Article

Comparison of Structure Determination Methods for Intrinsically Disordered Amyloid-β Peptides

K. Aurelia Ball,[†] David E. Wemmer,[†,‡] and Teresa Head-Gordon*[,†,‡,§,⊥]

[†]Graduate Group in Biophysics, Berkeley, California 94720, United States

[‡]Department of Chemistry, University of California, Berkeley, California 94720, United States

[§]Department of Bioengineering, University of California, Berkeley, California 94720, United States

[⊥]Department of Chemical and Biomolecular Engineering, University of California, Berkeley, California 94720, United States

Ⓢ *Supporting Information*

ABSTRACT: Intrinsically disordered proteins (IDPs) represent a new frontier in structural biology since the primary characteristic of IDPs is that structures need to be characterized as diverse ensembles of conformational substates. We compare two general but very different ways of combining NMR spectroscopy with theoretical methods to derive structural ensembles for the disease IDPs amyloid-β 1−40 and amyloid-β 1−42, which are associated with Alzheimer's Disease. We analyze the performance of *de novo* molecular dynamics and knowledge-based approaches for generating structural ensembles by assessing their ability to reproduce a range of NMR experimental observables. In addition to the comparison of computational methods, we also evaluate the relative value of different types of NMR data for refining or validating the IDP structural ensembles for these important disease peptides.
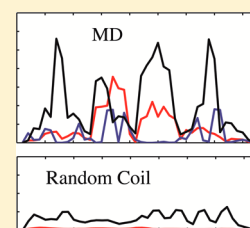
INTRODUCTION

Experimental approaches such as X-ray and electron crystallography and microscopy have traditionally excelled at determining the structure of single folded proteins[1,2] and large protein complexes.[3] However, intrinsically disordered proteins (IDPs) are not amenable to these static structural determination methods.[4] IDPs represent a new frontier in structural biology in that the IDP structure must be characterized as a diverse ensemble of interconverting conformational substates, as opposed to a single dominant 3D structure.[5] This necessitates an adjustment in the core methodology of protein structure determination for this class of protein.

The experimental *identification* of proteins with global intrinsic disorder can be performed using various spectroscopic techniques including circular dichroism (CD), NMR, infrared spectroscopy (IR), UV spectroscopy, and fluorescence spectroscopy.[6,7] CD and IR report on the amount of secondary structure, while lack of chemical shift dispersion in NMR spectra is a good indication of high flexibility. Hydrodynamic techniques such as SAXS, gel filtration, and dynamic light scattering can also aid in IDP identification as they report on the radius of the protein, which is often larger for an IDP or denatured protein than a folded protein of the same mass. Lack of a cooperative folding transition, solubility at high temperatures, and proteolytic sensitivity are also attributes of IDPs that are useful in forming a complete picture of a certain protein's level of disorder. A subset of these techniques is generally employed to determine that a protein is an IDP.

Recently increased importance has been placed on characterizing the conformational substates within IDP ensembles since they each may have distinct functional roles[7−13] or could lead to hypotheses about disease origin.[14] In order to achieve both better ensemble classification and a detailed description of conformational substates, we must critically assess how we build these complex structural ensembles from experimental data and theoretical models. NMR is the experimental tool of choice for characterizing the solution structure and dynamics of biological molecules since it reports on the native distribution of conformations in an aqueous environment, and more importantly is a dynamical experiment that probes the nanosecond to millisecond time scales of conformational motion.[4,15,16] Observables from these experiments include chemical shifts, which are characteristic of functional groups and their surrounding environment, and spin−spin couplings (J-couplings), which independently report on backbone dihedral angles. In addition, through-space dipole−dipole interactions give rise to the nuclear Overhauser effect (NOE) that reports on tertiary structure contacts, and more recently, residual dipolar couplings (RDCs) have been used to describe the relative orientation of spatially separated regions of a protein.[17−20] Paramagnetic relaxation enhancements (PREs), which can produce longer distance restraints than NOEs have also been used in the context of IDPs;[21−24] however, this measurement requires chemical modification of the protein with a nitroxide spin label or an amino-terminal copper binding motif, which sometimes requires sequence modification to

attach the probe, and which may perturb the monomeric IDP conformations.[25,26]

IDPs typically convert between conformations faster than the ca. nanoseconds−milliseconds time scale of the NMR experiment, leading to an averaging of the NMR observables across structural subpopulations. This uniform average hinders the structural characterization of all the conformational substates, and can even obscure the overall ensemble classification, as we will see for the amyloid peptides in this study. Building the connection between the averaged NMR observables and the complete IDP structural ensemble therefore depends critically on computational models.[27] The goal of the computational model is to provide a properly weighted set of the diverse subpopulations of the IDP most consistent with the NMR observables and perhaps other experimental measures such as circular dichrosim,[28] small-angle X-ray scattering,[29,30] or PREs.[23,31] Thus, multiple types of NMR or other experimental observables are necessary for validation of the computational model.[16,30]

Currently there are two primary but very different computational approaches to building an IDP structural ensemble, which can be loosely contrasted as first principle or *de novo* molecular dynamics (MD) methods versus knowledge-based approaches. The *de novo* approach implements MD simulations based on the theoretical foundations of statistical mechanical sampling and model-derived potential energy force fields. *De novo* MD generates a structural ensemble that is representative of given thermodynamic conditions according to the force field employed, i.e., a Boltzmann weighted ensemble of conformational subpopulations and their time scales, independent of experimental input. The MD trajectories also allow calculation of the time correlation functions that underlie the NMR experiment. The complementary use of MD and NMR data to determine structure and dynamics of folded and unfolded proteins has been a highly active area over the last two decades,[32,33] particularly for relaxation measurements that require a dynamical interpretation of the NMR data at the picosecond and nanosecond time scales.[34] For the *de novo* MD method, multiple NMR or other experimental data are necessary to *validate* the MD ensemble through direct back-calculation of observables, many of which depend on the time scales of motion, in order to directly compare to the experiment. Once validated, MD simulations provide a prediction of the complete IDP structural ensemble, allowing overall classification as well as the study of individual conformational substates, which can be analyzed with some confidence.

In contrast, we define knowledge-based approaches as those that use experimental NMR information directly to derive the structural ensemble. Such methods are the foundation of NMR structure determination of folded proteins using experimentally derived conformational constraints based on chemical shifts, J-couplings, and NOE data embodied in software packages such as CANDID,[35] CYANA,[36] and X-Plor-NIH.[37,38] While MD is often used to generate atomistic predictions independent of NMR experimental input, as in our *de novo* method, a number of researchers have advanced the combination of applying knowledge from NMR to restrain the MD ensemble.[22,24,39−41] For example, MD simulations have been combined with RDC restraint data for folded proteins[40] that then allows for the analysis of other features of the ensemble, such as conformational fluctuations. NMR restrained MD has also been applied to IDPs such as α-synuclein, a disease protein indicted in

Parkinson's disease. This study incorporated distance restraints derived from PRE experiments in order to guide the MD so that the protein's radius of gyration distribution is in good agreement with the experimental value.[39]

Other knowledge-based approaches for IDPs forego MD simulations altogether and instead use an extensive set of statistical coil conformations;[16,42] this starting pool, which can be generated using a variety of heuristics, can be thought of as a basis set of structures. Subsequently, the starting pool of structures is then culled for the subset of conformations that are in best agreement with experimental data to create the IDP ensemble. In the energy-minima mapping and weighting (EMW) method, Stultz and co-workers used end-to-end distance restraints to develop a pool of conformations with varying radii of gyration; they then selected, via Monte Carlo, a weighted ensemble of 15 structures to optimize the agreement with experimental ^{13}C and ^{15}N chemical shifts and J-couplings.[43,44] Blackledge and co-workers have developed the program *Flexible-Meccano* to create a pool of structures based on random coil backbone dihedral angles, on which they employ a genetic search algorithm in their ASTEROIDS software program to select structures that together best match experimental chemical shifts, PREs, or RDCs.[16,45,46]

The ENSEMBLE method, developed by Forman-Kay and co-workers, typically defines the starting pool of IDP conformational states as an ensemble of extended or random coil states generated using TraDES, with an option for biasing the secondary structure of the ensemble at certain places in the sequence that are known to be partially structured.[23,29,42,47] Structures are selected from this pool using a Monte Carlo selection algorithm with an energy-weighting scheme for each type of experimental input. The ENSEMBLE program includes modules for several different experimental data types including chemical shifts, RDCs, PREs, J-couplings, and contact distances derived from NOEs, and is a user-friendly and publically available software package.[42] Although there are some specific differences, ENSEMBLE is largely representative of the knowledge-based approaches and is qualitatively equivalent to the combination of *Flexible-Meccano* and ASTEROIDS software.[16,45,46] It is important to note that such techniques largely ignore the inherent dynamical information of certain types of the NMR data that can be important for discriminating between different IDP structural ensembles.

The primary objective of this work is to compare the *de novo* and knowledge-based approaches for deriving IDP structural ensembles in context of the intrinsically disordered Alzheimer's disease peptides amyloid-β 1−40 (Aβ40) and amyloid-β 1−42 (Aβ42). We implement the ENSEMBLE knowledge-based method by building an ensemble from a pool of statistical coil structures, and compare this knowledge-based ensemble to MD generated ensembles, which are qualitatively different in that they are comprised of mostly cooperative secondary structure and tertiary contacts. This comparison also exposes the relative utility of different types of NMR data for refining or validating the IDP computational ensemble. We find that chemical shifts and J-coupling constants are not particularly useful for distinguishing between qualitatively different IDP ensembles of the amyloid-β peptides. Finally we show that the combination of *de novo* MD methods that provide Boltzmann weighted samples with the ability to measure time correlation functions, and knowledge-based methods for conformation selection, provides the best agreement with the NMR data.

