Supplementary Information

Alzheimer's Protective A2T mutation changes the conformational landscape of the $A\beta_{1\text{-}42}$ monomer differently than does the A2V mutation

Payel Das[1*], Brian Murray[2], Georges Belfort[2]

[4]Soft Matter Theory and Simulations Group, Computational Biology Center, IBM Thomas J. Watson Research Center, Yorktown Heights, NY 10598.

[2]Howard P. Isermann Department of Chemical and Biological Engineering, and Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute, Troy, New York 12180-3590.

Model and methods

The monomeric conformational ensembles were generated by performing replica-exchange molecular dynamics (REMD) simulations of the solvated peptides using the GROMACS4 software [1]. REMD is an enhanced sampling algorithm that helps the system to escape the local minima in the free energy landscape by increasing temperature [2]. The method consists of several identical copies or replicas of the system, which are simulated in parallel over a range of temperatures. At frequent intervals, trials to exchange the temperature of all neighboring replicas are performed, according to a Metropolis Monte Carlo criterion. The swapping probability is chosen to satisfy a detailed balance. This method has been successfully applied to construct the ensemble of intrinsically disordered peptides that lack a single native conformation and instead populate multiple rapidly interchanging states, such as Aβ at atomic resolution [3-6].

The following protocol was used to construct the initial conformation for REMD simulations for each variant peptide: first, a 10 ns MD simulation at high temperature (~700 K) in vacuum was performed starting from a fully extended peptide conformation with charged termini. The collapsed peptide was solvated in a 56 x 56 x 56 $Å^3$ cubic box containing ~5,600 water molecules. The protonation states of the acidic and basic residues of the peptide were set at pH 7 and three Na^+ ions were added to neutralize the charge. The solvated peptide was equilibrated for 2 ns in an NPT ensemble (300 K and 1 atm) before the REMD run. Finally, constant volume REMD simulations were run for 175 ns per replica with an integration step of 2 ns, resulting in an aggregate simulation time of 11.2 μs per system. A total of 64 replicas within an exponentially distributed temperature range [7] of 276-592 K were used for each system and the replica exchange attempts were made every 4 ps. Use of this protocol results in the average exchange ratio of 19% that is constant over the temperature range. The system was coupled to a Nose-Hoover heat bath to maintain constant temperature between swaps. The particle-mesh Ewald (PME) method was used for the long-range electrostatic interactions [8], while the van der Waals interactions were treated with a cut-off distance of 10 Å. The bonds were constrained using LINCS [9] and SETTLE [10] algorithms. For all calculations, a combination of OPLS-AA force-field [11] and TIP3P water model [12] was used. This combination has been reported to generate the structural ensemble that is in agreement with NMR experiments [13].

All MD simulations were run using IBM BlueGene/Q supercomputers. The convergence of the simulations was checked by dividing the simulation data in two or four equal sets and estimating the similarity of the secondary/tertiary structure obtained from those sets. The secondary structure was estimated using the STRIDE program [14]. A cutoff distance of 8 Å between C_α atoms was considered to define a contact between two residues. Only non-sequential contacts ($|i-j| \geq 3$) were considered for tertiary interactions. For contacts between side-chains of two residues, a cutoff distance of 5 Å (heavy atom only) was used. Hydrogen bonds were determined with a cutoff of 3.5 Å for the donor-acceptor distance and a cutoff of 30° for the donor-hydrogen-acceptor angle.

Potential of mean force (PMF, W(X)) plots were obtained from a histogram analysis, using the equation $W(X) = -RT \log(p(X))$, where X is the set of reaction coordinates and p(X) is the probability. The number of contacts between the N-terminus (residues 1-5) and CHC (residues 16-21), termed as $NC_{Ntr\text{-}CHC}$, and the number of contacts between the CHC and C-terminus (residues 31-42), termed as $NC_{CHC\text{-}CTR}$, were used as the reaction coordinates for PMF estimation. A cluster analysis using the Daura algorithm [15] was performed. A 3 Å C_α-RMSD cut-off between two conformations was used for cluster analysis of the highly populated regions on the PMF plots. Regions on the PMF plots that individually represent ≥5% of total production ensemble were only considered for further analysis. The criteria used to define those regions can be found in Table S1. All peptide structure figures were rendered using VMD [16].

Equilibration and validation of the REMD ensemble

To make sure that the choice of the initial structure did not affect our results, C_α root-mean-square-distance (RMSD) from the initial structure was estimated as a function of the simulation time for the 305 K replica for all three variants (**Fig. S1a**). The RMSD reaches a sufficiently high value of 20 Å with first 15 ns of simulation and then fluctuates around that value for each variant, suggesting that the structural ensemble populated after 15 ns is independent of the initial structure.

