Role of Hydrophobic Residues for the Gaseous Formation of Helical Motifs

Supplementary Information

Methods

MD simulations of 2Aβ in the aqueous phase. The structure of Aβ (1-40) monomer in aqueous solution is available (PDB ID: 1AML), but the one for the 2Aβ is still unknown.[1] Therefore, we here followed the computational protocol as proposed in a recent MD work.[2] Briefly, two Aβ monomers was explicitly dissolved in an octahedron box, where the center-of-mass distance of two monomers was 45 Å. The orientations of monomers were randomly assigned. The protonation state of protein at pH=7 was determined using the H++ web server.[3] Six sodium cations were added to neutralize the simulation system. The AMBER ff99SB-ILDN force field and the TIP3P force field were used to describe proteins/ions and water.[4-6] The periodic boundary conditions were applied in all three dimensions. Coulombic and Lennard-Jones interactions cut off at 10 Å. The particle mesh Ewald algorithm was employed to calculate electrostatic interactions.[7] The SHAKE algorithm was used to constrain all chemical bonds.[8] The temperature was maintained using the Nosé-Hoover thermostat.[9, 10] The pressure was maintained using the Berendsen barostat.[11] All classical MD simulations in this study were conducted using the AMBER16 package.[12]

Each simulation was carried out for at least 2 μs or more, with a time step of 2 fs in an isothermal-isobaric ensemble (300 K, 1 bar). In all simulations, two Aβ monomers spontaneously aggregated into a dimer within 1 μs. The 2Aβ remained stable and did not disassociate as long as it formed. In the present study, we used the last 1 μs trajectory of each simulation for analysis. Because of the intrinsically disordered nature of the protein, the 2Aβ exhibited strong flexibility in the aqueous solution. Cluster analysis was carried out using the DBSCAN (density-based spatial clustering of applications with noise) algorithm to determine the representative conformers.[13] Each simulation gave one representative conformer. The representative conformers ($E1_s$, $E2_s$, and $E3_s$) accounted for 74%, 96%, and 74% of sampled conformers in the individual simulation, respectively.

Determination of the protonation states of 2Aβ in the gaseous phase. In this work, we used a well-established hybrid Monte Carlo/molecular dynamics (MC/MD) protocol, developed by some of us, to determine the most probable protonation state for each conformer in the gas phase.[14-16] This protocol iteratively generates new random protonation states, carries out MD simulations, and searches for energy minima. The optimized protonation states are next tested employing the Metropolis algorithm. The search of new protonation states continues until a default criterion was met.

A representative 2Aβ conformer was extracted from each equilibrated aqueous MD trajectories. Protonating/deprotonating was allowed at the following sites: the side chains of residue D, E, H, K, and R, together with N- and C-termini of monomers. Previous studies demonstrated that three different force fields (AMBER99, GROMOS41a1, and OPLS/AA) should give the same most probable protonation states regardless of protein species.[17-19] Given that predicted protonation states were supposed to be force field independent, the OPLS/AA force field was chosen because it could provide a complete set of conjugated acid/base pairs for protonatable residues and termini. Eventually, the protocol gave four different protonation states for the next step.

MD simulations of 2Aβ in the gaseous phase. We conducted four independent MD simulations in the gas phase (with the absence of water molecules) starting from representative aqueous

conformers with their corresponding lowest-energy protonation states. The calculations were again based on the AMBER ff99SB-ILDN force field because it has been widely used for gas-phase simulations.[16, 20] Note that there were no longer periodic boundary conditions employed in gas phase simulations. The temperature of the protein was maintained at 300 K. Each simulation carried out for 200 μs. The last 50 μs trajectory of each simulation was used for further analysis because the conformation remained generally unchanged during such stage. Similarly, four representative conformers, $E1a_g$, $E1b_g$, $E2_g$, and $E3_g$, were determined.