■ METHODS

Back-Calculation of NMR Observables. In order to evaluate the alternative ensembles produced by knowledge-based and *de novo* approaches, we need a method of calculating the chemical shifts, J-coupling constants, RDCs, and $^1H-^1H$ NOEs as averages over the entire computationally generated structural ensemble for comparison with experimental values.[48] General purpose chemical shift calculators such as SHIFTX[49] and SHIFTS[50] describe the isotropic shielding of the applied magnetic field for the given atom, a quantity that depends sensitively on the local electronic structure environment.[49−51] Even for folded proteins with a dominant native conformer, each atom type can exist in many different local environments, and for disordered peptides and proteins the ensemble average reflects an even more diverse set of chemical environments. This makes an accurate calculation of chemical shifts quite a challenge for IDPs. Whether one uses SHIFTX (used in ENSEMBLE) or SHIFTS to calculate chemical shifts, the results generated by the two programs are consistent when applied to amyloid-β and averaged over the structural ensembles.[14,48,52] We report results using SHIFTX in this work.

To calculate the scalar coupling constants, $^3J_{H_NH}{}^\alpha$, we used the Karplus equation[53]

$$J(\phi) = A \cos^2(\phi - 60) + B \cos(\phi - 60) + C \quad (1)$$

where ϕ indicates the protein backbone dihedral angle, with coefficients $A = 6.51$, $B = -1.76$, and $C = 1.60$ corresponding to the parameter set by Vuister and Bax.[54] However, Sgourakis and co-workers[55,56] and our own previous work on Aβ42[14,48] found that the MD results exhibited a systematic shift of calculated J-couplings with the experimental scalar couplings of Wang and co-workers. We delved into this issue more deeply and derived an analytical correction, described in recent work,[14] that must be applied to the original experimental J-coupling data from Wang and co-workers.[57] In particular, it has been corrected for a missing relaxation that makes scalar couplings determined from the H_NH_α 3D experiment consistently lower than those from COSY splittings by a small amount[54] (from ~1−5%). The J-coupling values are also averaged over all structures in the ensemble as in,[14,48,52] and then the calculations can be compared to the corrected experimental $^3J_{H_NH}{}^\alpha$ values for both amyloid peptides.

The standard method in the field for calculation of RDCs is the PALES[58] program, which we have used previously for the Aβ40 and Aβ42 MD ensembles.[14,48,52] The program computes the RDC by using steric properties of the molecule to generate a global alignment orientation. Then, the angle between the backbone amide bond vectors and the external magnetic field is used to calculate the RDC for each conformation, and the RDCs are averaged over all conformations of a given ensemble. The ENSEMBLE program by contrast, evaluates RDCs using a local alignment program developed in the Forman-Kay lab, where 15 residue segments along the protein are aligned separately over the ensemble of structures.[18] The local RDCs (L-RDCs) are also averaged over all conformations of a given ensemble. This local alignment has lower computational cost and has been shown to give similar results to PALES, hence L-RDCs, rather than RDCs generated from a global alignment algorithm, are optimized in the standard implementation of the ENSEMBLE approach. Similarly, the ASTEROIDS program by default employs a local alignment tensor to optimize ensemble agreement with experimental RDC data.[59] We also note that

the PALES alignment and RDC calculation were developed for folded proteins, and their application to IDPs assumes individual IDP conformations behave similarly to folded proteins during the RDC experiment, which may not be the case for IDPs such as amyloid-β.

We also evaluate the $^1H-^1H$ NOESY (or ROESY) spectra as we have described in previous studies[14,48,52] by calculating the intensity of the NOE cross-peaks

$$I(t_{mix}) = X e^{-\Lambda t_{mix}} X^{-1} I(0) \quad (2)$$

where X and Λ are the eigenvectors and eigenvalues of the full relaxation matrix, composed of diagonal elements

$$\rho_{ii} = \sum_{j=1,\neq i}^{n} \frac{1}{10} K^2 \left[3J_{ij}(2\omega_0) + \frac{3}{2}J_{ij}(\omega_0) + \frac{1}{2}J_{ij}(0) \right] \quad (3a)$$

and off-diagonal elements

$$\sigma_{ij} = \frac{1}{10} K^2 \left[3J_{ij}(2\omega_0) - \frac{1}{2}J_{ij}(0) \right] \quad (3b)$$

that are comprised of appropriate combinations of the spectral density functions

$$J(\omega) = \sum_{i=1}^{N} \left[a_i \frac{2\tau_i}{1 + \omega^2 \tau_i^2} \right] \quad (4)$$

evaluated at the relevant Larmor frequencies, ω, and where K is given by

$$K = \frac{\mu_0}{4\pi r_{eff}^3} \hbar \gamma_H^2 \quad (5)$$

γ_H is the gyromagnetic ratio of ^1H, μ_0 is the permeability of free space, and $\hbar$ is Planck's constant. r_{eff} is the distance between the hydrogen atoms raised to −6 power, averaged over all structures in the ensemble and then raised to the −1/6 power to convert back to units of distance. These calculations account for all hydrogen atoms explicitly (including all methyl or methylene groups) and hence r_{eff} and correlation functions for every pair of hydrogen atoms are evaluated.[14,48,52] The spectral density function for each atom pair is calculated as the Fourier transform of the correlation function for the pair vector and water proton coordinates are ignored, as is the standard assumption in the NMR experiment.

Finally we calculate $^1H-^{15}N$ NOEs as we did in refs 14 and 48 by evaluating the steady state NOE enhancement factor of the ^{15}N spin by the ^1H NOE according to

$$\varepsilon_{NOE} = 1 + \frac{\gamma_H \sigma_{HN}}{\gamma_N R_Z^{(N)}} \quad (6)$$

where γ_H and γ_N are the gyromagnetic ratios of ^1H and ^{15}N, respectively. The $^1H-^{15}N$ cross-relaxation rate constant is given by

$$\sigma_{HN} = \frac{1}{10} K^2 \left[3J_{HN}(\omega_{0,H} + \omega_{0,N}) - \frac{1}{2}J_{HN}(\omega_{0,H} - \omega_{0,N}) \right] \quad (7)$$

and the ^{15}N self-relaxation by

$$R_Z^{(N)} = \frac{1}{2} K^2 \left[\frac{3}{2}J_{HN}(\omega_{0,N}) + 3J_{HN}(\omega_{0,H} + \omega_{0,N}) \right.$$
$$\left. + \frac{1}{2}J_{HN}(\omega_{0,H} - \omega_{0,N}) \right] \quad (8)$$

In this case, $J_{HN}(\omega)$ is the spectral density function for the $^1H-^{15}N$ covalently bonded pair.

Note that the homonuclear and heteronuclear NOE calculations require correlation time information about the vector between each pair of atoms given by τ in eq 4. This dynamic information is naturally supplied by the *de novo* MD method, which allows direct measurement of the autocorrelation of the interatomic vector over the time of the simulation. However, dynamics are not considered in the generation of ensembles that are used in the knowledge-based approach. This is an inherent limitation of ensembles generated from a static perspective only, which we discuss further below.

■ **RESULTS**

To determine an IDP's level of disorder, we first generate several alternative ensembles, compare these ensembles to the available NMR data, and select the best validated ensemble. We consider the creation of three qualitatively different conformational ensembles that are typically used in the knowledge-based approaches[16,42,45,46] for the Aβ40 and Aβ42 peptides.

The common null hypothesis is that the disordered peptides can be well represented by a random coil (RC) ensemble. The second type of statistical coil (Pred-SS) ensemble is also random, but incorporates bioinformatics-based knowledge about what secondary structure category is more likely for a given residue in the amino acid sequence. In this case, the random ensemble is biased to contain a statistical probability of predicted secondary structure on a per residue basis, but no cooperative secondary structure such as α-helices, β-hairpins, or β-sheets are generated from the random secondary structure assignments. Computational methods such as TraDes and *Flexible-Meccano* are used to generate these type of random or statistical coil ensembles.[16,42] Finally, a fully knowledge-based approach is considered, which culls the RC or, in this case, Pred-SS ensembles to derive a subset of conformations that best agrees with the NMR data (Pred-SS-ENS). The ENSEMBLE software package provides a working example of the knowledge-based approach that performs this biased selection and which has been successfully applied to a range of IDPs. Each of the above three ensembles can then be compared against the ensembles generated by *de novo* MD for both of the IDPs Aβ40 and Aβ42. Details of the *de novo* MD approach applied to amyloid-β can be found in other publications.[14,48,52] We also consider an additional fifth ensemble (MD-ENS) that combines the knowledge-based and *de novo* MD approaches, by using ENSEMBLE to select structures from the *de novo* MD starting pool, and which is described in our recent study.[14]

Table 1 shows the average radius of gyration (Rg) values for each type of ensemble. We see that the order from most extended to most compact proceeds as Pred-SS > RC > Pred-SS-ENS > MD ∼ MD-ENS, and thus the alternative ensembles span a range of IDP classifications by the ⟨Rg⟩ measure.[14] Figure 1 provides the propensities for the Pred-SS, Pred-SS-ENS, *de novo* MD, and MD-ENS ensembles to form turns, antiparallel β-strands, or helical structure by residue for Aβ42. We do not show the secondary structure profiles for the RC ensemble since it is similar to the Pred-SS ensemble (see the Supporting Information).