Figure S**1b-c** shows the evolutions of radius of gyration, Rg, and turn propensities of the 305 K trajectory for each variant. Rg reaches an average value of ~10-11 Å after ~60 ns, at what point the turn propensity steadily fluctuates around ~50%, for all three variants. Thus, the first 60 ns of each replica was discarded as the equilibration time and the 60-175 ns portion of the twelve REMD simulation trajectories in 276-308 K temperature range was considered as the production ensemble comprised of ~28,000 conformations. We validate the simulated WT ensemble by comparing the $^3J_{HNHA}$ couplings derived from backbone ϕ dihedrals in simulations with NMR experiments [17]. **Figure S1d** shows the comparison between the simulated $^3J_{HNHA}$ couplings back-calculated from the production period of the 278 K replica of WT and experiments using the three different parameter sets for the Karplus equation [17]. The set proposed in Vuister and Bax [18] produces universally high MD-derived J-coupling values, implying that simulations might have over-sampled β-conformations, and yields an RMSD of 1.16 Hz. On the other hand, use of parameters for WT $A\beta_{1-42}$ proposed in Sgourakis et al. [19] and Rosenman et al. [17] produces lower RMSD values (0.91 Hz and 0.94 Hz, respectively) and better agreement with the experimental values. This is because of the fact that the Vuister and Bax's parameters were determined by fitting to well-ordered X-ray structures, whereas the parameters proposed in Sgourakis et al. [19] and Rosenman et al. [17] account for the motional averaging effect within a MD dataset. The residues that show strongest deviation (>1 Hz) from the corresponding experimental value are F4, H13, A21, and S26. It is interesting that the J-coupling values obtained for S26 from two different NMR pulse sequences – HNHA [19] and J-resolved SOFAST-HMQC [17] (shown in Fig. 1c, black line) - also vary significantly. The discrepancy between experimental and MD-derived J-couplings for F4, H13, and A21 could be due to the force-field limitation.

Additional ensemble properties

The Rg distributions show a single peak at ~10-11 Å for all three variants (**Fig. S2a**), indicating a collapsed coil structure (the value for the statistical coil is ~20 Å). This value is also consistent with what has been reported before for the WT monomer in simulations and experiments [20, 21]. A cluster analysis on the production ensemble (**Fig. S2b**) reveals that the top 50 clusters only represent ~20% of the population for all three systems, suggesting their intrinsic disordered nature. The fact that a lower C_α-RMSD cutoff (4.5 Å) for the A2V system can be used to produce similar results compared to the WT and A2T peptides (6 Å cutoff) suggests relative lowering of disorder in A2V.

Convergence of the secondary and tertiary structure estimated from REMD

To determine whether the sampling has converged or not in terms of the secondary structure propensity, the 60-175 ns production ensemble was divided in four equal ~25 ns long segments and the standard deviation values were estimated (**Fig. S3a**). Overall, the values of standard deviations appear small. It was earlier reported that the β-content is the slowest to converge [17]. **Fig. S3a** confirms that the standard error for the β-element composition is small enough (1.6%, 1.4%, and 1.9% for WT, A2V, and A2T, respectively). The sampling convergence in terms of the tertiary structure was checked by dividing

the simulation data in two ~55 ns equal segments and comparing the tertiary structure obtained from those sets. The results are shown in **Fig. S3c-e**. No new major features were obtained by doing so, suggesting reasonable convergence of the tertiary structure. To confirm that 175 ns per replica produces a good representation of the structural ensemble of the monomeric variants, we extended the simulations for each system for another 50 ns, resulting 225 ns per replica and compared with results obtained from 60-175ns portion. The secondary and tertiary structure distributions from the extended runs are shown in Figures **S3b** and **S3c-e** (bottom). Overall, the composition and major features remain unchanged and only subtle changes are detected, suggesting that the structural ensemble generated by the procedure described in the manuscript has reached a steady state in terms of secondary and tertiary structures and provides a good representation of the peptide variant monomers in solution.

