using revisited force fields. Very long all-atom MD simulations in explicit solvent using Anton and the modified AMBER and CHARMM force fields folded the WW domain with 1 Å RMSD.^{2,3} All-atom REMD simulations in explicit solvent using the AMBER ff03* force field and 32 replicas each of 1.4 μ s found native-like (RMSD of 2–3 Å) lowest free energy structure.¹⁰³

Some other groups have also reached medium resolution. Combining the zipping and assembly mechanism with all-atom REMD and an implicit solvent model, Ozkan et al. folded two WW domains within 2.2 Å RMSD.¹⁰⁴ Using all-atom replica exchange Monte Carlo simulations and a knowledge-based potential, Yang et al. and Xu et al. found the WW domain with a RMSD of 2.7 Å.^{105,106} Using all-atom DMD simulations, Ding et al. also achieved similar accuracy.¹⁰⁷ However, we note that all these simulations only studied the region spanning the β -strands (i.e., with 25 or 26 amino acids and the unstructured tails removed). It is also instructive to compare with another CG force field using a full-length 37-residue WW domain. The CG UNRES force field coupled to multiplexed REMD simulations produced most native-like β -sheet structures with a RMSD value of 4.8 Å and a population of 10% at 280 K.³⁷

$\text{cc}\beta$ Peptide. Analytical ultracentrifugation indicates that at 277 K the peptide forms a trimer and the CD spectrum confirms the presence of substantial α -structure. At the same temperature, the crystal structure at pH 6.5 shows a three-stranded coiled coil conformation with three parallel α -helices. When the temperature is increased to 310 K, the CD minimum shifts to 217 nm, representative of β -rich structure, and transmission electron microscopy and green birefringence under cross polarized light reveal amyloid fibrils.¹⁰⁸

A REMD simulation using 16 replicas between 250 and 615 K, each of 200 ns, was launched starting from the crystal structure. Figure 7 shows the heat capacity profile and the most populated clusters at three temperatures based on the 25–200 ns time window. The dominant cluster at 266 K with a population of 64.5% is an α -helical trimer deviating by 0.9 Å RMSD from the crystal structure. The presence of a topology with $\alpha\beta$ character and a population of 13% at 266 K does not demonstrate full convergence of the REMD sampling but clearly indicates that the REMD sampling is efficient within 200 ns. In addition, extending each replica by 50 ns does not change the population of the α -helical trimer by more than 1%. Above the melting temperature at 345 K, the peptides undergo an α -to- β conversion. Figure 7 shows the first four clusters at 380 K with a total population of 41.4%. The first cluster with a population of 13% is characterized by three β -hairpin-shaped peptides with β -strands at positions 5–8 and 13–16 in chain 1 and positions 6–8 and 13–15 in chains 2 and 3. The other three clusters illustrate the variety of irregular β -sheet aggregates. In the second (11%) and third (9%) clusters, one

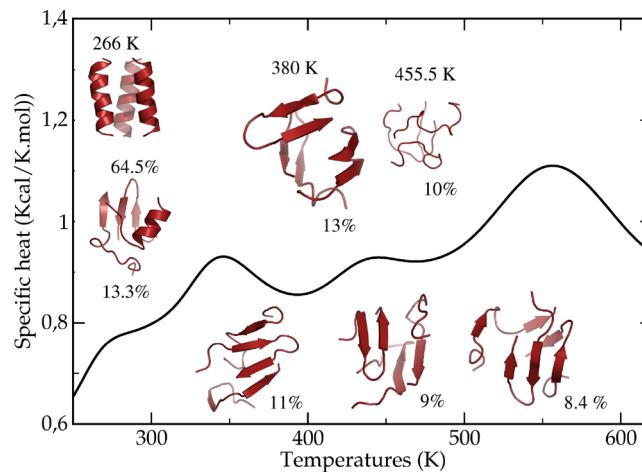


Figure 7. Heat capacity of the $\text{cc}\beta$ peptide as obtained from REMD-OPEP. The most populated clusters at 266, 380, and 455 K are shown (upper). At 380 K, we also show the second, third, and fourth clusters with their populations of 11, 9, and 8%, respectively.

β -hairpin peptide is either sitting near the two other peptides which form a short antiparallel β -sheet (cluster 3) or is sandwiched by the two other peptides allowing the formation of four-stranded β -sheet (cluster 2). In the fourth cluster, the β -sheet organization is slightly more complex, though one peptide adopts a β -hairpin.