We used the trajectory method to compute the collisional cross section (CCS) of the protein via the ion mobility spectrometry suite (IMoS).[21] The secondary structure of the protein was decided employing define secondary structure of protein (DSSP) algorithm.[22] The hydrophobicity index was calculated in Abraham and Leo's scale.[23] A 5 Å sidechain contact cutoff was employed in the present study.[24]

Figures

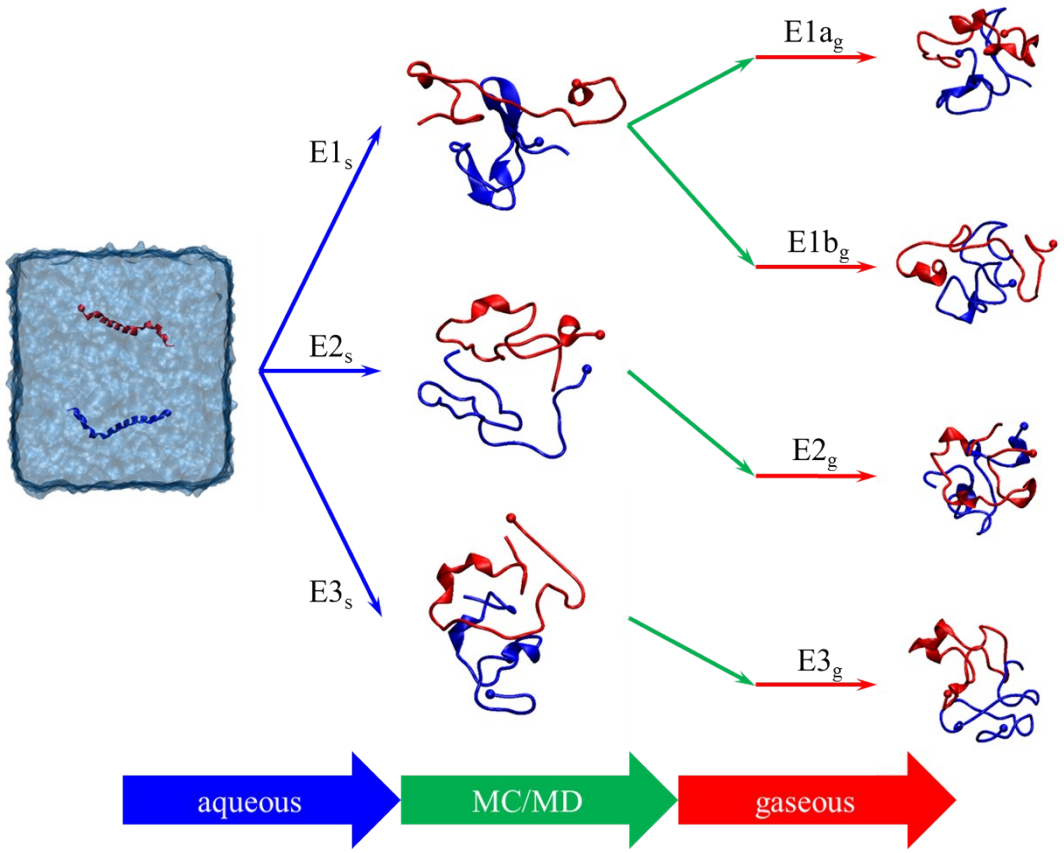

Figure S1. Schematic illustration of the simulations conducted in this study. Blue arrows are the aqueous MD simulations; green arrows are predictions of protonation states using the MC/MD protocol; and red arrows are the gaseous MD simulations. Two Aβ monomers are blue and red, respectively. The ball indicates the N-terminus of the monomer.