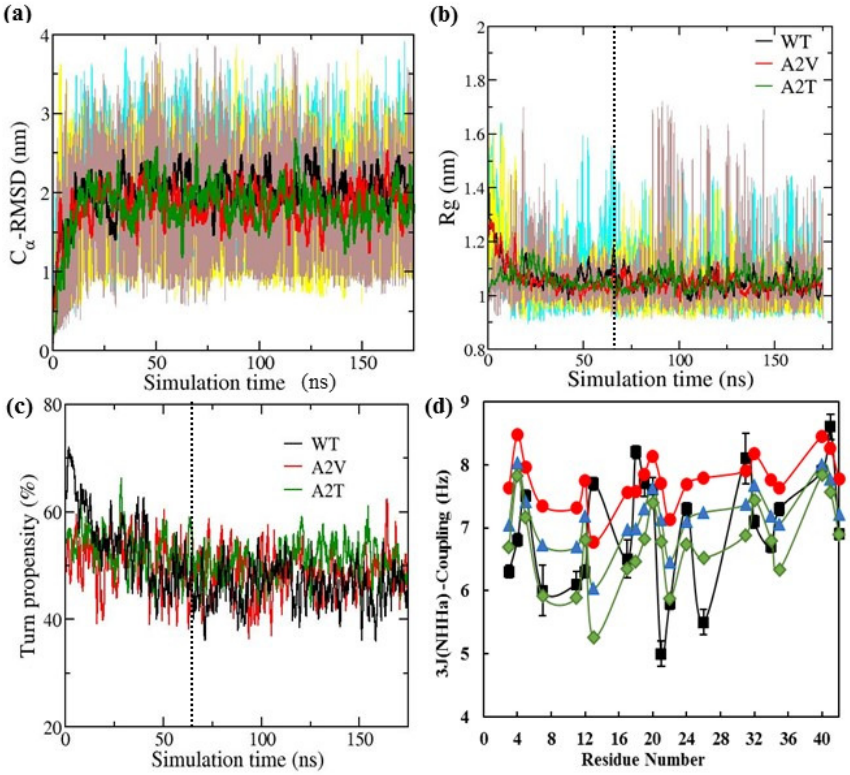

Figure S1: Equilibration and validation of REMD trajectories. (a) Evolution of the C_α root-mean-square-distance (RMSD) from the initial structure for the trajectories at 305 K (black: WT; red: A2V; green: A2T). Moving averages of 1 ns are shown. The raw data (cyan: WT; yellow: A2V; brown: A2T) are also shown. **(b-c)** Evolutions of the radius of gyration, Rg, and turn propensity, for the trajectories at 305 K (black: WT; red: A2V; green: A2T). Moving averages of 1 ns are shown. The raw data (cyan: WT; yellow: A2V; brown: A2T) are also shown in (b). The black dotted lines indicate equilibration around 60 ns. **(d)** Comparison between experimental $^3J_{HNHA}$ couplings (black) and MD-derived corresponding values (red: Vuister parameters; green: Sgourakis parameters; blue: Rosenman parameters) back-calculated from the production period of the WT 278K replica.

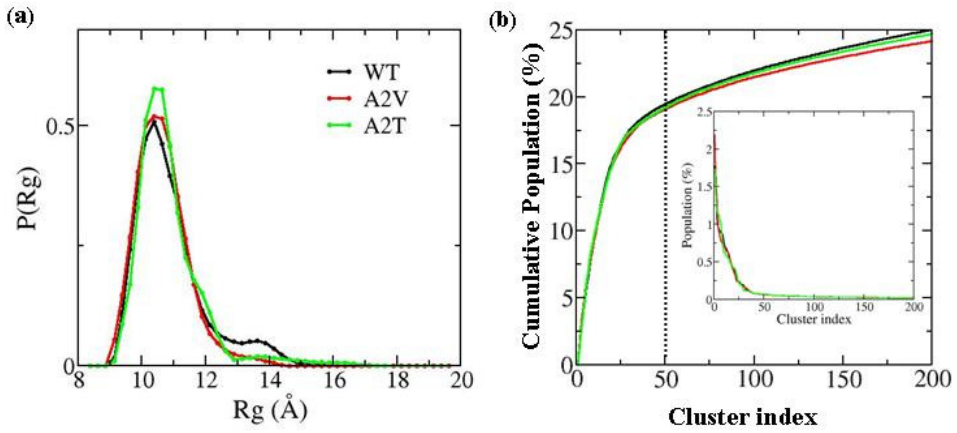

Figure S1. (a). Rg distributions and (b) Cluster analysis. **(a)** Normalized Rg histograms for three variant monomers (black: WT; red: A2V; green: A2T), as obtained from REMD. **(b)** Cumulative population (in %) of the top 200 clusters for each system (black: WT; red: A2V; green: A2T), as obtained using the greedy algorithm with a Cα–RMSD cutoff of 6 Å for the WT and A2T systems and of 4.5 Å for the A2V systems. The black dotted line indicates that top 50 clusters represent only ~20% of the total population for all three systems. **Inset.** Individual population (in %) of the top 200 clusters as a function of cluster index.