It is instructive to compare our results with previous computational studies. Starting from the crystal structure and using REMD coupled to the CHARMM all-atom force field and the EFF1 implicit solvent model, Wales et al. found several fully extended, antiparallel β -sheet trimers in equilibrium with irregular β and β -hairpin aggregates.¹⁰⁹ Starting from random conformations and using DMD simulations with a coarse-grained model, Ding et al. found that the equilibrium ensemble consists of compact coiled mixed β -sheets with a few residues only partially in the β -sheet conformation, and β -hairpin aggregates.²² Interestingly, it is only by studying the aggregation of six peptides that DMD found the existence of fully aligned β -sheets consistent with solid-state NMR of the fibrils. Our β -hairpins are very similar to those observed by Ding²² and Wales,¹⁰⁹ and our irregular β -sheet aggregates were also discussed by Ding.²²

Helix 1 of Prion Protein. The last system studied is the short helix H1 of 10 residues spanning the amino acids 144–153 in the mouse prion protein PrP. The NMR structure of the recombinant monomeric PrP from human, mouse, and many species spanning residues 90–231 or 121–231 displays, in addition to H1, a two-helix bundle H2–H3 at the C-terminal connected by a disulfide bond.¹¹⁰

Structural studies have shown that peptides covering H1¹¹¹ and encompassing H1 and flanking sequences¹¹² are stable under different pH and solvent conditions. This high α -helix propensity of H1 in aqueous solution was confirmed by long all-atom MD trajectories using various force fields.¹¹³ Starting from a random state, our REMD with 14 replicas each of 50 ns reveals that H1 is indeed stable at 300 K, with 95.4% of all conformations forming a full α -helix. Note that simulating each replica on a longer time scale (100 ns) does not change the results. The high stability of helix H1 in PrP monomer has also been discussed by many computational studies using all-atom^{113–118} and OPEP^{119,120} simulations. OPEP-MD simu-

lations have also shown that helix H1 is very robust in PrP dimer.¹²¹

Turbidimetric analysis of the human PrP peptide spanning residues 140–158 and thus covering H1 has shown no aggregation tendency within 180 h.¹²² Aggregation was only observed for the destabilized human variant PrP(140–158)R151G with an approximate lag-time of 101 h. The same experiment on the peptide 137–157 forms, however, amyloids, leading to the hypothesis that interconversion of helix H1 is rather a barrier than a nucleus.^{112,122} All-atom minimizations of various aggregates with two, three, and four H1 peptides from mouse in vacuo and in explicit solvent have also shown that there is a large barrier for the formation of a dimeric β -sheet or β -nucleus from two α helices, but once the barrier is overcome, the addition of H1 peptide chains to the aggregate is energetically favorable.¹²³

The dimer and tetramer of mouse H1 were subjected to REMD-OPEP simulations of 600 and 800 ns per replica starting from two and four parallel α -helices, respectively. The first 100 and 300 ns are excluded from the analysis of the dimer and tetramer, respectively. Figure 8 shows the evolution of α -

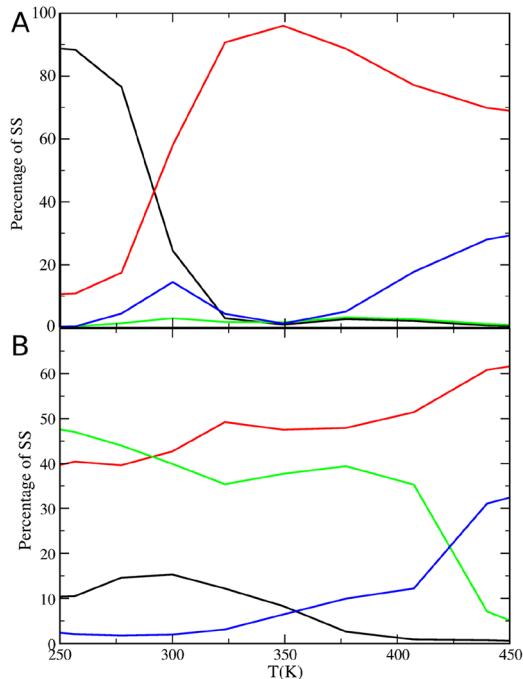


Figure 8. H1 dimer and tetramer: Percentages of secondary structures as a function of T as obtained from REMD-OPEP. H1 dimer (A), H1 tetramer (B), α -helix (black line), coil (red line), β -strand (green line), and turn (blue line).

helix, β -strand, turn, and coil as a function of temperature. Figure 9 reports the center of the three most populated clusters at 277, 299, and 349 K. Note that clustering here uses a chain-independent C_{α} RMSD cutoff of 4 Å (i.e., the numbering of the chains is not considered).