This plot emphasizes that the MD-based ensembles are qualitatively different from the RC or Pred-SS ensembles, in that the Aβ40 and Aβ42 peptides samples some type of structured conformations in ∼99% of the MD ensemble,

Table 1. Comparison between Random Coil (RC), Predicted Secondary Structure (Pred-SS), *de novo* MD (MD), and ENSEMBLE Optimized Pred-SS-ENS and MD-ENS Ensembles[a]

Aβ40 peptide	average property					
ensemble type	Rg (Å)	χ^2_δ (Hα)	χ^2_δ (HN)	χ^2_δ (Cα)	χ^2_δ (Cβ)	$^3J_{H_NH}{}^a$
RC	16.9 ± 3.2	0.20	0.13	0.29	0.34	0.80 (1.20)
Pred-SS	19.3 ± 3.6	0.45	0.10	0.67	0.49	1.09 (2.23)
Pred-SS-ENS	15.6 ± 3.3	0.41	0.11	0.52	0.45	0.88 (1.46)
MD	14.7 ± 4.8	0.58	0.36	0.69	0.70	0.99 (1.82)
MD-ENS	15.0 ± 4.1	0.30	0.34	0.46	0.36	0.62 (0.72)

Aβ42 peptide	average property					
ensemble type	Rg (Å)	χ^2_δ (Hα)	χ^2_δ (HN)	χ^2_δ (Cα)	χ^2_δ (Cβ)	$^3J_{H_NH}{}^a$
RC	17.3 ± 3.3	0.22	0.15	0.41	0.25	0.49 (0.46)
Pred-SS	19.9 ± 3.8	0.49	0.10	1.06	0.38	0.85 (1.34)
Pres-SS-ENS	16.2 ± 3.6	0.40	0.12	0.71	0.29	0.67 (0.85)
MD	13.1 ± 4.3	0.54	0.48	0.98	0.52	0.99 (1.83)
MD-ENS	13.1 ± 2.8	0.33	0.37	0.51	0.34	0.60 (0.67)

[a]For the radius of gyration (Rg) values, we report both the ensemble average and RMSD. For chemical shifts, we report χ^2 that measures agreement between the computational ensembles and the experimentally measured chemical shifts: $\chi^2 < 1$ indicates no disagreement with experiment within SHIFTX calculator error. We also report the $^3J_{H_NH}{}^a$ RMSD (χ^2). Some data reproduced from ref 14.

including complex βstrand formation.[14,48] From this, we conclude that the radius of gyration trends stem from the much larger propensity for the MD ensembles to form *cooperative* secondary structure and collapsed tertiary contacts, as opposed to the random or knowledge-based ensembles that do not generate contiguous blocks of secondary structure, and hence are more extended on average. Although the secondary structure content of the MD-ENS ensemble resembles that of the MD ensemble, Figure 1 shows that there is some variation in the percentages with which certain residues adopt different types of secondary structure.[14]

For analysis of Aβ40 and Aβ42 IDPs considered here, we have utilized a wide range of previously published NMR data including chemical shifts from the Zagorski group[60] as well as J-coupling constants, RDCs, and heteronuclear $^1H-^{15}N$ NOEs for backbone amides from Wang and co-workers.[57,61,62] Our group has collected ^1H chemical shifts and NOESY $^1H-^1H$ homonuclear spectra for the full length Aβ40 and Aβ42 peptides as reported elsewhere.[14,48] The data for the longer peptides were processed as described in ref 48 in a similar approach to that used for the Aβ21−30 fragment.[52]

First we consider the chemical shift data, for which we note that the calculated chemical shifts have an uncertainty that is independent of the quality or type of structural ensemble, and results from approximations of the SHIFTX[49] or SHIFTS[50] calculators themselves. Other research groups have reported the uncertainty, σ^2 (ppm), for these calculators, with the value depending on the atom type and its bonding chemistry.[49,50] Therefore the best way validate the various IDP ensembles with chemical shift data is to calculate the difference between the experimental chemical shift and the shift calculated from each
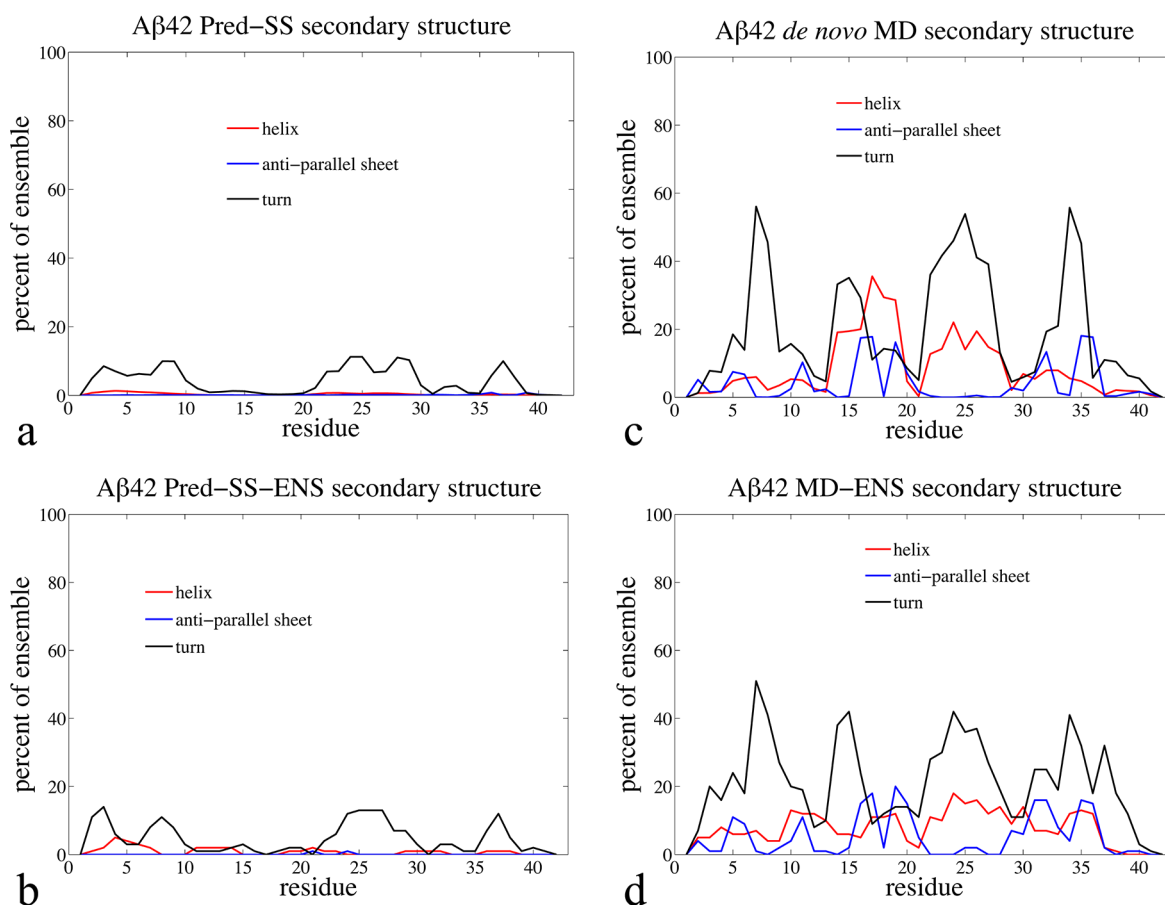
Figure 1. Percentage of Aβ42 simulated ensemble in different types of secondary structure by residue for (a) the Pred-SS, (b) Pred-SS-ENS, (c) *de novo* MD, and (d) MD-ENS ensembles. The red line represents helix, the blue line for antiparallel sheet, and the black line for β-turns. We note that the blue line represents only antiparallel sheet structure (the most common) and not all sheets.

of the structural ensembles, normalizing it by the calculator uncertainty, to generate χ_δ^2 values

$$\chi_\delta^2 = \frac{1}{N} \sum_{i=1}^{N} \frac{(\langle \delta_{i,calc} \rangle - \delta_{i,exp})^2}{\sigma_\delta^2} \qquad (9)$$

Reported uncertainties (root mean squared difference, RMSD, from experiment) for the SHIFTX calculator[49] are σ = 0.23 ppm for H_α, σ = 0.49 ppm for H_N, σ = 0.98 ppm for C_α, and σ = 1.10 ppm for $C\beta$. Any dominant error due to the underlying structural ensemble would then correspond to values of $\chi_\delta^2 > 1$.