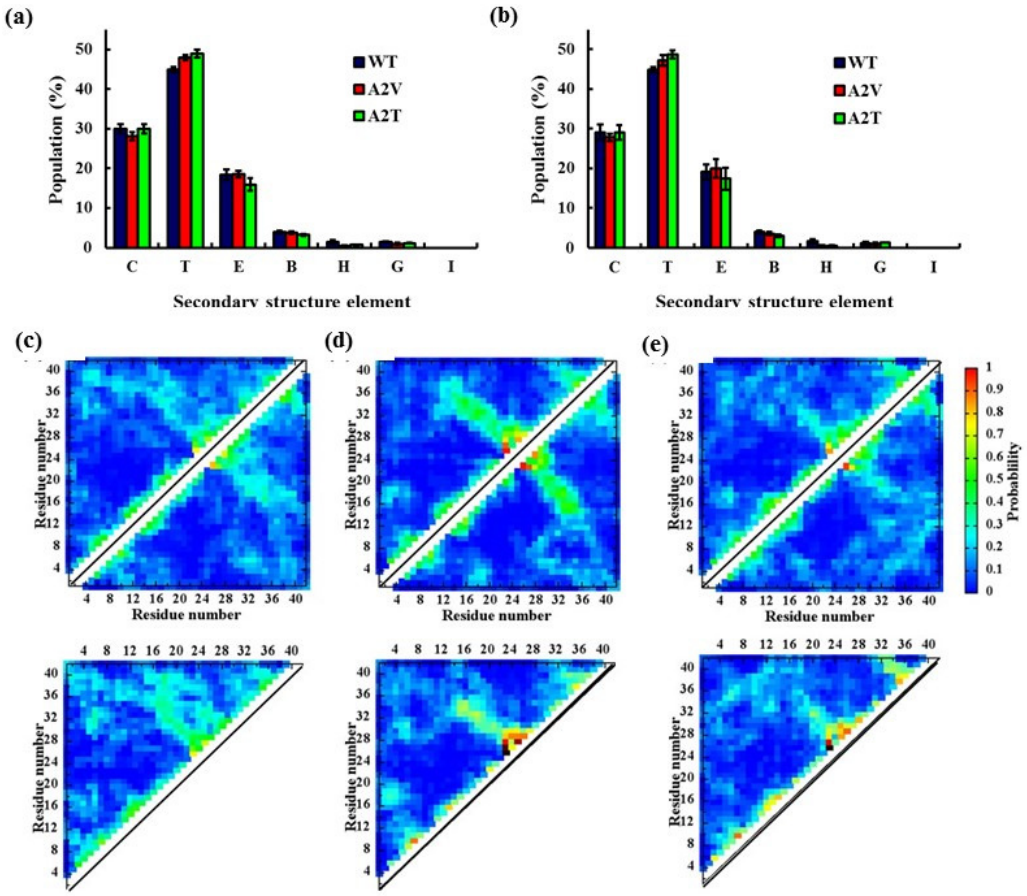

Figure S3: Convergence of the secondary and tertiary structure. The ensemble-averaged percentage population of each secondary structural element averaged over all residues. Results obtained from the production ensemble (60-175 ns of 276-308 K replicas) are shown in **(a)**, whereas results obtained from the extended runs (60-225 ns of 276-308 K replicas) are shown in **(b)**. The standard deviations were estimated by splitting the data in four (b) or six (c) equal ~25 ns long segments. Results on the convergence of tertiary contact distribution is shown in **(c-e)**, WT **(c)**, A2V **(d)**, A2T **(e)**. Top panel: The data from the production ensemble (60-175 ns of 276-308 K replicas) were divided into two equal ~55 ns long subsets (shown in upper and lower triangles of the contact maps) and the probabilities were averaged over each of the subsets. Bottom panel: Results obtained from the extended runs (175-225 ns portion of 276-308 K replicas). Color scheme used is same as in Fig.1d-f. No new major features were obtained by doing so, suggesting reasonable sampling convergence of the REMD simulations.