We first analyze the results of the dimer. As seen in Figure 8, between 275 and 300 K, there is an abrupt decrease of α -helix content going from 85 ± 1 to $25 \pm 1\%$ that is compensated by an increase of coil from 17 ± 1 to $58 \pm 1\%$. Over the temperature range between 250 and 450 K, the β -sheet content remains constant at 1%. At 277 K, the two H1 peptides form two parallel α -helices with a population of 85%, while, at 299 K,

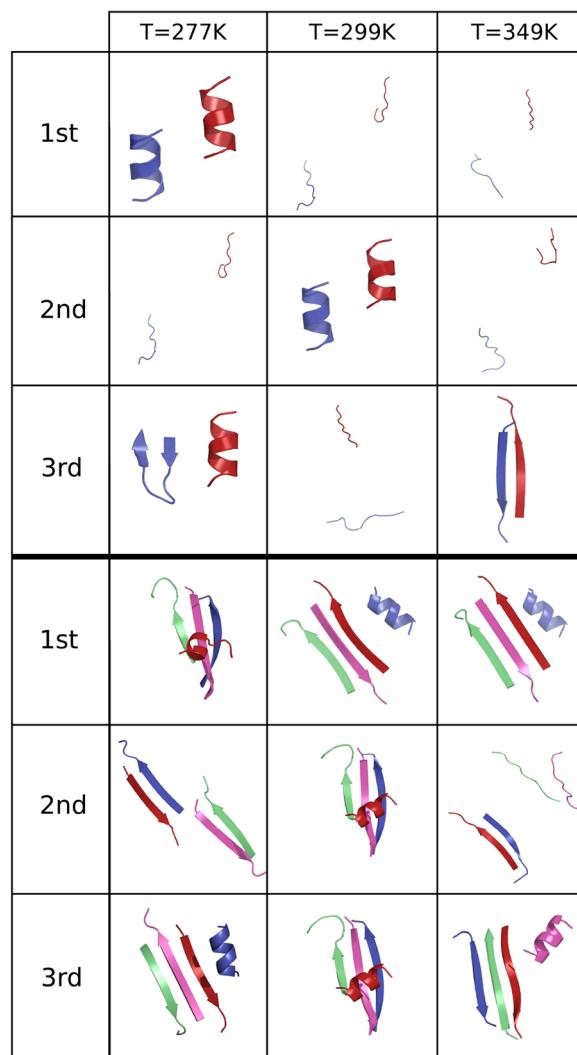


Figure 9. The first three centers of the most populated clusters for H1 dimer and tetramer at 277, 299, and 349 K as obtained from REMD-OPEP. A dark horizontal line separates the dimer from the tetramer. For the tetramer, the populations of clusters 1, 2, and 3 are 17.7, 13.2, and 11.2% at 277 K; 17.3, 13.1, and 7.6% at 299 K; and 10.3, 8.6, and 6.8% at 349 K.

they are primarily random coil with a population of 54% (clusters 1 and 3 in Figure 9). At 349 K, the two H1 chains consist of two separated random coil monomers (clusters 1 and 2) albeit there is a two-stranded β -sheet with a marginal probability of 2% and thus within statistical errors (cluster 3 in Figure 9). Looking at the interpeptide side-chain and main-chain contacts and considering a cutoff distance below 7 Å for association, we find that two H1 chains form a dimer with a population of 89% at 277 K, 33% at 299 K, and 5% at 349 K. All these results indicate that the formation of β -sheets for two H1 peptides is very low at all temperatures, in contrast to the REMD-OPEP results of the $cc\beta$ peptide, where a transition from α to β is observed above the melting temperature. Our results are consistent with previous all-atom minimizations of the H1 dimer.¹²³

The conformational ensemble changes drastically for the tetramer. Averaged over all conformations and amino acids, the secondary structure at 277 K shown on Figure 8 is mostly β -strand ($44 \pm 1\%$) and coil ($40 \pm 1\%$) with substantial α -helix ($33 \pm 1\%$). At 299 and 349 K, the α -helix decreases to 15 and

8% and the random coil content increases to 43 and 48%, but the β -strand remains on the order of 40 and 38%, while the turn content is <10%. The quaternary structures can be described by a total of (6, 6, and 10) clusters at (277, 299, and 349 K), representing (61.3, 54.5, and 52.8%) of all conformations. As seen in Figure 9, the first three clusters at 277 K display an α -helix lying above a mixed parallel/antiparallel three-stranded β -sheet (clusters 1 and 2) and two antiparallel stranded- β -sheets (cluster 3). At 299 K, the first three dominant clusters consist of various mixed parallel/antiparallel three-stranded β -sheets with an α helix, while, at 349 K, we observe the same topology for two clusters and a two-stranded β -sheet + a dimer of two disordered peptides in cluster 2. Looking at the oligomer size distribution and using the same cutoff distance of 7 Å for association, we find that four H1 peptides form a tetramer with a population of 62, 60, and 48% at 279, 299, and 349 K, respectively.