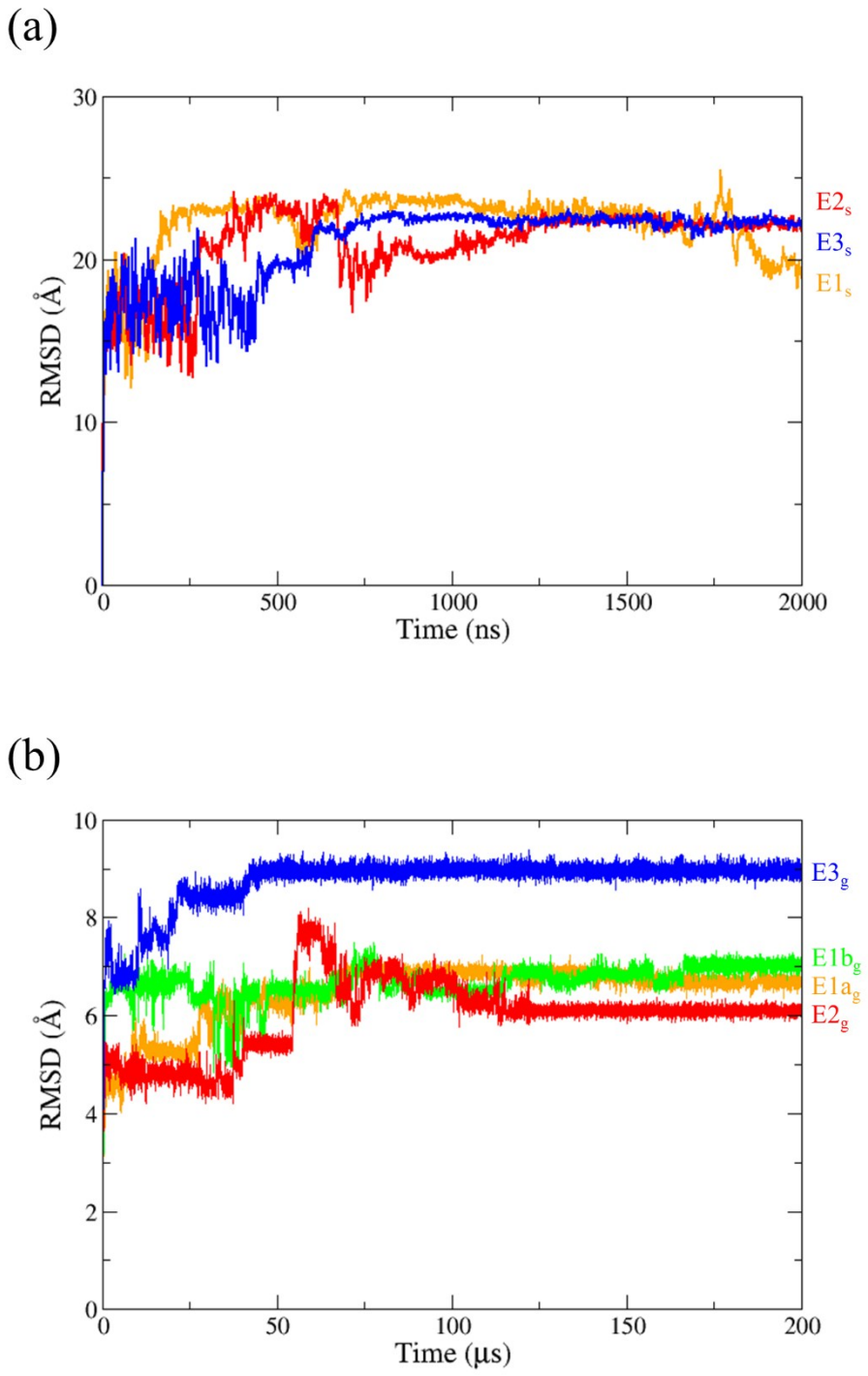

Figure S2. RMSD of 2Aβ simulations in the aqueous phase (a) and in the gas phase (b), plotted as a function of simulated time. For cluster analysis, last 1000 ns trajectories of the aqueous simulations and last 50 μs trajectories of the gaseous ones were used in this study.