Table 1 displays the χ_δ^2 agreement between experimentally measured proton[48] and carbon[60] chemical shifts with those generated from each candidate ensemble for both Aβ40 and Aβ42.[14,48,52] Experimental chemical shift data reported for the monomeric Aβ40 and Aβ42 peptides do not differ greatly from random coil values, and therefore the RC ensemble falls within χ_δ^2 uncertainty.[14,48] Since the Pred-SS ensemble shows almost no DSSP defined secondary structure (Figure 1a), it remains largely equivalent to the RC ensemble as deduced by chemical shifts. The *de novo* MD structural ensemble is also in good agreement with the chemical shift data;[14,48] however, ~99% of the MD generated Aβ conformations contain one or more elements of cooperative secondary structure somewhere along the peptide sequence (Figure 1b). The reason that the MD ensemble is also in good agreement with the experimental chemical shifts is that averaging over a large ensemble of cooperatively formed secondary structure and tertiary contacts

yields average chemical shifts that are consistent with random coil values.[14,48] For example, averaging the chemical shifts of all folded proteins in the PDB results in averages very similar to random coil values.[48,63] We have found that the ENSEMBLE optimization of the Pred-SS and MD starting pools improves the χ_δ^2 values, but all are within the calculator uncertainty. Not surprisingly, if the knowledge-based ENSEMBLE approach were biased by chemical shift data alone, they would show little deviation from their starting "soup", and the structural interpretation would be highly dependent on the starting ensemble. For this reason we conclude that NMR chemical shifts alone do not provide any qualitative discrimination between the alternative ensembles, at least not for the Aβ40 and Aβ42 disease IDPs.[14,48] It may still be useful to apply chemical shift constraints in combination with other experimental observables to optimize an IDP ensemble, as we have done when generating the MD-ENS ensemble. In this context the chemical shift constraints might provide a 'sanity check' against ensembles that fit other observables, such as NOEs, but lead to unphysical chemical shift values.

Similarly, J-couplings alone also do not discriminate between random coil IDPs and those that are more structured with cooperative secondary structure and tertiary structure contacts. Figure 2 illustrates this by plotting the agreement between experimentally measured $^3J(H_N,H_\alpha)$,[55,57] and those calculated from the RC, Pred-SS, Pred-SS-ENS, *de novo* MD, and MD-ENS ensembles for Aβ40 and Aβ4. Table 1 shows that all
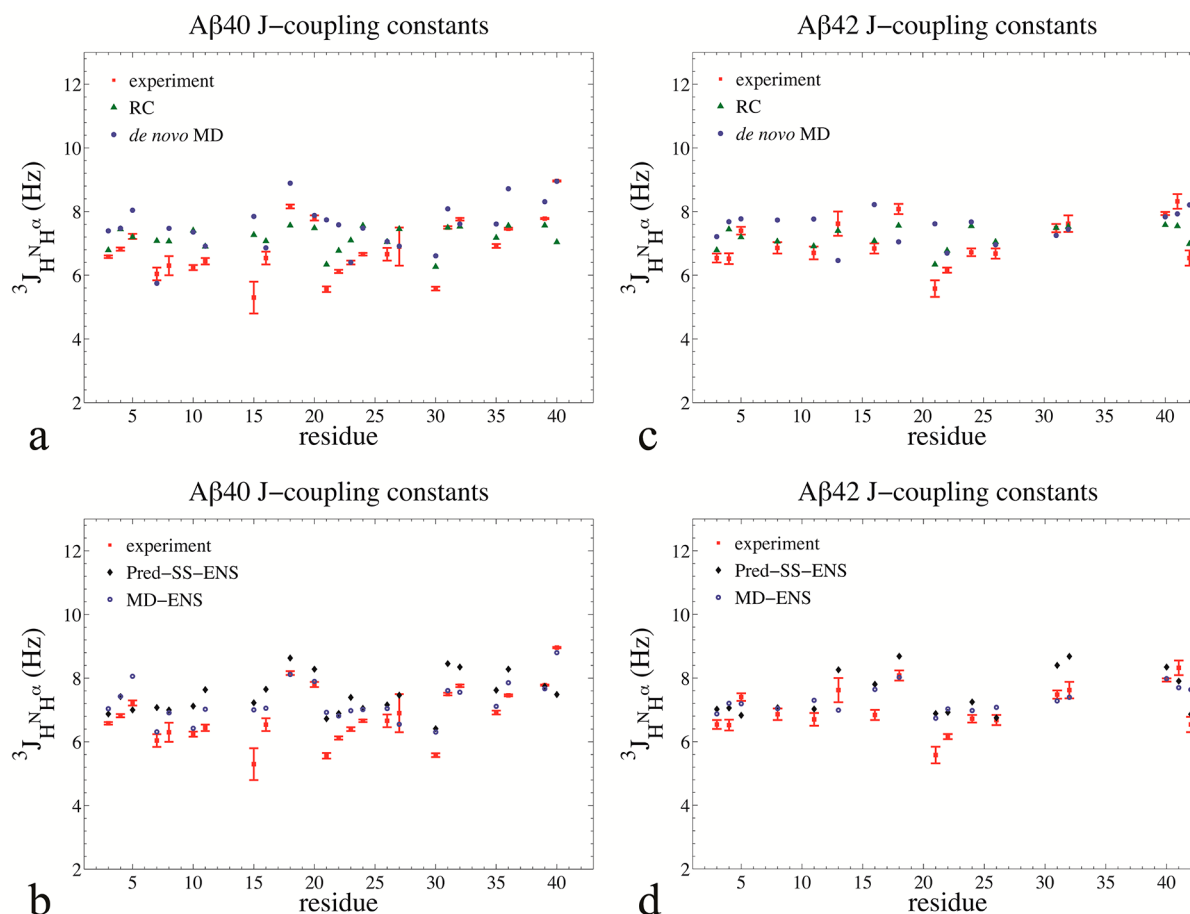
Figure 2. J-coupling constants for backbone amides for Aβ40 and Aβ42. (a) Aβ40 experimental J-coupling constants (red squares) compared to RC (green triangles) and *de novo* MD (solid blue circles). (b) Aβ40 experimental J-coupling constants (red squares) compared to Pred-SS-ENS (black diamonds) and MD-ENS (blue circles). (c) Aβ42 experimental, RC, and *de novo* MD J-coupling constants. (d) Aβ42 experimental, Pred-SS-ENS, and MD-ENS J-coupling constants. The experimental data are from Yan et al.[57] has been corrected to account for T_{1sel} relaxation and bring J-couplings determined from a $H_N H_\alpha$ 3D experiment to be consistent with those from COSY splittings.[60]

ensembles yield an RMSD across all residues of 0.60−1.09 Hz, and also reports

$$\chi_J^2 = \frac{1}{N} \sum_{i=1}^{N} \frac{(\langle J_{i,\text{calc}} \rangle - J_{i,\text{exp}})^2}{\sigma_J^2} \qquad (10)$$

with $\sigma_J = 0.73$ Hz. J-couplings report on the backbone ϕ dihedral angle, and therefore could in principle distinguish between an unstructured peptide and a peptide with a defined secondary structure; however, in the case of the disease IDP Aβ, the presence of diverse secondary structure in the MD ensemble is not apparent from the calculated J-couplings. We believe that this stems from the fact that good agreement with scalar coupling data for IDPs can largely be predicted by sampling over the allowed regions of residue-specific Ramachandran plots without needing to assume any structure adopted by the full length sequence. Thus J-couplings also do not provide an experimental measure for discriminating among qualitatively different structural ensembles for the amyloid peptides.

Table 2 provides the assessment of the five alternative ensembles for Aβ40 and Aβ42 using RDC values evaluated residue by residue using the PALES program[58] and L-RDCs based on local alignments.[18] While the RC and Pred-SS ensembles yield lower RMSD values, 1.3−1.5 Hz, they are

marginally better than the *de novo* MD RMSD of 2.2 Hz.[14,48] This is in part due to the fact that experimental RDC uncertainties for IDPs are larger (∼0.9 Hz for Aβ40 and ∼0.5 Hz for Aβ42) than the uncertainty observed for folded proteins of ∼0.1 Hz.[64] In addition, there are large uncertainties in the accuracy of RDC calculators using programs such as PALES.[58] In fact, the reported RMSD of the PALES calculator for folded proteins is ∼2.0 Hz, on the same order as the RMSD for the *de novo* MD ensemble. While the ENSEMBLE method does significantly lower the RMSD for L-RDCs for the Pred-SS-ENS and MD-ENS ensembles, the corresponding RMSD based on the global alignment using PALES is marginally better than the Pred-SS and *de novo* MD starting pools.[14,48] Hence for this particular application on disordered amyloid peptides, we have found that RDCs are not a particularly good experimental metric for differentiating among the different ensembles, and substantial disagreement between RDCs based on local and global alignments are observed.

Finally, we consider the performance of the different ensemble methods for reproducing ^{1}H−^{1}H homonuclear NOE cross-peaks. We have presented the NOE data collection for the Aβ42 peptide in which ∼700 cross-peaks are observed in the NOE spectra, but only ∼200 can be uniquely assigned from experimental information alone.[14,48] The remaining cross-peaks do not have a clear independent assignment (and in fact

Table 2. Comparison between Random Coil (RC), Predicted Secondary Structure (Pred-SS), *de novo* MD (MD), and ENSEMBLE Optimized Pred-SS-ENS, and MD-ENS Ensembles[a]

$A\beta40$ peptide	average property			
ensemble type	RDC-PALES (Hz)	RDC-Local (Hz)	H_2O NOEs	D_2O NOEs
RC	1.49	1.56	11.75 (0.47)	4.61 (0.54)
Pred-SS	1.54	1.36	4.68 (0.50)	3.75 (0.54)
Pres-SS-ENS	1.85	0.48	1.85 (0.68)	3.54 (0.52)
MD	2.22	1.88	1.15 (0.74)	3.22 (0.55)
MD-ENS	1.69	0.18	1.22 (0.70)	3.66 (0.51)
$A\beta42$ peptide	average property			
ensemble type	RDC-PALES (Hz)	RDC-Local (Hz)	H_2O NOEs	D_2O NOEs
RC	1.35	1.42	7.25 (0.38)	3.77 (0.44)
Pred-SS	1.37	1.27	3.28 (0.52)	1.26 (0.67)
Pres-SS-ENS	1.17	0.42	1.81 (0.59)	0.75 (0.76)
MD	2.25	2.14	1.25 (0.67)	0.58 (0.80)
MD-ENS	2.13	0.33	1.51 (0.62)	0.73 (0.76)