Overall, our findings have three implications. First, the formation of a four-stranded β -sheet is a very improbable event, unlike the formation of a trimeric β -sheet. This is in partial agreement with all-atom minimizations of the H1 trimer and tetramer.¹²³ Second, our oligomer size distribution indicates that formation of a (dimer, tetramer) is not a favorable event (on the order of 33 and 50%) between 300 and 350 K in agreement with turbidimetric analysis of the human PrP peptide spanning residues 140–158, showing no aggregation tendency within 180 h.¹²² Third, our results support the idea that H1 conformational transition might not drive the early steps of full-length PrP interconversion,^{112,122} and the H2–H3 domain is a candidate for the conversion.¹²⁴

CONCLUSIONS

We have presented the coarse-grained OPEP4 force field for proteins, which differs from the OPEP3.2 version, by using a new analytical formulation for the nonbonded interactions and adding specific side-chain–side-chain interactions for α -helix. The OPEP4 force field has been tested on a total of 23 proteins with 10–158 amino acids using MD or REMD simulations. Our results can be summarized as follows.

First, even though we did not optimize the full set of parameters, OPEP4-MD simulations at 300 K on 17 proteins with various secondary structure compositions are satisfactory, leading to a mean C_α RMSD of 3.1 Å for the generated 20–30 ns conformations from the experimental rigid cores. Extending the time scales to 100 ns on five proteins does not change the results. OPEP4 is in between the flexible Majek's CG model of two beads per amino acid (C_α and the side-chain center of mass) that leaves 58% of tested proteins at 300 K after 20 ns within 5 Å from the native fold,¹⁶ and the more complex mapping of MARTINI that imposes, however, restraints on the native α and β secondary structures.^{30,125} Comparison of our RMSDs with all-atom MD simulations in explicit solvent is not an easy task, since implicit solvent and coarse graining accelerate protein dynamics by 1 or 2 orders of magnitude.¹²⁶ In addition, all-atom force fields do not describe similarly the NMR data for folded proteins,¹²⁷ and lead to dissimilar equilibrium ensembles for amyloid oligomers.¹²⁸ Using the revised CHARMM22 force field, the RMSD varies from 1.5 to 3.5 Å from the experimental structures for an ensemble of 12 proteins.³

Second, in contrast to most CG force fields which are tested by MD or Monte Carlo simulations,^{16,17,19,26,33} OPEP4 when coupled to REMD simulations is able to predict the free energy

structures of four prototypical α and β -hairpin sequences (trpzip1, trpzip2, p16-31, and H1 fragment from prion protein) and the $cc\beta$ peptide but overestimates the melting temperatures of the trpzip1, trpzip2, and $cc\beta$ peptides. However, many coarse-grained and all-atom force fields also fail to reproduce the correct melting temperatures of these trpzip peptides^{96–99} or other monomeric proteins.^{98,129} Even using long all-atom MD trajectories in explicit solvent, Shawn et al. found a shift of 40 K between the experimental and calculated T_m 's for many proteins.^{2,3} In addition, we know that the folding of short β -sheet monomeric peptides is heterogeneous and the transition temperature is highly sensitive to the experimental probe employed.¹³⁰

Third, the OPEP4-REMD simulations are able to capture the experimental temperature dependence of the $cc\beta$ peptide from a coiled-coil conformation to amyloids and provide some physical insights into the slow amyloid or non-amyloid character of the H1 prion peptide. The propensity of CG force fields to deal with both amyloid and non-amyloid peptides is not very frequent. To our knowledge, only DMD simulations with seven coarse-grained beads have been applied successfully and extensively to protein structure prediction and amyloid formation,^{22,131} the efficiency of the CG UNRES model¹³² and a generic CG model¹³³ remaining to be explored for amyloids.

In summary, the present study indicates that the OPEP4 force field is effective for protein folding and aggregation, and its combination with the HiRE-RNA force field¹¹ could open the door to the study of protein/RNA assemblies. There is still room for OPEP improvement to reduce the RMSD by MD simulations and better reproduce the melting temperatures. The first direction involves the use of a systematic desolvation barrier.^{59,134} The second direction consists of deriving potentials of mean force for salt-bridges at neutral and acidic pH.¹³⁵ These ideas are currently being tested.

AUTHOR INFORMATION

Corresponding Author

*E-mail: philippe.derreumaux@ibpc.fr.

Present Address

[§]Département de Biologie et de Génomique Structurales, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), Institut National de Santé et de Recherche Médicale (INSERM) U964/Centre National de Recherche Scientifique (CNRS) UMR 1704/Université de Strasbourg, Illkirch, France.

Notes

The authors declare no competing financial interest.

REFERENCES

- (1) Karplus, M.; McCammon, J. A. *Nat. Struct. Biol.* **2002**, *9*, 646–652.
- (2) 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. *Science* **2010**, *330*, 341–346.
- (3) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. *Science* **2011**, *334*, 517–520.
- (4) Snow, C. D.; Nguyen, H.; Pande, V. S.; Gruebele, M. *Nature* **2002**, *420*, 102–106.
- (5) Lane, T. J.; Bowman, G. R.; Beauchamp, K.; Voelz, V. A.; Pande, V. S. *J. Am. Chem. Soc.* **2011**, *133*, 18413–18419.
- (6) Hansmann, U. H.; Okamoto, Y. *Curr. Opin. Struct. Biol.* **1999**, *9*, 177–183.
- (7) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. *J. Phys. Chem. B* **2007**, *111*, 7812–7824.