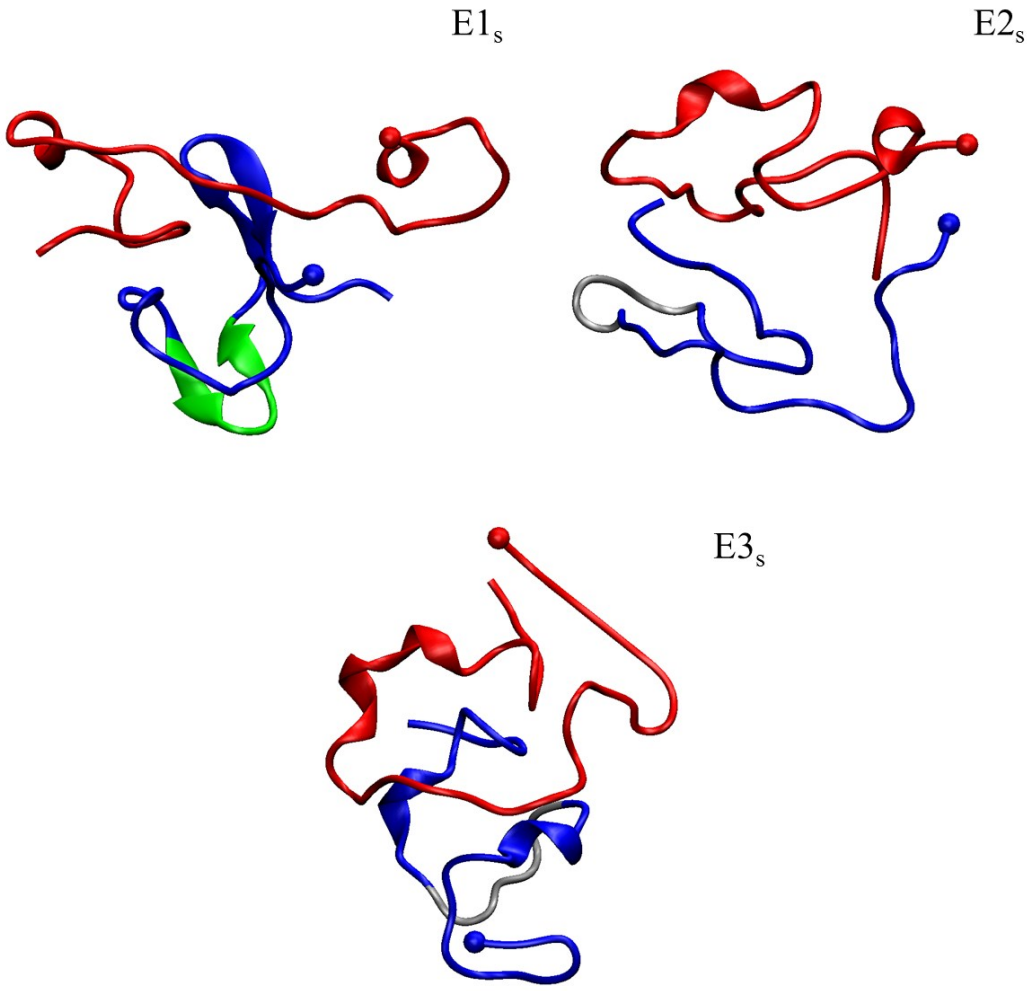

Figure S3. Selected conformers ($E1_s$, $E2_s$, and $E3_s$) obtained from MD simulation in the solution. Two Aβ monomers are colored in blue and red, respectively. The ball indicates the N-terminus of the monomer. The β hairpin motif (sequence ^{20}FAEDVG^{25}) found in the $E1_s$ conformer is colored in green. The same regions in the $E2_s$ and $E3_s$ conformers, which are random coils, are colored in gray.

Figure S4. Protonation states predicted by the MC/MD hybrid protocol. Colored letters represent protonatable sites. Blue letters are positively charged sites, red are negatively charged, and green are neutral. The letter N and C indicates, respectively, the N- and C-terminal.