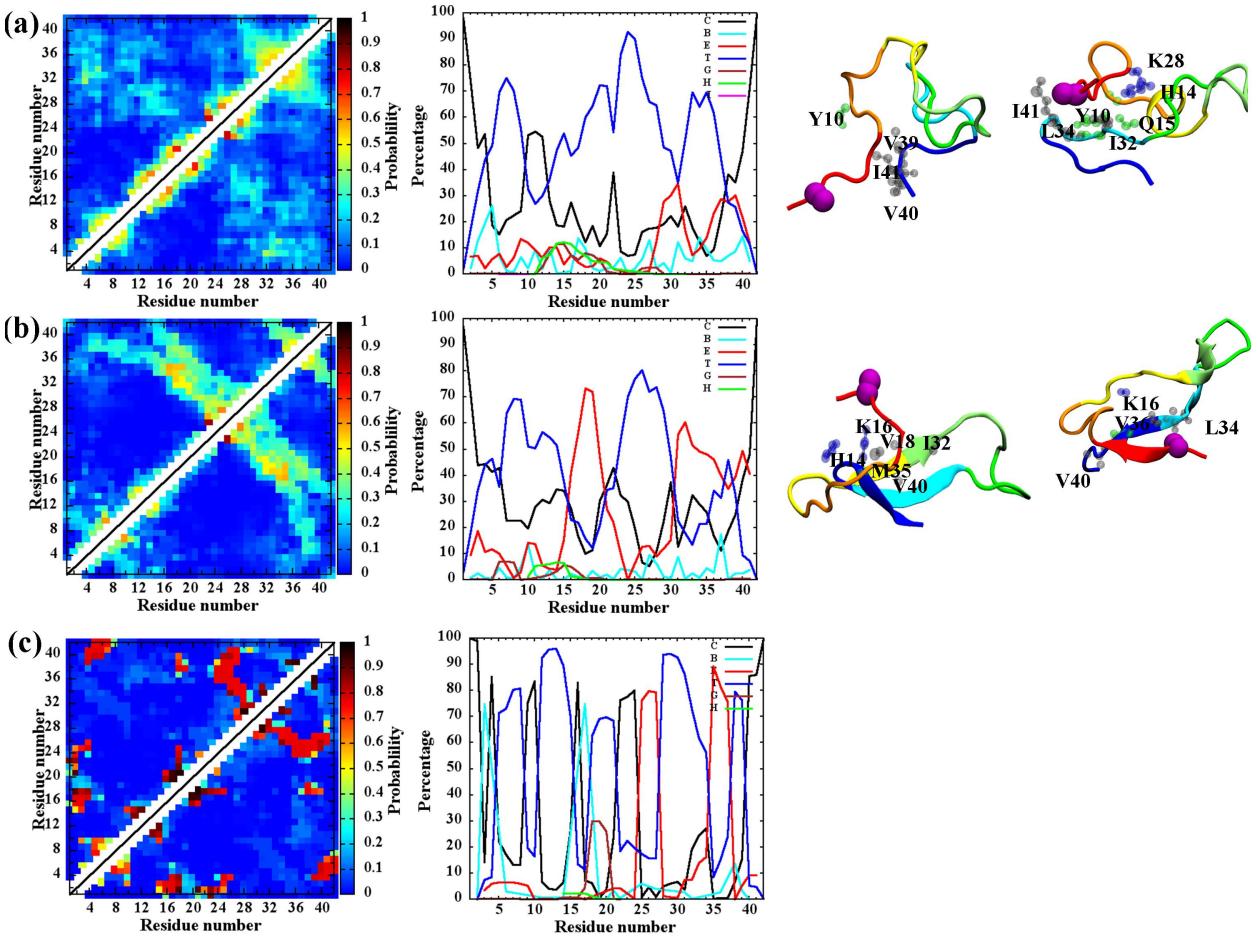

Figure S4: **Structural analyses of the most populated regions of the WT PMF plot (Fig. 2a).** The tertiary contact maps and the residue-based secondary structure for conformations corresponding to the **(a)** S1 region, **(b)** S2 region, and **(c)** S3 region for the WT peptide. S1 region represents structures with limited CHC-CTR and Ntr-CHC contacts. The structures with robust CHC-CTR interaction and a few (≤5) Ntr-CHC contacts are represented by the S2 region. Conformations with stronger Ntr-CHC interaction that lacked substantial CHC-CTR contacts are represented by the weakly populated S3 region. The representative conformations of the second and third largest clusters are also shown for the S1 and S2 regions along with residues forming contacts with the Ntr region. Color schemes used for the figures are the same as in Fig. 1-2.

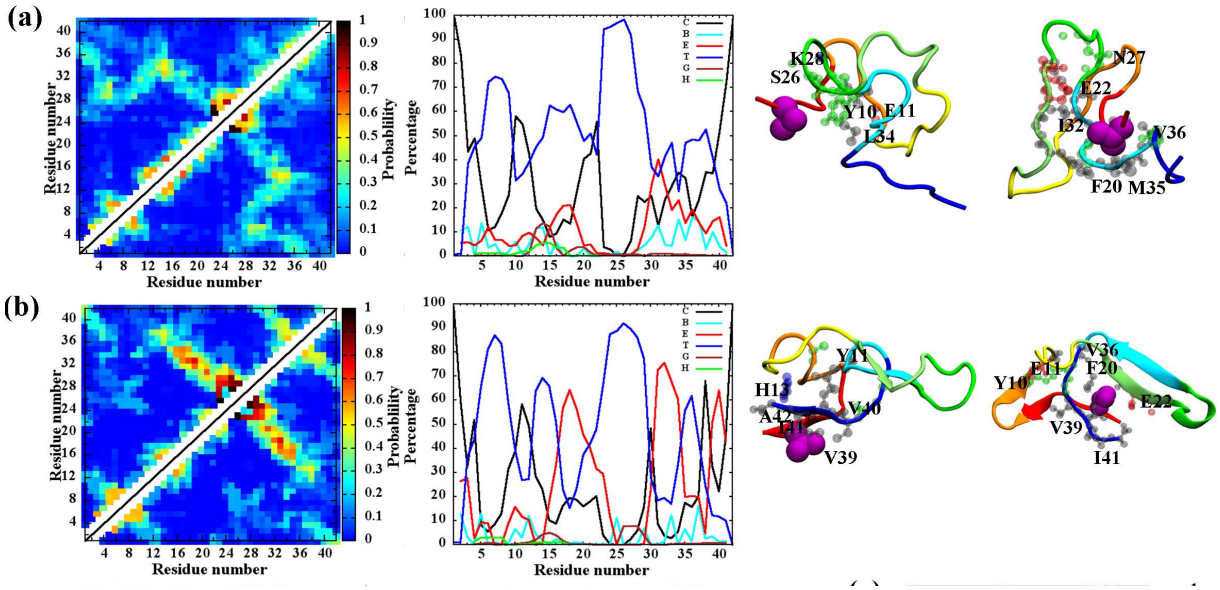

Figure S5: **Structural analyses of the S1 and S2 regions of the A2V PMF plot (Fig. 2b).** The tertiary contact maps and the residue-based secondary structure for conformations corresponding to the **(a)** S1 region, and **(b)** S2 region for the A2V peptide. The representative conformations of the second and third largest cluster are also shown for the S1 and S2 regions along with residues forming contacts with the Ntr region. Color schemes used for the figures are the same as in Fig 1-2.