[a]We report RMSDs for the RDC calculator PALES and L-RDCs evaluated with ENSEMBLE using local alignments. The NOEs are back-calculated from the structural ensembles as described in Section 4. We evaluate the RMSD normalized by the largest NOE intensity, RMSDN and (correlation coefficient, r) with the H_2O and D_2O experiments. Some data reproduced from ref 14.

require a computational model to interpret them[14,48]). Therefore we have only compared the different methods against the NOE cross-peaks that can be assigned by experiment alone. We note that quantitatively reproducing NOE intensities is a very high bar since peak volumes are extremely sensitive to r^{-6} distance averaging, that also involve an appropriate time scale that is heterogeneous across proton pairs. Geometric imperfections in the conformational ensemble where contact differences differ by a factor of $2^{1/6}$ (difference between 1 Å and 1.12 Å) will double the corresponding intensity value, thereby driving up the RMSD error for all ensembles.[14,48,52] Large absolute NOE intensities especially tend to dominate the RMSD error, and therefore we have mitigated this effect by normalizing the RMSD (RMSDN) by the experimental intensity for each NOE as in ref 14.

Table 2 shows that the predicted set of $^1H-^1H$ NOEs from *de novo* MD is better than any other ensemble, with RMSDNs that are lower than the RC and Pred-SS ensembles values by 2−3 fold and with much higher correlation coefficients.[14] The Pred-SS-ENS ensemble performs better for $A\beta40$ and $A\beta42$ than the randomly generated ensembles because the NOE restraints are used in the knowledge-based ensemble selection. However, the Pred-SS-ENS ensemble still does not reproduce the data as well as the *de novo* MD ensemble. The NOE validation clearly indicates that the *de novo* MD ensemble with its cooperative secondary structure is a better representation of $A\beta40$ and $A\beta42$ than are the RC, Pred-SS, or Pred-SS-ENS ensembles, which have no cooperative secondary structure. Since time information is not available for the static ensembles, we can only evaluate the NOEs for the statistical ensembles under the assumption of one uniform correlation time applied to all pairs of protons, for which we use a 1 ns correlation time,

which is on the same order as those observed in the MD simulations. Of course the *de novo* MD method can account for the time scales explicitly and more importantly for the fact that different pairs of hydrogen atoms do decay on different time scales. Thus the statistically generated and knowledge-based ensembles agree relatively poorly with the NOE observables since the heterogeneity in correlation times are unknown, and hence even the MD-ENS ensemble is in somewhat worse agreement with the experimental NOEs than the *de novo* MD ensemble (Table 2). The NOE validation emphasizes that an IDP's diverse set of conformations gives rise to a heterogeneous set of correlation times that must be described in order to validate against experimental NOEs.

We further emphasize that the calculation of heteronuclear NOEs, being a purely dynamical measurement, is only possible with the *de novo* MD method. Figure 3 shows a comparison of the experimental $^1H-^{15}N$ NOE intensities, measured by Yan and Wang,[62] and those derived from our MD simulation for $A\beta42$ and $A\beta40$, showing overall excellent agreement.[14,48] Unlike the $^1H-^1H$ NOEs, these assignments are unambiguous from experiment. We find that, as in the experiment, there is an increase in $^1H-^{15}N$ NOE intensities calculated from simulation for residues 35−40 for $A\beta42$ compared to $A\beta40$, indicating that the longer peptide experiences slower dynamics at the C-terminus.[14,48] This difference in experimental $^1H-^{15}N$ NOEs for $A\beta42$ and $A\beta40$ has previously been interpreted as evidence that $A\beta42$ has greater structural rigidity in the C-terminus compared to $A\beta40$[62,65] and we provide more analysis on this point in recently published work.[7]

■ DISCUSSION

We have shown that the MD and MD-ENS structural ensembles for the IDPs $A\beta40$ and $A\beta42$ previously characterized[14,48] yield substantially better agreement with a range of NMR data than the random coil or statistical ensembles that are typically used with knowledge based approaches. The MD ensembles are qualitatively different from random coil or statistical ensembles in that the subpopulations are richly structured, contain a diverse set of secondary structures including α-helix, β-turns, and β-strands, and span the full range of compact to fully extended conformations. Furthermore, while MD generated ensembles are Boltzmann weighted, the knowledge-based approaches give equal statistical weight to all conformations and thus are likely inconsistent with statistical mechanical weightings that are inherent to the NMR experiment.

We have also shown that some types of NMR data may not be helpful for discriminating among qualitatively different structural ensemble of IDPs. In particular, averages over a diverse set of cooperative secondary structure conformations yield experimental values of chemical shifts that are superficially consistent with values expected from a random coil ensemble. Furthermore, if the chemical shifts are not highly dispersed along the sequence of a particular IDP, such as is found for the amyloid-β peptides, then the chemical shifts have limited value as experimental refinement input or as a validation measure. J-couplings also do not provide discrimination between randomly generated conformations and a diverse population of cooperative secondary structure. In fact, we found that scalar couplings calculated as averages over the allowed regions of each residue-specific Ramachandran plot gave as good agreement with the experimental J-couplings for $A\beta40$ and $A\beta42$ as did averages over the structural populations.
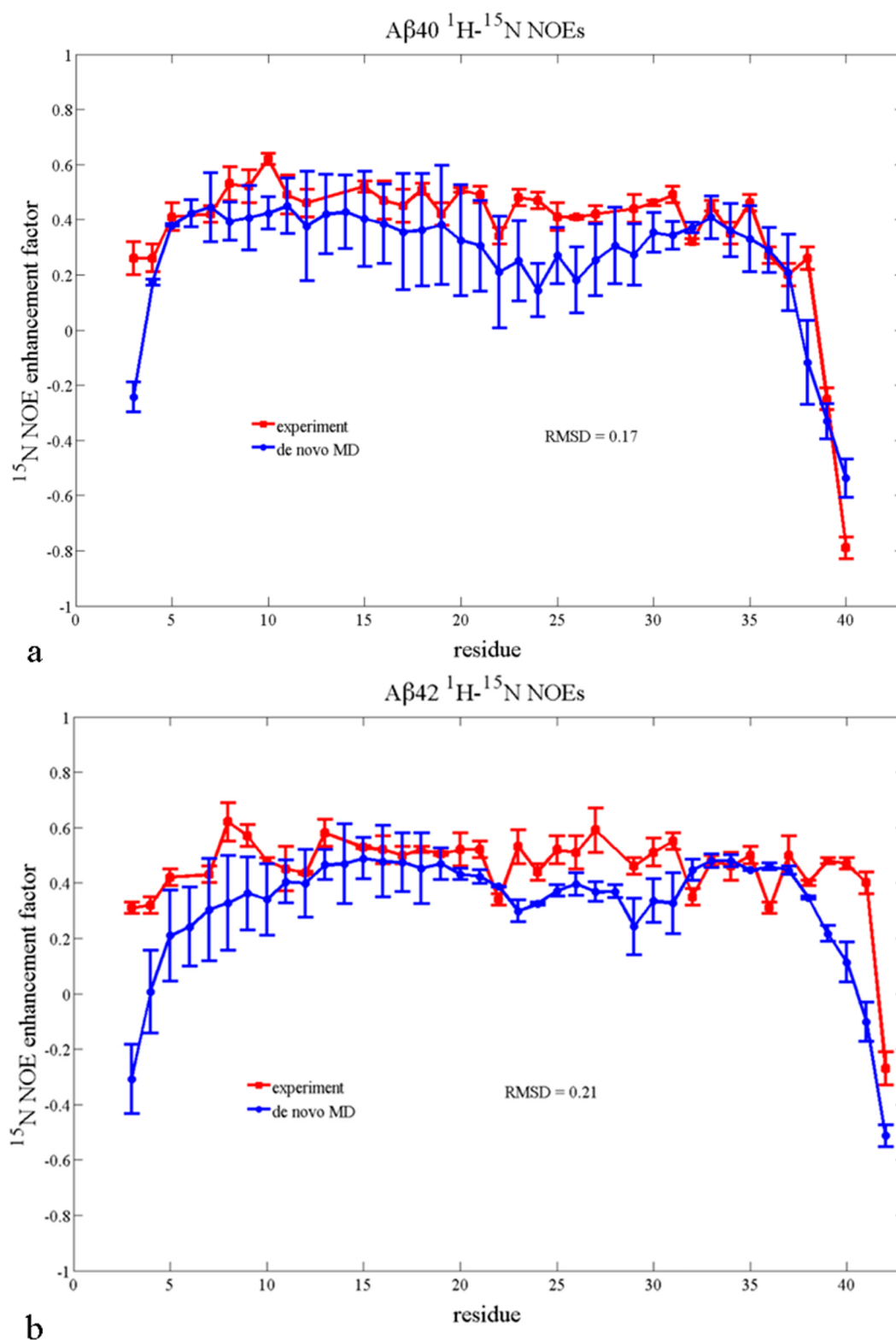
Figure 3. Agreement with experiment of simulated (a) Aβ40 and (b) Aβ42 ^{1}H−^{15}N NOE. The red squares are experimental data from Yan and Wang.[66] The blue circles are the data calculated from the *de novo* MD ensemble.

