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

Katherine Aurelia Ball[1], David E. Wemmer[1,2], Teresa Head-Gordon[1,2,3,4*]

[1]*Graduate Group in Biophysics, Berkeley, CA 94720*
[2]*Department of Chemistry, University of California, Berkeley, CA 94720*
[3]*Department of Bioengineering, University of California, Berkeley, CA 94720*
[4] *Department of Chemical and Biomolecular Engineering, University of California, Berkeley, CA 94720y*

SUPPLEMENTARY INFORMATION

Details of generating the different structural ensembles

RC ensemble. To generate a RC or Pred-SS structural ensemble we used the TraDES software package[1]. TraDES generates a random coil (RC) ensemble of structures by building each conformation one residue at a time and picking the dihedral angles according to their probability from the Ramachandran plot. This procedure also avoids steric clashes, but does not attempt to distinguish between energetically favorable and unfavorable conformations. We used the -kT flag during TraDES runs in order to avoid generating right handed structures due to a known bug.

Pred-SS ensemble. There is also an option to bias the TraDES ensemble to preferentially sample α-helix or β-sheet regions along the sequence according to a bioinformatics prediction of secondary structure, which we label the Pred-SS ensemble. Many bioinformatics tools allow prediction of a protein's secondary structure based on its sequence alone by comparing the sequence to known structures in the PDB. We employ the Psi-Pred V3.0 server[2], for this purpose, feeding it the Aβ40 and Aβ42 sequences and receiving a prediction of either extended, helical, or coil structure for each residue, as shown in Table S1. For both peptides, PsiPred predicts blocks of extended, β-structure with high confidence, consistent with the structure that Aβ adopts in the aggregated fibril state[3]. This does not guarantee that the resulting conformations will contain true cooperative secondary structure (such as β-sheet), since the secondary structure

state of each residue is picked independently of other residues by TraDES. The RC and Pred-SS ensembles each contain 100,000 structures as recommended in [4].

Pred-SS-ENS ensemble. As an example of a knowledge-based approach we consider the ENSEMBLE software package, which selects from a large starting pool (basis set) of structures, typically generated by TraDES, a subset of 100 conformations that best conform to various NMR experimental data supplied to it (see Section 5 of the paper on calculation of experimental observables from structures). The Pred-SS-ENS, was selected by the ENSEMBLE program from a starting 'soup' consisting of the Pred-SS ensemble structures. We supplied H_α, H_N, C_α, and C_β chemical shifts, J-coupling constants, RDCs and NOEs for both Aβ40 and Aβ42. We used default values of the ENSEMBLE program input parameters and the default output of a 100-structure ensemble. We ran each ENSEMBLE optimization for 48 hours on a Cray XE6 at National Energy Research Scientific Computing Center (NERSC), during which ~500 rounds of ENSEMBLE optimization steps were completed. We found that the resulting Pred-SS-ENS ensemble satisfied most of the experimental criteria besides RDCs according to the ENSEMBLE software. Based on our subsequent analysis, the default convergence criteria for RDCs appear to be too strict. J-coupling criteria were not satisfied for Aβ40, while Cα chemical shifts were not satisfied for Aβ42, however we selected those ensembles with the best fit to the NMR data overall, which we found to be sufficient for quantitative analysis.

MD ensemble. We created the fourth ensemble with *de novo* molecular dynamics simulations of Aβ40 and Aβ42 using the Amber ff99SB force field[5] and aqueous solvent represented by the TIP4P-Ew water model[6], which we chose because previous studies support its clear superiority relative to other biomolecular simulation force fields[7]. We simulated each amyloid-β peptide in a cubic box containing 6,251 water molecules for Aβ42 and 6,136 water molecules for Aβ40, with three Na+ ions to neutralize the charge of the peptide. The sander module of AMBER[8] was used in conjunction with Multi-Reservoir Replica Exchange (MRRE) method[9] to generate ~2 μs of MD trajectories, from which we created a Boltzmann weighted ensemble of 72,632 – 89,469 structures each for Aβ40 and Aβ42 respectively at the experimental temperature of 287 K.