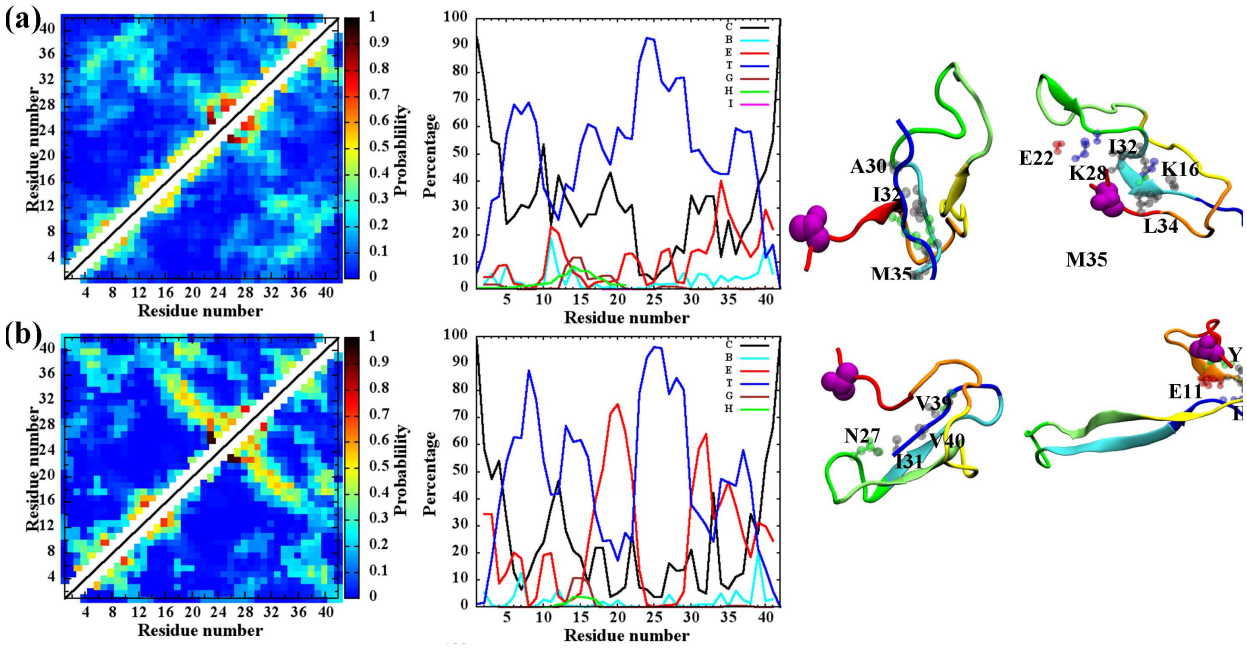

Figure S6: **Structural analyses of the S1 and S2 regions of the A2T PMF plot (Fig. 2c).** The tertiary contact maps and the residue-based secondary structure for conformations corresponding to the **(a)** S1 region, and **(b)** S2 region for the A2T peptide. The representative conformations of the second and third largest cluster are also shown for the S1 and S2 regions along with residues forming contacts with the Ntr region. Color schemes for the figures are the same as in Fig 1-2.

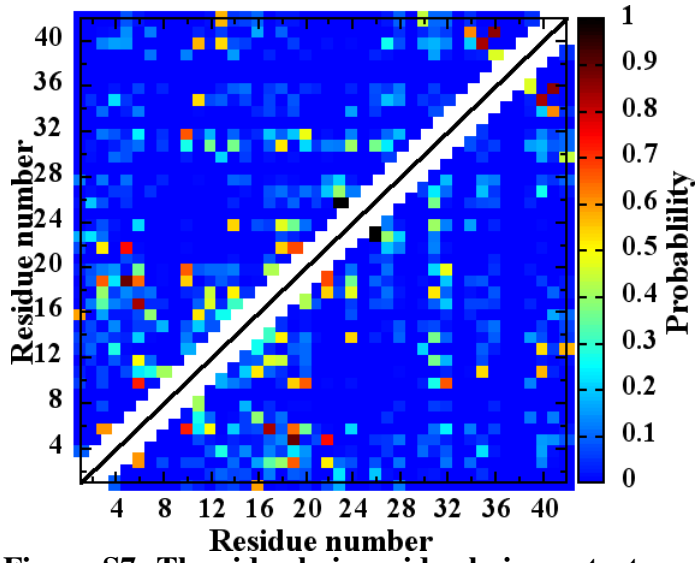

Figure S7: The side-chain—side-chain contact map for the S3 population of A2T. A cutoff of 5 Å was used for defining a contact between side-chains of two residues.