Unlike others who have used RDC data to help interpret IDP or unfolded protein structural ensembles, we found RDCs to be only marginally useful for Aβ40 and Aβ42. This may be due to limitations of RDC calculators such as PALES,[58] which were originally developed and successfully applied to folded proteins, but which are reported to have large uncertainties in their predicted RDC values. Furthermore, calculated RDCs based on global alignment algorithms such as PALES[58] diverge significantly from RDCs evaluated from localized alignments[18] for Aβ40 and Aβ42, indicating that in cases like this the ENSEMBLE package should be employed using the PALES calculator to fit RDCs, which is possible though not standard.[66]

More research may be necessary to apply programs like PALES to disordered proteins, which likely do not align in an anisotropic medium in the same way as folded proteins, in part due to the time scale of interconversion of the conformational substates. For example, conventional methods for calculating RDCs cannot be applied to the motion of multidomain biomolecules,[67] and the local conformational sampling and long-range structure need to be simultaneously accounted for because they both affect the experimental RDC data.[16] However, progress is being made in using RDCs to provide meaningful structural information for other IDPs.[16,17,40,45,67,68] We speculate that success is greatest when all subpopulations of the IDP ensemble are homogeneously classifiable (as extended disordered conformations for example), so that the IDP global alignment properties are uniform and resulting averages provide meaningful and consistent structural information.

We have demonstrated that homonuclear ^{1}H$-^1$H NOE intensities and heteronuclear ^{15}N$-^1$H NOEs are by themselves discriminating with regard to the tertiary contacts and backbone dynamics, respectively, that define the important validation of the MD-based ensembles over the statistical coil ensembles.[14,48] Furthermore, a correct picture of the IDP ensemble based on the experimental NOE data would not be possible without a computational model providing both details of individual structures and the time scales for their interconversion. In turn, although the homonuclear NOEs are averaged over all subpopulations, they are still vital for deducing whether a given ensemble contains subpopulations of structure with the right tertiary contacts to give rise to the observed cross-peaks in the spectra. Because these cross-peak intensities rely directly on the decay time scales of correlated proton distances, the NOEs for IDPs are reporting on a heterogeneous population of time scales. One of the primary limitations of the statically generated ensembles is that they are not associated with any information about motional time scales that can be used to calculate NOE observables. Relaxation times can be used with the ENSEMBLE method, although they are incorporated as structural rather than dynamic constraints.[29,30,42] This dynamic information is a genuine strength of the *de novo* MD methods, especially for ^{15}N$-^1$H NOEs, which cannot be calculated from the static ensembles.

We believe that the primary limitation of knowledge-based methods applied to the difficult amyloidβ case is 3-fold. We note that while there will be quantitative differences between ENSEMBLE and other knowledge-based approaches such as ASTEROIDS, qualitatively the problems will be similar. First, there is no requirement for generating a complete and representative starting "basis set" of conformations to select the final ensemble from; i.e., these methods cannot use the NMR data effectively to select for compact structures with elements of cooperative secondary structure if the initial pool of structures is largely composed of extended random coil structures. Both ENSEMBLE and ASTEROIDS have relied on statistical coil ensembles as the starting pool of structures, and while some "on-the-fly" addition of new structures is possible with these methods, they do not yet support formation of complicated β-sheet motifs.[16,42] Metrics of ensemble heterogeneity, such as those developed by the Onuchic and Stultz research groups, will continue to be useful as we explore the range of IDPs that cannot be easily classified based on their level of disorder.[69−71] Second, for certain classes of IDPs such as amyloidβ optimization of structures to reproduce chemical

shifts and scalar couplings does not discriminate among qualitatively different structural ensembles. Third, the optimization phase of the knowledge-based approaches relies on approximations to NMR observables, which may diverge from a global property, as for L-RDCs, or from the dynamical origins of NOE intensities.

At the same time, the *de novo* MD method is not quantitatively perfect, and therefore the MD ensemble provides an excellent start state for subsequent refinement by knowledge-based approaches. An unambiguous future direction for the structural biology of IDPs is the combined use of knowledge-based approaches and MD that supplies Boltzmann weighted conformational substates as well as heterogeneous time scales of motion.

All together, the productive interplay between NMR experiments, *de novo* MD simulations, and knowledge-based approaches, along with supporting models, algorithms, and computer hardware, gives us an ability to accurately identify structures present in IDP ensembles and use that knowledge to gain previously inaccessible functional insights.[14,22,24,39,43,44,48,52,55,72−83] To further improve techniques for studying disordered proteins, we as a community could establish a high throughput computational infrastructure to predict IDP structural ensembles using a combination of MD and NMR. This would be similar to the establishment of X-ray crystallography beamlines for the rapid solution of folded protein structures that was launched during the structural genomics era. The ultimate goal in both cases is to use structural information to drive the formation of hypotheses about protein function. Based on the success of using structural information for functional characterization of folded proteins and complexes, we hope and expect that structural knowledge of IDP ensembles can provide similar insight into IDP function and enable development of molecular hypotheses for disease IDPs.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

A detailed description of how the five candidate .amyloid-β structural ensembles were generated, details on the use of the ENSEMBLE software, a table showing how the Pred-SS ensemble was biased to contain extended dihedral angles preferentially, and a figure showing the secondary structure present in the RC ensemble. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Notes
The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Kendrew, J. C.; Bodo, G.; Dintzis, H. M.; Parrish, R. G.; Wyckoff, H.; Phillips, D. C. A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-ray Analysis. *Nature* **1958**, *181*, 662−666.

(2) Henderson, R.; Baldwin, J. M.; Ceska, T. A.; Zemlin, F.; Beckmann, E.; Downing, K. H. Model for the Structure of Bacteriorhodopsin based on High-Resolution Electron Cryo-microscopy. *J. Mol. Biol.* **1990**, *213*, 899−929.

(3) Lange, O. F.; Schäfer, L. V.; Grubmüller, H. Flooding in GROMACS: Accelerated Barrier Crossings in Molecular Dynamics. *J. Comput. Chem.* **2006**, *27*, 1693−1702.

(4) Wright, P. E.; Dyson, H. J. Linking Folding and Binding. *Curr. Opin. Struct. Biol.* **2009**, *19* (1), 31−38.

(5) Dunker, A. K.; Gough, J. Sequences and Topology: Intrinsic Disorder in the Evolving Universe of Protein Structure. *Curr. Opin. Struct. Biol.* **2011**, *21* (3), 379−381.

(6) Wright, P. E.; Dyson, H. J. Intrinsically Unstructured Proteins: Re-assessing the Protein Structure-function Paradigm. *J. Mol. Biol.* **1999**, *293* (2), 321−331.

(7) Tompa, P. Intrinsically Unstructured Proteins. *Trends Biochem. Sci.* **2002**, *27* (10), 527−533.

(8) Dunker, A. K.; Brown, C. J.; Lawson, J. D. Intrinsic Disorder and Protein Function. *Biochemistry* **2002**, *41*, 6573−6582.

(9) Iakoucheva, L. M.; Brown, C. J.; Lawson, J. D.; Obradović, Z.; Dunker, A. K. Intrinsic Disorder in Cell-signaling and Cancer-associated Proteins. *J. Mol. Biol.* **2002**, *323* (3), 573−584.

(10) Uversky, V. N.; Oldfield, C. J.; Dunker, A. K. Showing your ID: Intrinsic Disorder as an ID for Recognition, Regulation and Cell Signaling. *J. Mol. Recognition* **2005**, *18* (5), 343−384.

(11) Dyson, H. J.; Wright, P. E. Intrinsically Unstructured Proteins and their Functions. *Nat. Rev.: Mol. Cell Biol.* **2005**, *6* (3), 197−208.

(12) Xie, H.; Vucetic, S.; Iakoucheva, L. M.; Oldfield, C. J.; Dunker, A. K.; Uversky, V. N.; Obradović, Z. Functional Anthology of Intrinsic Disorder. 1. Biological Processes and Functions of Proteins with Long Disordered Regions. *J. Proteome Res.* **2007**, *6* (5), 1882−1898.

(13) Oldfield, C. J.; Meng, J.; Yang, J. Y.; Yang, M. Q.; Uversky, V. N.; Dunker, A. K. Flexible nets: Disorder and Induced Fit in the Associations of p53 and 14−3−3 with their Partners. *BMC Genomics* **2008**, *S1*.

(14) Ball, K. A.; Phillips, A. H.; Wemmer, D. E.; Head-Gordon, T. Differences in β-strand Populations of Monomeric Amyloid-β 40 and Amyloid-β 42. *Biophys. J.* **2013**, *104* (12), 2714−2724.

(15) Mittag, T.; Orlicky, S.; Choy, W.-Y.; Tang, X.; Lin, H.; Sicheri, F. Dynamic Equilibrium Engagement of a Polyvalent Ligand with a Single-site Receptor. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (46), 17772−17777.

(16) Schneider, R.; Huang, J.-R.; Yao, M.; Communie, G.; Ozenne, V.; Mollica, L.; Salmon, L.; Jensen, M. R.; Blackledge, M. Towards a Robust Description of Intrinsic Protein Disorder using Nuclear Magnetic Resonance Spectroscopy. *Mol. BioSys.* **2012**, *8* (1), 56−68.

(17) Esteban-Martín, S.; Fenwick, R. B.; Salvatella, X. Refinement of Ensembles Describing Unstructured Proteins Using NMR Residual Dipolar Couplings. *J. Am. Chem. Soc.* **2010**, *132* (13), 4626−4632.