MD-ENS ensemble. After reaching the conclusion that no true cooperative secondary structure can be generated from the purely knowledge-based method that selects from a pool of random structures, we decided to construct a fifth ensemble that uses a knowledge-based

approach combined with MD. For this ensemble (MD-ENS), we employed the same ENSEMBLE procedure described above for the Pred-SS-ENS ensemble except that we used the *de novo* MD ensemble as the starting pool of structures from which the experimentally optimized 100-structure ensemble was selected. NOE data was not included in our final MD ENSEMBLE refinement because a more structured ensemble could not satisfy the large number of distance restraints. We again had the same problem satisfying RDC convergence criteria, but otherwise the MD-ENS ensemble was optimized well to the NMR data after ~500 ENSEMBLE rounds. We note that in this paper the MD-ENS ensemble consists of only 100 structures, and not 2,000 structures as in our previous study[10]..

Details of using the ENSEMBLE software

For knowledge-based approach we used the ENSEMBLE software package to select a final ensemble of 100 structures from a starting pool of structures. The ENSEMBLE method formulates energy functions that score structures favorably when they agree with an experimental observable and unfavorably when they do not. To generate the Pred-SS-ENS and MD-ENS ensembles we performed two sets of ENSEMBLE optimization. The first used the Pred-SS ensemble as the starting 'soup' of structures while the second used our de novo MD ensemble. In our first attempt to select a final ensemble that agreed with experiment we supplied multiple atom type chemical shifts, J-coupling constants, RDCs, and NOE contacts that we could assign directly from the experimental data for both Aβ40 and Aβ42. For the NOEs, we did not have specific distance restraints for the contacts, so we set the distance to a maximum of 7.0 Å for each contact. We used default values of the ENSEMBLE program for the experimental observable target energies. As recommended by the Forman-Kay group, we set chemical shifts and NOE distances to converge first, before converging J-coupling constants or RDCs. We ran each ENSEMBLE optimization for 48 hours on a Cray XE6 at NERSC, during which ~500 rounds of ENSEMBLE optimization steps were completed.

After this first ENSEMBLE attempt we saw that while the both chemical shifts and NOE distance restraints converged for the Pred-SS-ENS ensemble, in the MD-ENS calculation the NOE distance restraints did not converge. For the MD-ENS optimization the NOE distance restraint energies were extremely high (~800 in the best ensemble, compared with the target energy of 49). This was probably due to the approximation of the NOE cross-peaks as simple

distance restrains, the large number of NOE distance restraints being optimized (177 for Aβ42 and 340 for Aβ40), and the fact that the MD starting pool was more diverse than the Pred-SS pool. We then attempted a second MD-ENS ENSEMBLE run without the NOE distance restraints, and this time the chemical shifts did converge.

The second stage of the ENSEMBLE runs optimized against J-coupling constants and RDCs. This round was able to successfully converged the J-coupling constants to the default energy tolerance (except for the Aβ40 Pred-SS-ENS ensemble), but not the RDCs. Although the RDC energies had improved (from ~80 to 51 on energy units), their convergence had stagnated so as to not meet the ENSEMBLE convergence tolerance of >0.2 Hz, keeping a constant energy of ~51, even with additional computation time of ~5000 ENSEMBLE optimization rounds. In actuality, the RDCs were sufficiently converged to permit the analysis in sections 5 and 6 of the paper, suggesting that the default RDC tolerance was too tight. We therefore selected the final ensembles of 100 structures with the lowest energy according to the ENSEMBLE weighting function, in order to give the overall best performance against the chemical shift, J-coupling, RDC, and, for the Pred-SS-ENS ensemble, NOE data. We then used these final ensembles as the Pred-SS-ENS and MD-ENS ensembles in our subsequent analyses.