Table S1: Definition of summary of cluster analysis results for S1-S5 regions of all three variants (Fig. 2). NC1 and NC2 represent number of contacts between Ntr-CHC and CHC-CTR, respectively. Regions on the PMF plots that individually represent ≥5% of total production ensemble are only considered. The numbers within parentheses represent the population using the following criteria: S1 (NC1=[0-2], NC2=[0-13]); S2 (NC1=[0-2], NC2=[14-25]); S3=(NC1=[2-15], NC2=[0-10]).

WT	S1 NC1 = [0-2] NC2 = [0-13]	S2 NC1 = [0-2] NC2 = [14-25]	S3 NC1 = [8-11] NC2 = [0-9]	S4	S5
% of total population	32.6	35.9	6.6 (11.4)	N/A	N/A
# clusters	105	107	8		
Size of top five clusters (%)	7.6 6.5 6.0 5.4 5.2	14.3 11.1 7.7 6.9 6.3	76.4 8.0 5.8 4.5 3.3		
A2V	S1 NC1 = [0-3] NC2 = [0-13]	S2 NC1 = [0-1] NC2 = [14-25]	S3 NC1 = [11-15] NC2 = [4-11]	S4 NC1 = [5-13] NC2 = [13-17]	S5 NC1 = [10-13] NC2 = [19-25]
% of total population	40.1 (38.6)	24.9 (25.6)	5.9 (9.1)	9.1	5.5
# clusters	127	78	7	23	6
Size of top five clusters (%)	10.7 8.9 8.6 6.7 5.7	15.6 13.3 13.1 11.6 10.3	79.0 8.6 6.2 4.2 1.6	57.8 11.6 5.5 3.6 3.6	83.7 14.9 0.8 0.3 0.1
A2T	S1 NC1 = [0-2] NC2 = [0-13]	S2 NC1 = [0-3] NC2 = [14-25]	S3 NC1 = [2-10] NC2 = [0-5]	S4 NC1 = [5-7] NC2 = [17-20]	S5 NC1 = [2-4] NC2 = [24-26]
% of total population	40.1 (40.1)	25.6 (24.9)	17.2 (21.5)	N/A	N/A
# clusters	137	68	36		
Size of top five clusters (%)	11.4 9.7 8.3 7.2 4.3	20.9 18.2 10.5 8.7 4.2	41.2 9.2 7.7 6.8 6.5		

Table S2: Probability (in %) of non-sequential salt-bridges between all charged residues for S1-S5 regions for all three variants. Those with probability <1% re marked as "-". The salt-bridges involving D1 and R5 (N-terminal residues) that are frequently formed in the S3 structures of the WT or the A2T monomer (**Fig. 2a & Fig. 2c**) are marked red.

	WT	S1	S2	S3	S4	S5
WT	D1-R5	12.8	2.7	-	N/A	N/A
	D1-K16	2.4	20.6	67.4		
	D1-K28	5.8	2.9	7.8		
	E3-K16	3.0	4.2	-		
	E3-K28	1.7	-	2.0		
	R5-E11	-	4.7	-		
	R5-E22	4.0	2.4	-		
	R5-D23	4.0	-	-		
	D7-K16	6.2	8.2	1.9		
	D7-K28	4.1	-	-		
	E11-K16	4.9	-	3.0		
	E11-K28	4.7	-	-		
	K16-E22	8.3	1.7	5.4		
	K16-D23	6.0	-	-		
	E22-K28	16.7	14.3	3.2		
	D23-K28	3.2	9.7	-		
A2V	D1-R5	-	-	-	9.7	-
	D1-K16	7.6	1.4	2.7	8.6	13.1
	D1-K28	3.8	6.3	-	-	-
	E3-K16	-	-	3.3	-	-
	E3-K28	-	-	-	-	-
	R5-E11	3.2	4.9	3.7	-	-
	R5-E22	1.8	-	-	-	-
	R5-D23	-	-	-	-	-
	D7-K16	2.5	-	2.9	40.5	-
	D7-K28	5.7	-	-	-	-
	E11-K16	15.0	21.9	4.1	5.8	-
	E11-K28	9.9	-	-	-	-
	K16-E22	11.6	-	-	-	-
	K16-D23	-	-	-	-	-
	E22-K28	29.3	25.8	16.4	12.6	9.6
	D23-K28	-	-	-	-	-
A2T	D1-R5	4.5	-	-	N/A	N/A
	D1-K16	4.2	8.5	33.8		
	D1-K28	6.2	2.0	5.7		
	E3-K16	-	-	-		
	E3-K28	4.0	-	-		
	R5-E11	1.2	-	-		
	R5-E22	1.5	1.8	7.8		
	R5-D23	1.0	6.3	-		
	D7-K16	-	-	3.3		
	D7-K28	-	6.3	-		
	E11-K16	18.6	3.2	15.3		
	E11-K28	9.1	5.4	-		
	K16-E22	7.2	1.2	-		
	K16-D23	-	-	-		
	E22-K28	35.5	12.4	5.6		
	D23-K28	1.5	2.0	1.6		