(18) Marsh, J. A.; Baker, J. M. R.; Tollinger, M.; Forman-Kay, J. D. Calculation of Residual Dipolar Couplings from Disordered State Ensembles using Local Alignment. *J. Am. Chem. Soc.* **2008**, *130* (25), 7804−7805.

(19) Montalvao, R. W.; Simone, A. D.; Vendruscolo, M. Determination of Structural Fluctuations of Proteins from Structure-based Calculations of Residual Dipolar Couplings. *J. Biomol. NMR* **2012**, *53*, 281.

(20) Showalter, S. A.; Brüschweiler, R. Quantitative Molecular Ensemble Interpretation of NMR Dipolar Couplings without Restraints. *J. Am. Chem. Soc.* **2007**, *129* (14), 4158−4159.

(21) Felitsky, D. J.; Lietzow, M. A.; Dyson, H. J.; Wright, P. E. Modeling Transient Collapsed States of an Unfolded Protein to Provide Insights into Early Folding Events. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (17), 6278−6283.

(22) Vise, P.; Baral, B.; Stancik, A.; Lowry, D. F.; Daughdrill, G. W. Identifying Long-range Structure in the Intrinsically Unstructured Transactivation Domain of p53. *Proteins: Struct., Funct., Bioinf.* **2007**, *67* (3), 526−530.

(23) Marsh, J. A.; Neale, C.; Jack, F. E.; Choy, W.-Y.; Lee, A. Y.; Crowhurst, K. A.; Forman-Kay, J. D. Improved Structural Characterizations of the drkN SH3 Domain Unfolded State suggest a Compact Ensemble with Native-like and Non-native Structure. *J. Mol. Biol.* **2007**, *367* (5), 1494−1510.

(24) Allison, J. R.; Várnai, P.; Dobson, C. M.; Vendruscolo, M. Determination of the Free Energy Landscape of Alpha-Synuclein Using Spin Label Nuclear Magnetic Resonance Measurements. *J. Am. Chem. Soc.* **2009**, *131* (51), 18314−18326.

(25) Ganguly, D.; Chen, J. Structural Interpretation of Paramagnetic Relaxation Enhancement-derived Distances for Disordered Protein States. *J. Mol. Biol.* **2009**, *390* (3), 467−477.

(26) Chen, J. Towards the Physical Basis of how Intrinsic Disorder Mediates Protein Function. *Arch. Biochem. Biophys.* **2012**, *524* (2), 123−131.

(27) Vendruscolo, M. Determination of Conformationally Heterogeneous States of Proteins. *Curr. Opin. Struct. Biol.* **2007**, *17* (1), 15−20.

(28) Lawrence, C. W.; Bonny, A.; Showalter, S. A. The Disordered C-terminus of the RNA Polymerase II Phosphatase FCP1 is Partially Helical in the Unbound State. *Biochem. Biophys. Res. Commun.* **2011**, *410* (3), 461−465.

(29) Marsh, J. A.; Forman-Kay, J. D. Structure and Disorder in an Unfolded State under Nondenaturing Conditions from Ensemble Models Consistent with a Large Number of Experimental Restraints. *J. Mol. Biol.* **2009**, *391* (2), 359−374.

(30) Marsh, J. A.; Forman-Kay, J. D. Ensemble Modeling of Protein Disordered States: Experimental Restraint Contributions and Validation. *Proteins: Struct., Func., Bioinf.* **2011**, DOI: 10.1002/prot.23220.

(31) Marsh, J. A.; Dancheck, B.; Ragusa, M. J.; Allaire, M.; Forman-Kay, J. D.; Peti, W. Structural Diversity in Free and Bound States of Intrinsically Disordered Protein Phosphatase 1 Regulators. *Structure* **2010**, *18* (9), 1094−1103.

(32) Prompers, J. J.; Brüschweiler, R. General Framework for Studying the Dynamics of Folded and Nonfolded Proteins by NMR Relaxation Spectroscopy and MD Simulation. *J. Am. Chem. Soc.* **2002**, *124* (16), 4522−4534.

(33) Zandarashvili, L.; Li, D. W.; Wang, T.; Brüschweiler, R.; Iwahara, J. Signature of Mobile Hydrogen Bonding of Lysine Side Chains from Long-Range ^{15}N-^{13}C Scalar J-Couplings and Computation. *J. Am. Chem. Soc.* **2011**, *133*, 9192−9195.

(34) Case, D. A. Molecular Dynamics and NMR Spin Relaxation in Proteins. *Acc. Chem. Res.* **2002**, *35* (6), 325−331.

(35) Herrmann, T.; Güntert, P.; Wüthrich, K. Protein NMR Structure Determination with Automated NOE Assignment using the new Software CANDID and the Torsion Angle Dynamics Algorithm DYANA. *J. Mol. Biol.* **2002**, *319*, 209−227.

(36) López-Méndez, B.; Güntert, P. Automated Protein Structure Determination from NMR Spectra. *J. Am. Chem. Soc.* **2006**, *128*, 13112−13222.

(37) Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Clore, G. M. The Xplor-NIH NMR Molecular Structure Determination Package. *J. Magn. Reson.* **2003**, *160*, 66−74.

(38) Schwieters, C. D.; Kuszewski, J. J.; Clore, G. M. Using Xplor-NIH for NMR Molecular Structure Determination. *Prog. Nucl. Magn. Reson. Spectrosc.* **2006**, *48*, 47–62.

(39) Dedmon, M. M.; Lindorff-Larsen, K.; Christodoulou, J.; Vendruscolo, M.; Dobson, C. M. Mapping long-range Interactions in Alpha-synuclein using Spin-label NMR and Ensemble Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **2005**, *127* (2), 476–477.

(40) Richter, B.; Gsponer, J.; Várnai, P.; Salvatella, X.; Vendruscolo, M. The MUMO (Minimal Under-restraining Minimal Over-restraining) Method for the Determination of Native State Ensembles of Proteins. *J. Biomol. NMR* **2007**, *37* (2), 117–135.

(41) Lindorff-Larsen, K.; Kristjansdottir, S.; Teilum, K.; Fieber, W.; Dobson, C. M.; Poulsen, F. M.; Vendruscolo, M. Determination of an Ensemble of Structures Representing the Denatured State of the Bovine Acyl-coenzyme a Binding Protein. *J. Am. Chem. Soc.* **2004**, *126* (10), 3291–3299.

(42) Krzeminski, M.; Marsh, J. A.; Neale, C.; Choy, W.-Y.; Forman-Kay, J. D. Characterization of Disordered Proteins with ENSEMBLE. *Bioinformatics* **2013**, *29* (3), 398–399.

(43) Huang, A.; Stultz, C. M. The effect of a DeltaK280 Mutation on the Unfolded State of a Microtubule-binding Repeat in Tau. *PLoS Comput. Biol.* **2008**, *4* (8), e1000155.

(44) Yoon, M.; Venkatachalam, V.; Huang, A.; Choi, B.; Stultz, C.; Chou, J. Residual Structure within the Disordered C-terminal Segment of p21(Waf1/Cip1/Sdi1) and its Implications for Molecular Recognition. *Protein Sci.* **2009**, *18*, 337–347.

(45) Salmon, L.; Nodet, G.; Ozenne, V.; Yin, G.; Jensen, M. R.; Zweckstetter, M.; Blackledge, M. NMR Characterization of Long-range Order in Intrinsically Disordered Proteins. *J. Am. Chem. Soc.* **2010**, *132* (24), 8407–8418.

(46) Jensen, M. R.; Salmon, L.; Nodet, G.; Blackledge, M. Defining Conformational Ensembles of Intrinsically Disordered and Partially Folded Proteins Directly from Chemical Shifts. *J. Am. Chem. Soc.* **2010**, *132* (4), 1270.

(47) Choy, W. Y.; Forman-Kay, J. D. Calculation of Ensembles of Structures Representing the Unfolded State of an SH3 Domain. *J. Mol. Biol.* **2001**, *308* (5), 1011–1032.

(48) Ball, K. A.; Phillips, A. H.; Nerenberg, P. S.; Fawzi, N. L.; Wemmer, D. E.; Head-Gordon, T. L. Homogeneous and Heterogeneous Tertiary Structure Ensembles of Amyloid-β Peptides. *Biochemistry* **2011**, *50* (35), 7612–7628.

(49) Neal, S.; Nip, A. M.; Zhang, H.; Wishart, D. S. Rapid and Accurate Calculation of Protein 1H, 13C and 15N Chemical Shifts. *J. Biomol. NMR* **2003**, *26* (3), 215–40.

(50) Xu, X. P.; Case, D. A. Automated Prediction of ^{15}N, ^{13}C$^{\alpha}$, ^{13}C$^{\beta}$, and ^{13}C$'$ Chemical Shifts in Proteins using a Density Functional Database. *J. Biomol. NMR* **2001**, *21* (4), 321–33.

(51) Sitkoff, D.; Case, D. A. Density Functional Calculations of Proton Chemical Shifts in Model Peptides. *J. Am. Chem. Soc.* **1997**, *119* (50), 12262–12273.

(52) Fawzi, N. L.; Phillips, A. H.; Ruscio, J. Z.; Doucleff, M.; Wemmer, D. E.; Head-Gordon, T. Structure and dynamics of the Aβ(21–30) Peptide from the Interplay of NMR Experiments and Molecular Simulations. *J. Am. Chem. Soc.* **2008**, *130* (19), 6145–58.