REFERENCES

1. (a) Feldman, H. J.; Hogue, C. W. V., Probabilistic sampling of protein conformations: new hope for brute force? *Proteins: Struct., Func., Bioinform.* **2002,** *46* (1), 8-23; (b) Feldman, H. J.; Hogue, C. W., A fast method to sample real protein conformational space. *Proteins: Struct., Func., Bioinform.* **2000,** *39* (2), 112-131.
2. (a) Jones, D. T., Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **1999,** *292* (2), 195-202; (b) Buchan, D. W. A.; Ward, S. M.; Lobley, A. E.; Nugent, T. C. O.; Bryson, K.; Jones, D. T. Protein annotation and modelling servers at University College London *Nuc. Acid. Res.* [Online], 2010, p. W563-568.
3. Petkova, A. T.; Ishii, Y.; Balbach, J. J.; Antzutkin, O. N.; Leapman, R. D.; Delaglio, F.; Tycko, R., A structural model for Alzheimer's beta -amyloid fibrils based on experimental constraints from solid state NMR. *Proc Natl Acad Sci U S A* **2002,** *99* (26), 16742-7.
4. 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.
5. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C., Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* **2006,** *65* (3), 712-25.
6. Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T., Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J. Chem. Phys.* **2004,** *120* (20), 9665-78.

7. (a) Wickstrom, L.; Okur, A.; Simmerling, C., Evaluating the Performance of the ff99SB Force Field Based on NMR Scalar and Coupling Data. *Biophysical Journal* **2009,** *97* (3), 853-856; (b) 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 ; (c) Fawzi, N. L.; Phillips, A. H.; Ruscio, J. Z.; Doucleff, M.; Wemmer, D. E.; Head-Gordon, T., Structure and dynamics of the Abeta(21-30) peptide from the interplay of NMR experiments and molecular simulations. *J. Am. Chem. Soc.* **2008,** *130* (19), 6145-58; (d) 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. *Biochem.* **2011,** *50* (35), 7612-7628.

8. Case, D. A.; Darden, T. A.; III, T. E. C.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M. J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A., AMBER 11. **2010**.

9. Ruscio, J.; Fawzi, N.; Head-Gordon, T., How hot? Systematic convergence of the replica exchange method using multiple reservoirs. *J. Comput. Chem.* **2010,** *31* (3), 620-627.

10. 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.

Table S1. *Aβ40 and Aβ42 predicted secondary structure.* For Aβ40 and Aβ42 the predicted secondary structure and confidence in that prediction are presented[29]. This corresponds to a TraDES ensemble generated with extended dihedral angles for those residues predicted to have extended structure. The percent of the TraDES ensemble that has that residue extended corresponds to the confidence of the prediction, and the rest of the TraDES ensemble is generated with random coil structure.

	Aβ40				Aβ42			
Residue	structure prediction	prediction confidence	TraDES % extended	TraDES % coil	structure prediction	prediction confidence	TraDES % extended	TraDES % coil
1	C	9	0	100	C	9	0	100
2	C	4	0	100	C	4	0	100
3	C	2	0	100	C	2	0	100
4	C	1	0	100	C	1	0	100
5	C	1	0	100	C	1	0	100
6	C	1	0	100	C	0	0	100
7	C	4	0	100	C	4	0	100
8	C	6	0	100	C	6	0	100
9	C	5	0	100	C	5	0	100
10	E	0	0	100	E	0	0	100
11	E	5	50	50	E	5	50	50
12	E	7	70	30	E	7	70	30
13	E	7	70	30	E	8	80	20
14	E	6	60	40	E	6	60	40
15	E	4	40	60	E	5	50	50
16	E	8	80	20	E	8	80	20
17	E	9	90	10	E	9	90	10
18	E	8	80	20	E	8	80	20
19	E	8	80	20	E	8	80	20
20	E	7	70	30	E	7	70	30
21	E	5	50	50	E	5	50	50
22	E	0	0	100	E	0	0	100
23	C	1	0	100	C	1	0	100
24	C	2	0	100	C	2	0	100
25	C	7	0	100	C	7	0	100
26	C	8	0	100	C	8	0	100
27	C	8	0	100	C	8	0	100
28	C	4	0	100	C	4	0	100
29	C	2	0	100	C	2	0	100
30	E	1	10	90	E	1	10	90
31	E	8	80	20	E	9	90	10
32	E	9	90	10	E	9	90	10
33	E	9	90	10	E	9	90	10
34	E	9	90	10	E	9	90	10
35	E	7	70	30	E	7	70	30
36	E	5	50	50	E	6	60	40
37	E	0	0	100	E	0	0	100
38	E	0	0	100	E	0	0	100
39	E	2	20	80	E	8	80	20
40	C	9	0	100	E	9	90	10
41					E	4	40	60
42					C	9	0	100