- (8) Bond, P. J.; Sansom, M. S. P. *J. Am. Chem. Soc.* **2006**, *128*, 2697–2704.
- (9) Knotts, T. A.; Rathore, N.; Schwartz, D. C.; de Pablo, J. J. *J. Chem. Phys.* **2007**, *126*, 084901.
- (10) Ding, F.; Sharma, S.; Chalasani, P.; Demidov, V. V.; Broude, N. E.; Dokholyan, N. V. *RNA* **2008**, *14*, 1164–1173.
- (11) Pasquali, S.; Derreumaux, P. *J. Phys. Chem. B* **2010**, *114*, 11957–11966.
- (12) Levitt, M.; Warshel, A. *Nature* **1975**, *253*, 694–698.
- (13) Tama, F.; Wriggers, W.; Brooks, C. L. *J. Mol. Biol.* **2002**, *321*, 297–305.
- (14) Alemani, D.; Collu, F.; Cascella, M.; Peraro, M. D. *J. Chem. Theory Comput.* **2010**, *6*, 315–324.
- (15) Liwo, A.; Khalili, M.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 2362–2367.
- (16) Májek, P.; Elber, R. *Proteins* **2009**, *76*, 822–836.
- (17) Tozzini, V.; McCammon, J. *Chem. Phys. Lett.* **2005**, *413*, 123–128.
- (18) Arkhipov, A.; Freddolino, P. L.; Imada, K.; Namba, K.; Schulter, K. *Biophys. J.* **2006**, *91*, 4589–4597.
- (19) Basdevant, N.; Borgis, D.; Ha-Duong, T. *J. Phys. Chem. B* **2007**, *111*, 9390–9399.
- (20) Zacharias, M. *Protein Sci.* **2003**, *12*, 1271–1282.
- (21) Ha-Duong, T. *J. Chem. Theory Comput.* **2010**, *6*, 761–773.
- (22) Ding, F.; LaRocque, J. J.; Dokholyan, N. V. *J. Biol. Chem.* **2005**, *280*, 40235–40240.
- (23) Fujitsuka, Y.; Takada, S.; Luthey-Schulten, Z. A.; Wolynes, P. G. *Proteins* **2004**, *54*, 88–103.
- (24) van Giessen, A. E.; Straub, J. E. *J. Chem. Theory Comput.* **2006**, *2*, 674–684.
- (25) Cheon, M.; Chang, I.; Hall, C. K. *Proteins* **2010**, *78*, 2950–2960.
- (26) Rohl, C. A.; Strauss, C. E. M.; Misura, K. M. S.; Baker, D. *Methods Enzymol.* **2004**, *383*, 66–93.
- (27) Derreumaux, P. *J. Chem. Phys.* **1997**, *107*, 1941–1947.
- (28) Irbäck, A.; Sjönnesson, F.; Wallin, S. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 13614–13618.
- (29) Devane, R.; Shinoda, W.; Moore, P. B.; Klein, M. L. *J. Chem. Theory Comput.* **2009**, *5*, 2115–2124.
- (30) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S.-J. *J. Chem. Theory Comput.* **2008**, *4*, 819–834.
- (31) Gopal, S. M.; Mukherjee, S.; Cheng, Y.-M.; Feig, M. *Proteins* **2010**, *78*, 1266–1281.
- (32) Mullinax, J. W.; Noid, W. G. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 19867–19872.
- (33) Moritsugu, K.; Smith, J. C. *Biophys. J.* **2008**, *95*, 1639–1648.
- (34) Thorpe, I. F.; Zhou, J.; Voth, G. A. *J. Phys. Chem. B* **2008**, *112*, 13079–13090.
- (35) Liwo, A.; Czaplewski, C.; Pillardy, J.; Scheraga, H. A. *J. Chem. Phys.* **2001**, *115*, 2323–2347.
- (36) Liwo, A.; Khalili, M.; Czaplewski, C.; Kalinowski, S.; Oldziej, S.; Wachucik, K.; Scheraga, H. A. *J. Phys. Chem. B* **2007**, *111*, 260–285.
- (37) He, Y.; Xiao, Y.; Liwo, A.; Scheraga, H. A. *J. Comput. Chem.* **2009**, *30*, 2127–2135.
- (38) Derreumaux, P. *J. Chem. Phys.* **1999**, *111*, 2301–2310.
- (39) Maupetit, J.; Tufféry, P.; Derreumaux, P. *Proteins* **2007**, *69*, 394–408.
- (40) Derreumaux, P. *Phys. Rev. Lett.* **2000**, *85*, 206–209.
- (41) Forcellino, F.; Derreumaux, P. *Proteins* **2001**, *45*, 159–166.
- (42) Derreumaux, P. *Biophys. J.* **2001**, *81*, 1657–1665.
- (43) Barducci, A.; Bonomi, M.; Derreumaux, P. *J. Chem. Theory Comput.* **2011**, *7*, 1928–1934.
- (44) Wei, G.; Derreumaux, P.; Mousseau, N. *J. Chem. Phys.* **2003**, *119*, 6403–6406.
- (45) Chen, W.; Mousseau, N.; Derreumaux, P. *J. Chem. Phys.* **2006**, *125*, 084911.
- (46) Maupetit, J.; Derreumaux, P.; Tufféry, P. *Nucleic Acids Res.* **2009**, *37*, W498–W503.