(53) Karplus, M.; Grant, D. M. A Criterion For Orbital Hybridization and Charge Distribution in Chemical Bonds. *Proc. Natl. Acad. Sci. U. S. A.* **1959**, *45* (8), 1269–1273.

(54) Vuister, G. W.; Bax, A. Quantitative *J* Correlation: A New Approach For Measuring Homonuclear Three-Bond $J(H^N H^\alpha)$ Coupling-Constants In ^{15}N-Enriched Proteins. *J. Am. Chem. Soc.* **1993**, *115* (17), 7772–7777.

(55) Sgourakis, N. G.; Merced-Serrano, M.; Boutsidis, C.; Drineas, P.; Du, Z.; Wang, C.; Garcia, A. Atomic-Level Characterization of the Ensemble of the Aβ(1–42) Monomer in Water Using Unbiased Molecular Dynamics Simulations and Spectral Algorithms. *J. Mol. Biol.* **2011**, *405* (2), 570–583.

(56) Sgourakis, N. G.; Yan, Y.; McCallum, S. A.; Wang, C.; Garcia, A. E. The Alzheimer's Peptides Aβ40 and 42 adopt Distinct

Conformations in Water: A Combined MD/NMR Study. *J. Mol. Biol.* **2007**, *368* (5), 1448–57.

(57) Yan, Y.; McCallum, S. A.; Wang, C. M35 Oxidation Induces Aβ40-like Structural and Dynamical Changes in Aβ42. *J. Am. Chem. Soc.* **2008**, *130* (16), 5394–5.

(58) Zweckstetter, M.; Bax, A. Prediction of Sterically induced Alignment in a Dilute Liquid Crystalline Phase: Aid to Protein Structure Determination by NMR. *J. Am. Chem. Soc.* **2000**, *122* (15), 3791–3792.

(59) Nodet, G.; Salmon, L.; Ozenne, V.; Meier, S.; Jensen, M. R.; Blackledge, M. Quantitative Description of Backbone Conformational Sampling of Unfolded Proteins at Amino Acid Resolution from NMR Residual Dipolar Couplings. *J. Am. Chem. Soc.* **2009**, *131* (49), 17908–17918.

(60) Hou, L.; Shao, H.; Zhang, Y.; Li, H.; Menon, N. K.; Neuhaus, E. B.; Brewer, J. M.; Byeon, I. J.; Ray, D. G.; Vitek, M. P.; Iwashita, T.; Makula, R. A.; Przybyla, A. B.; Zagorski, M. G. Solution NMR Studies of the Aβ(1–40) and Aβ(1–42) Peptides Establish that the Met35 Oxidation State Affects the Mechanism of Amyloid Formation. *J. Am. Chem. Soc.* **2004**, *126* (7), 1992–2005.

(61) Yan, Y.; Liu, J.; McCallum, S.; Yang, D.; Wang, C. Methyl Dynamics of the Amyloid-beta Peptides Aβ40 and Aβ42. *Biochem. Biophys. Res. Commun.* **2007**, *362* (2), 410–414.

(62) Yan, Y.; Wang, C. Aβ42 is more Rigid than Aβ40 at the C terminus: Implications for Aβ Aggregation and Toxicity. *J. Mol. Biol.* **2006**, *364* (5), 853–62.

(63) Ulrich, E. L.; Akutsu, H.; Doreleijers, J. F.; Harano, Y.; Ioannidis, Y. E.; Lin, J.; Livny, M.; Mading, S.; Maziuk, D.; Miller, Z.; Nakatani, E.; Schulte, C. F.; Tolmie, D. E.; Wenger, R. K.; Yao, H.; Markley, J. L. BioMagResBank. *Nuc. Acid. Res.* **2008**, *36*, D402–D408.

(64) Clore, G. M.; Gronenborn, A. M.; Tjandra, N. Direct Structure Refinement Against Residual Dipolar Couplings in the Presence of Rhombicity of Unknown Magnitude. *J. Magn. Reson.* **1998**, *131*, 1590162.

(65) Lim, K.; Henderson, G.; Jha, A.; Louhivuori, M. Structural, Dynamic Properties of Key Residues in Abeta Amyloidogenesis: Implications of an Important Role of Nanosecond Timescale Dynamics. *ChemBioChem* **2007**, *8* (11), 1251–1254.

(66) Pinheiro, A. S.; Marsh, J. A.; Forman-Kay, J. D.; Peti, W. Structural Signature of the MYPT1-PP1 Interaction. *J. Am. Chem. Soc.* **2011**, *133* (1), 73–80.

(67) Fenwick, R. B.; Esteban-Martín, S.; Salvatella, X. Understanding Biomolecular Motion, Recognition, and Allostery by use of Conformational Ensembles. *Euro. Biophys. J.* **2011**, *40* (12), 1339–1355.

(68) Valafar, H.; Prestegard, J. H. REDCAT: A Residual Dipolar Coupling Analysis Tool. *J. Magn. Reson.* **2004**, *167* (2), 228–241.

(69) Plotkin, S. S.; Onuchic, J. N. Investigation of Routes and Funnels in Protein Folding by Free Energy Functional Methods. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97* (12), 6509–6514.

(70) Fisher, C. K.; Stultz, C. M. Protein Structure along the Order–Disorder Continuum. *J. Am. Chem. Soc.* **2011**, *133* (26), 10022–10025.

(71) Chavez, L. L.; Onuchic, J. N.; Clementi, C. Quantifying the Roughness on the Free Energy Landscape: Entropic Bottlenecks and Protein Folding Rates. *J. Am. Chem. Soc.* **2004**, *126* (27), 8426–8432.

(72) Sivakolundu, S. G.; Bashford, D.; Kriwacki, R. W. Disordered p27Kip1 Exhibits Intrinsic Structure Resembling the Cdk2/cyclin A-bound Conformation. *J. Mol. Biol.* **2005**, *353* (5), 1118–1128.

(73) Chen, Y.-X.; Du, J.-T.; Zhou, L.-X.; Liu, X.-H.; Zhao, Y.-F.; Nakanishi, H.; Li, Y.-M. Alternative O-GlcNAcylation/O-phosphorylation of Ser16 induce Different Conformational Disturbances to the N Terminus of Murine Estrogen Receptor Beta. *Chem. Biol.* **2006**, *13* (9), 937–944.

(74) Ma, K.; Forbes, J. G.; Gutierrez-Cruz, G.; Wang, K. Titin as a Giant Scaffold for Integrating Stress and Src Homology Domain 3-mediated Signaling Pathways: the Clustering of novel Overlap Ligand Motifs in the Elastic PEVK Segment. *J. Biol. Chem.* **2006**, *281* (37), 27539–27556.

(75) Wu, K.-P.; Weinstock, D. S.; Narayanan, C.; Levy, R. M.; Baum, J. Structural Reorganization of Alpha-synuclein at Low pH Observed by NMR and REMD Simulations. *J. Mol. Biol.* **2009**, *391* (4), 784−796.

(76) Terakawa, T.; Takada, S. Multiscale Ensemble Modeling of Intrinsically Disordered Proteins: p53 N-terminal Domain. *Biophys. J.* **2011**, *101* (6), 1450−1458.

(77) Higo, J.; Nishimura, Y.; Nakamura, H. A Free-energy Landscape for Coupled Folding and Binding of an Intrinsically Disordered Protein in Explicit Solvent from Detailed all-atom Computations. *J. Am. Chem. Soc.* **2011**, *133* (27), 10448−10458.

(78) Hazy, E.; Bokor, M.; Kalmar, L.; Gelencser, A.; Kamasa, P.; Han, K.-H. Distinct Hydration Properties of Wild-type and Familial Point Mutant A53T of α-synuclein Associated with Parkinson's Disease. *Biophys. J.* **2011**, *101* (9), 2260−2266.

(79) Wostenberg, C.; Kumar, S.; Noid, W. G.; Showalter, S. A. Atomistic Simulations Reveal Structural Disorder in the RAP74-FCP1 Complex. *J. Phys. Chem. B* **2011**, *115* (46), 13731−13739.

(80) Savard, P.-Y.; Daigle, R.; Morin, S.; Sebilo, A.; Meindre, F.; Lagüe, P. Structure and Dynamics of Mycobacterium Tuberculosis Truncated Hemoglobin N: Insights from NMR Spectroscopy and Molecular Dynamics Simulations. *Biochemistry* **2011**, *50* (51), 11121−11130.

(81) Shan, B.; Li, D.-W.; Brüschweiler-Li, L.; Brüschweiler, R. Competitive Binding between Dynamic p53 Transactivation Sub-domains to Human MDM2 Protein: Implications for Regulating the p53·MDM2/MDMX Interaction. *J. Biol. Chem.* **2012**, *287* (36), 30376−30384.

(82) Kumar, P.; Chimenti, M. S.; Pemble, H.; Schönichen, A.; Thompson, O.; Jacobson, M. P.; Wittmann, T. Multisite Phosphorylation Disrupts Arginine-glutamate Salt Bridge Networks Required for Binding of Cytoplasmic Linker-associated Protein 2 (CLASP2) to end-binding Protein 1 (EB1). *J. Biol. Chem.* **2012**, *287* (21), 17050−17064.

(83) Knott, M.; Best, R. B. A Preformed Binding Interface in the Unbound Ensemble of an Intrinsically Disordered Protein: Evidence from Molecular Simulations. *PLoS Comput. Biol.* **2012**, *8* (7), e1002605.