FIGURES

Figure S1. *Percentage of Aβ42 simulated ensemble in different types of secondary structure by residue for the RC ensemble.* The red line represents helix, the blue line for anti-parallel sheet, and the black line for β-turns.

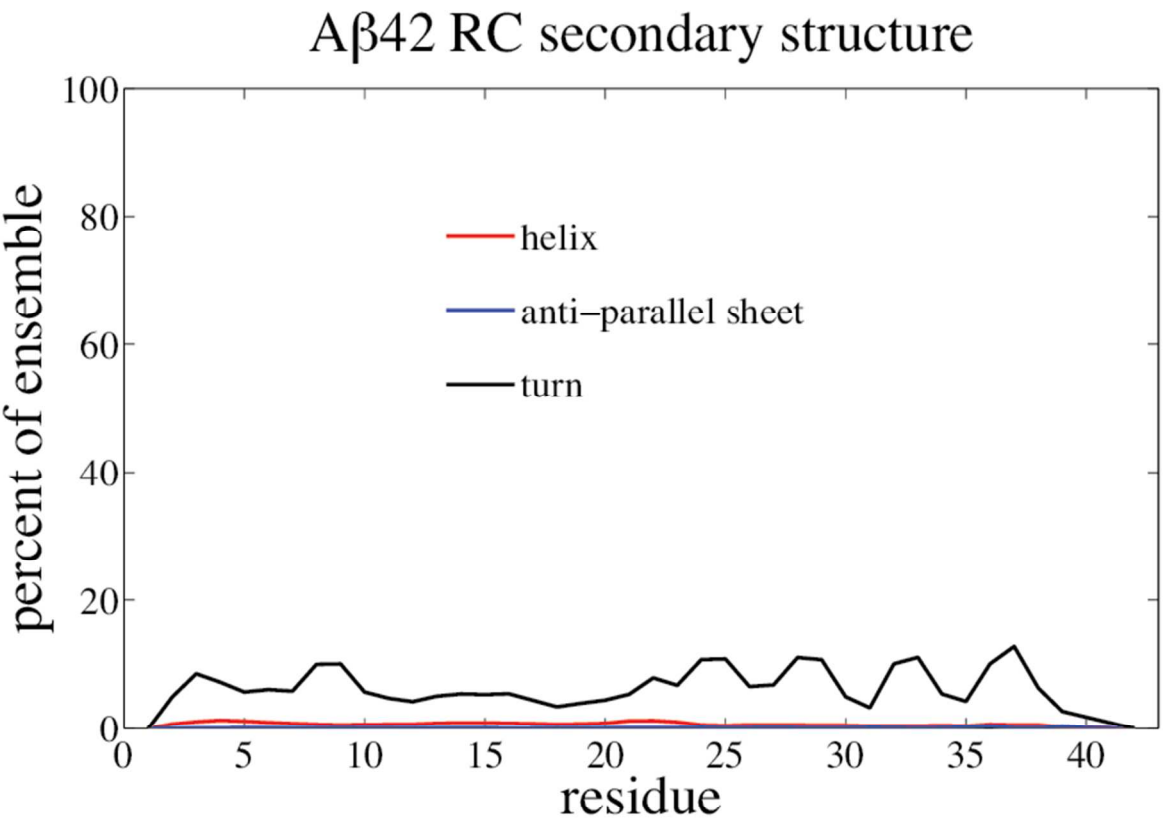

Figure S1. Ball and co-workers