- (47) Maupetit, J.; Derreumaux, P.; Tufféry, P. *J. Comput. Chem.* **2010**, *31*, 726–738.
- (48) Thévenet, P.; Shen, Y.; Maupetit, J.; Guyon, F.; Derreumaux, P.; Tufféry, P. *Nucleic Acids Res.* **2012**, *40*, W288–293.
- (49) Santini, S.; Wei, G.; Mousseau, N.; Derreumaux, P. *Structure* **2004**, *12*, 1245–1255.
- (50) Santini, S.; Mousseau, N.; Derreumaux, P. *J. Am. Chem. Soc.* **2004**, *126*, 11509–11516.
- (51) Mousseau, N.; Derreumaux, P. *Acc. Chem. Res.* **2005**, *38*, 885–891.
- (52) Melquiond, A.; Mousseau, N.; Derreumaux, P. *Proteins* **2006**, *65*, 180–191.
- (53) Melquiond, A.; Boucher, G.; Mousseau, N.; Derreumaux, P. *J. Chem. Phys.* **2005**, *122*, 174904.
- (54) Derreumaux, P.; Mousseau, N. *J. Chem. Phys.* **2007**, *126*, 025101.
- (55) Song, W.; Wei, G.; Mousseau, N.; Derreumaux, P. *J. Phys. Chem. B* **2008**, *112*, 4410–4418.
- (56) Wei, G.; Mousseau, N.; Derreumaux, P. *Prion* **2007**, *1*, 3–8.
- (57) Mo, Y.; Lu, Y.; Wei, G.; Derreumaux, P. *J. Chem. Phys.* **2009**, *130*, 125101.
- (58) Lu, Y.; Derreumaux, P.; Guo, Z.; Mousseau, N.; Wei, G. *Proteins* **2009**, *75*, 954–963.
- (59) Chebaro, Y.; Dong, X.; Laghaei, R.; Derreumaux, P.; Mousseau, N. *J. Phys. Chem. B* **2009**, *113*, 267–274.
- (60) Melquiond, A.; Dong, X.; Mousseau, N.; Derreumaux, P. *Curr. Alzheimer Res.* **2008**, *5*, 244–250.
- (61) Chebaro, Y.; Mousseau, N.; Derreumaux, P. *J. Phys. Chem. B* **2009**, *113*, 7668–7675.
- (62) Chebaro, Y.; Derreumaux, P. *Proteins* **2009**, *75*, 442–452.
- (63) Derreumaux, P.; Wilson, K.; Vergoten, G.; Peticolas, W. *J. Phys. Chem.* **1989**, *93*, 1338–1350.
- (64) Derreumaux, P.; Vergoten, G.; Lagant, P. *J. Comput. Chem.* **1990**, *11*, 560–568.
- (65) Raibaud, S.; Lebars, I.; Guillier, M.; Chiaruttini, C.; Bontems, F.; Rak, A.; Garber, M.; Allemand, F.; Springer, M.; Dardel, F. *J. Mol. Biol.* **2002**, *323*, 143–151.
- (66) Neri, D.; Billeter, M.; Wüthrich, K. *J. Mol. Biol.* **1992**, *223*, 743–767.
- (67) Grummitt, C. G.; Townsley, F. M.; Johnson, C. M.; Warren, A. J.; Bycroft, M. *J. Biol. Chem.* **2008**, *283*, 23326–23332.
- (68) Iwahara, J.; Kigawa, T.; Kitagawa, K.; Masumoto, H.; Okazaki, T.; Yokoyama, S. *EMBO J.* **1998**, *17*, 827–837.
- (69) Berjanskii, M. V.; Riley, M. I.; Xie, A.; Semenchenko, V.; Folk, W. R.; Doren, S. R. V. *J. Biol. Chem.* **2000**, *275*, 36094–36103.
- (70) Evans, S. P.; Bycroft, M. *J. Mol. Biol.* **1999**, *291*, 661–669.
- (71) Bateman, A.; Bycroft, M. *J. Mol. Biol.* **2000**, *299*, 1113–1119.
- (72) Ross, S.; Sarisky, C.; Su, A.; Mayo, S. *Protein Sci.* **2001**, *10*, 450–454.
- (73) Skelton, N. J.; Kordel, J.; Chazin, W. *J. Mol. Biol.* **1995**, *249*, 441–462.
- (74) Gallagher, T.; Alexander, P.; Bryan, P.; Gilliland, G. *Biochemistry* **1994**, *33*, 4721–4729.
- (75) Mercier, K. A.; Mueller, G. A.; Acton, T. B.; Xiao, R.; Montelione, G. T.; Powers, R. *Biomol. NMR Assignments* **2009**, *3*, 191–194.
- (76) Stoldt, M.; Wohner, J.; Gorlach, M.; Brown, L. *EMBO J.* **1998**, *17*, 6377–6384.
- (77) Macias, M. J.; Gervais, V.; Civera, C.; Oschkinat, H. *Nat. Struct. Biol.* **2000**, *7*, 375–379.
- (78) Park, S.; Takeuchi, K.; Wagner, G. *J. Biomol. NMR* **2006**, *34*, 203–208.
- (79) Campos-Olivas, R.; Bruix, M.; Santoro, J.; Lacadena, J.; del Pozo, A. M.; Gavilanes, J. G.; Rico, M. *Biochemistry* **1995**, *34*, 3009–3021.
- (80) Musacchio, A.; Noble, M.; Paupit, R.; Wierenga, R.; Saraste, M. *Nature* **1992**, *359*, 851–855.
- (81) Cochran, A. G.; Skelton, N. J.; Starovasnik, M. A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 5578–5583.