References

1. Hess, B., et al., *GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation*. Journal of Chemical Theory and Computation, 2008. **4**(3): p. 435-447.
2. Sugita, Y. and Y. Okamoto, *Replica-exchange molecular dynamics method for protein folding*. Chemical Physics Letters, 1999. **314**(1-2): p. 141-151.
3. Fukunishi, H., O. Watanabe, and S. Takada, *On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction*. Journal of Chemical Physics, 2002. **116**(20): p. 9058-9067.
4. Miyashita, N., J.E. Straub, and D. Thirumalai, *Structures of beta-amyloid peptide 1-40, 1-42, and 1-55-the 672-726 fragment of APP-in a membrane environment with implications for interactions with gamma-secretase*. J Am Chem Soc, 2009. **131**(49): p. 17843-52.
5. Qiao, Q., G.R. Bowman, and X. Huang, *Dynamics of an intrinsically disordered protein reveal metastable conformations that potentially seed aggregation*. J Am Chem Soc, 2013. **135**(43): p. 16092-101.
6. Sgourakis, N.G., et al., *Atomic-level characterization of the ensemble of the Abeta(1-42) monomer in water using unbiased molecular dynamics simulations and spectral algorithms*. J Mol Biol, 2011. **405**(2): p. 570-83.
7. Patriksson, A. and D. van der Spoel, *A temperature predictor for parallel tempering simulations*. Physical Chemistry Chemical Physics, 2008. **10**(15): p. 2073-2077.
8. Deserno, M., Holm, C. , *How to mesh up Ewald sums. II. An accurate error estimate for the particleâ€“particleâ€“particle-mesh algorithm*. The Journal of Chem.. Phys., 1998. **109**: p. 8.
9. Hess, B., et al., *LINCS: A linear constraint solver for molecular simulations*. Journal of Computational Chemistry, 1997. **18**(12): p. 1463-1472.
10. Miyamoto, S. and P.A. Kollman, *Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models*. Journal of Computational Chemistry, 1992. **13**(8): p. 952-962.
11. Jorgensen, W.L., D. Maxwell, and J. Tirado-Rives, *Development and testing of the OPLS all-atom force field on confromational energetics and properties of organic liquids*. J. Am. Chem. Soc., 1996. **118**: p. 11225-11236.
12. Jorgensen, W.L., et al., *Comparison of simple potential functions for simulating liquid water*. J Chem Phys, 1983. **79**: p. 926-935.
13. Rosenman, D.J., et al., *Abeta Monomers Transiently Sample Oligomer and Fibril-Like Configurations: Ensemble Characterization Using a Combined MD/NMR Approach*. Journal of Molecular Biology, 2013. **425**(18): p. 3338-3359.
14. Frishman, D. and P. Argos, *Knowledge-based protein secondary structure assignment*. Proteins, 1995. **23**(4): p. 566-579.
15. Daura, X., et al., *Peptide Folding: When Simulation Meets Experiment*. Angewandte Chemie International Edition, 1999. **38**(1-2): p. 236-240.
16. Humphrey, W., A. Dalke, and K. Schulten, *VMD: visual molecular dynamics*. Journal of molecular graphics, 1996. **14**(1): p. 33-38.

17. Rosenman, D.J., et al., *Abeta Monomers Transiently Sample Oligomer and Fibril-Like Configurations: Ensemble Characterization Using a Combined MD/NMR Approach*. J Mol Biol, 2013. **425**(18): p. 3338-3359.
18. Vuister, G.W. and A. Bax, *Quantitative J Correlation - a New Approach for Measuring Homonuclear 3-Bond J(H(N)H(Alpha) Coupling-Constants in N-15-Enriched Proteins*. Journal of the American Chemical Society, 1993. **115**(17): p. 7772-7777.
19. Sgourakis, N.G., et al., *Atomic-level characterization of the ensemble of the Abeta (1-42) monomer in water using unbiased molecular dynamics simulations and spectral algorithms*. Journal of molecular biology, 2011. **405**(2): p. 570-583.
20. Nag, S., et al., *Nature of the Amyloid-beta Monomer and the Monomer-Oligomer Equilibrium*. Journal of Biological Chemistry, 2011. **286**(16): p. 13827-13833.
21. Baumketner, A., et al., *Amyloid β-protein monomer structure: A computational and experimental study*. Protein Science, 2006. **15**(3): p. 420-428.