- (82) Steinmetz, M. O.; Jelesarov, I.; Matousek, W. M.; Honnappa, S.; Jahnke, W.; Missimer, J. H.; Frank, S.; Alexandrescu, A. T.; Kammerer, R. A. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 7062–7067.
- (83) Kammerer, R. A.; Kostrewa, D.; Zurdo, J.; Detken, A.; Garcia-Echeverria, C.; Green, J. D.; Müller, S. A.; Meier, B. H.; Winkler, F. K.; Dobson, C. M.; Steinmetz, M. O. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 4435–4440.
- (84) Andersen, H. C. *J. Comput. Phys.* **1983**, *52*, 24–34.
- (85) Berendsen, H.; Postma, J.; van Gunsteren, W.; Nola, A. D.; Haak, J. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- (86) Spill, Y.; Pasquali, S.; Derreumaux, P. *J. Chem. Theory Comput.* **2011**, *7*, 1502–1510.
- (87) Chodera, J. D.; Swope, W. C.; Pitera, J. W.; Seok, C.; Dill, K. A. *J. Chem. Theory Comput.* **2007**, *3*, 26–41.
- (88) Frishman, D.; Argos, P. *Proteins* **1995**, *23*, 566–579.
- (89) Spoel, D. V. D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comput. Chem.* **2005**, *26*, 1701–1718.
- (90) McLachlan, A. *Acta Crystallogr., Sect. A* **1982**, *38*, 871–873.
- (91) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577–2637.
- (92) Derreumaux, P. *J. Chem. Phys.* **1997**, *106*, 5260–2370.
- (93) Liwo, A.; Pincus, M. R.; Wawak, R. J.; Rackovsky, S.; Scheraga, H. A. *Protein Sci.* **1993**, *2*, 1715–1731.
- (94) Snow, C. D.; Qiu, L.; Du, D.; Gai, F.; Hagen, S. J.; Pande, V. S. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 4077–4082.
- (95) Settanni, G.; Fersht, A. R. *Biophys. J.* **2008**, *94*, 4444–4453.
- (96) Irbäck, A.; Mitternacht, S.; Mohanty, S. *PMC Biophys.* **2009**, *2*.
- (97) Vitalis, A.; Pappu, R. V. *J. Comput. Chem.* **2009**, *30*, 673–699.
- (98) Han, W.; Wan, C.-K.; Wu, Y.-D. *J. Chem. Theory Comput.* **2010**, *6*, 3390–3402.
- (99) Chen, J.; Im, W.; Brooks, C. L. *J. Am. Chem. Soc.* **2006**, *128*, 3728–3736.
- (100) Zheng, W.; Qi, B.; Rohrdanz, M. A.; Caflisch, A.; Dinner, A. R.; Clementi, C. *J. Phys. Chem. B* **2011**, *115*, 13065–13074.
- (101) Freddolino, P. L.; Harrison, C. B.; Liu, Y.; Schulter, K. *Nat. Phys.* **2010**, *6*, 751–758.
- (102) Freddolino, P. L.; Liu, F.; Gruebele, M.; Schulter, K. *Biophys. J.* **2008**, *94*, L75–L77.
- (103) Mittal, J.; Best, R. B. *Biophys. J.* **2010**, *99*, L26–L28.
- (104) Ozkan, S. B.; Wu, G. A.; Chodera, J. D.; Dill, K. A. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 11987–11992.
- (105) Yang, J. S.; Chen, W. W.; Skolnick, J.; Shakhnovich, E. I. *Structure* **2007**, *15*, 53–63.
- (106) Xu, J.; Huang, L.; Shakhnovich, E. I. *Proteins* **2011**, *79*, 1704–1714.
- (107) Ding, F.; Tsao, D.; Nie, H.; Dokholyan, N. V. *Structure* **2008**, *16*, 1010–1018.
- (108) Kammerer, R. A.; Steinmetz, M. O. *J. Struct. Biol.* **2006**, *155*, 146–153.
- (109) Strodel, B.; Fitzpatrick, A. W.; Vendruscolo, M.; Dobson, C. M.; Wales, D. J. *J. Phys. Chem. B* **2008**, *112*, 9998–10004.
- (110) Riek, R.; Hornemann, S.; Wider, G.; Billeter, M.; Glockshuber, R.; Wüthrich, K. *Nature* **1996**, *382*, 180–182.
- (111) Liu, A.; Riek, R.; Zahn, R.; Hornemann, S.; Glockshuber, R.; Wüthrich, K. *Biopolymers* **1999**, *51*, 145–152.
- (112) Ziegler, J.; Sticht, H.; Marx, U. C.; Müller, W.; Rösch, P.; Schwarzinger, S. *J. Biol. Chem.* **2003**, *278*, 50175–50181.
- (113) Dima, R. I.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 15335–15340.
- (114) Simone, A. D.; Zagari, A.; Derreumaux, P. *Biophys. J.* **2007**, *93*, 1284–1292.
- (115) Campos, S. R. R.; Machuqueiro, M.; Baptista, A. M. *J. Phys. Chem. B* **2010**, *114*, 12692–12700.
- (116) Rossetti, G.; Giachin, G.; Legname, G.; Carloni, P. *Proteins* **2010**, *78*, 3270–3280.
- (117) Gorfe, A. A.; Caflisch, A. *FASEB J.* **2007**, *21*, 3279–3287.
- (118) van der Kamp, M. W.; Daggett, V. *J. Mol. Biol.* **2010**, *404*, 732–748.
- (119) Santini, S.; Derreumaux, P. *Cell. Mol. Life Sci.* **2004**, *61*, 951–960.
- (120) Santini, S.; Claude, J.-B.; Audic, S.; Derreumaux, P. *Proteins* **2003**, *51*, 258–265.
- (121) Chebaro, Y.; Derreumaux, P. *J. Phys. Chem. B* **2009**, *113*, 6942–6948.
- (122) Ziegler, J.; Viehrig, C.; Geimer, S.; Rösch, P.; Schwarzinger, S. *FEBS Lett.* **2006**, *580*, 2033–2040.
- (123) Morrissey, M. P.; Shakhnovich, E. I. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 11293–11298.
- (124) Chakroun, N.; Prigent, S.; Dreiss, C. A.; Noinville, S.; Chapuis, C.; Frernali, F.; Rezaei, H. *FASEB J.* **2010**, *24*, 3222–3231.
- (125) Bond, P. J.; Sansom, M. S. P. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 2631–2636.
- (126) Rao, F.; Settanni, G.; Guarnera, E.; Caflisch, A. *J. Chem. Phys.* **2005**, *122*, 184901.
- (127) Lindorff-Larsen, K.; Maragakis, P.; Piana, S.; Eastwood, M. P.; Dror, R. O.; Shaw, D. E. *PLoS One* **2012**, *7*, e32131.
- (128) Nguyen, P. H.; Li, M. S.; Derreumaux, P. *Phys. Chem. Chem. Phys.* **2011**, *13*, 9778–9788.
- (129) Czaplewski, C.; Kalinowski, S.; Liwo, A.; Scheraga, H. A. *J. Chem. Theory Comput.* **2009**, *5*, 627–640.
- (130) Yang, W. Y.; Pitera, J. W.; Swope, W. C.; Gruebele, M. *J. Mol. Biol.* **2004**, *336*, 241–251.
- (131) Ding, F.; Buldyrev, S. V.; Dokholyan, N. V. *Biophys. J.* **2005**, *88*, 147–155.
- (132) Rojas, A. V.; Liwo, A.; Scheraga, H. A. *J. Phys. Chem. B* **2011**, *115*, 12978–12983.
- (133) Bereau, T.; Deserno, M. *J. Chem. Phys.* **2009**, *130*, 235106.
- (134) Kaya, H.; Chan, H. S. *J. Mol. Biol.* **2003**, *326*, 911–931.
- (135) Makowski, M.; Liwo, A.; Scheraga, H. A. *J. Phys. Chem. B* **2011**, *115*, 6130–6137.