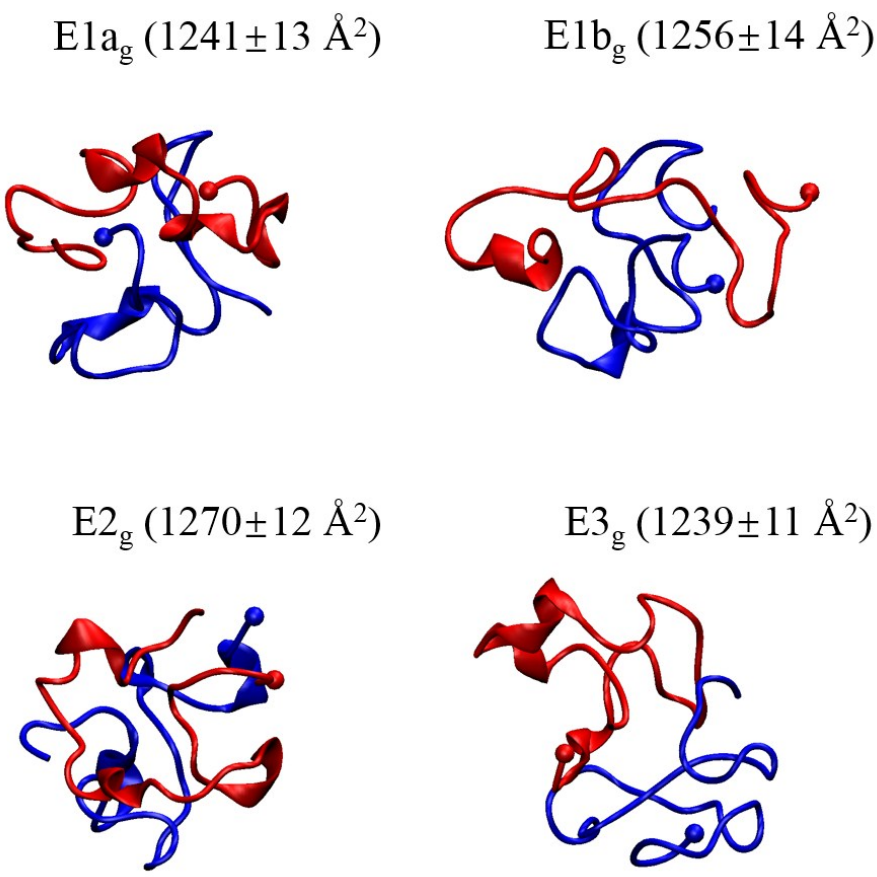

Figure S5. Selected conformers (E1a$_g$, E1b$_g$, E2$_g$, and E3$_g$) obtained from MD simulation in the vacuum. Two Aβ monomers are colored in blue and red, respectively. The ball indicates the N-terminus of the monomer.

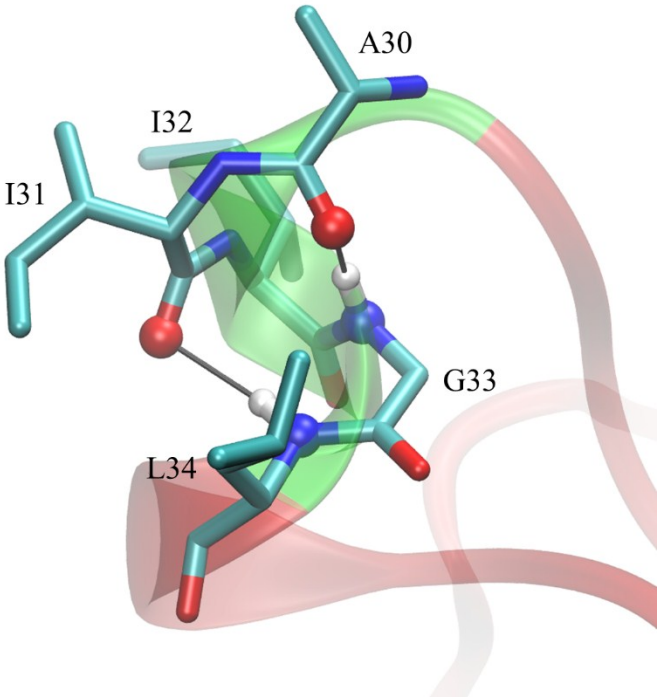

Figure S6. Selected local conformer of 2Aβ in the gas phase. The ^{30}AIIGL^{34} region is highlighted in transparent green. Carbon atoms are colored in cyan, hydrogen are white, oxygen are red, and nitrogen are blue. Hydrogen bonds are depicted as black continuous lines. Only hydrogen atoms forming these interactions are shown.

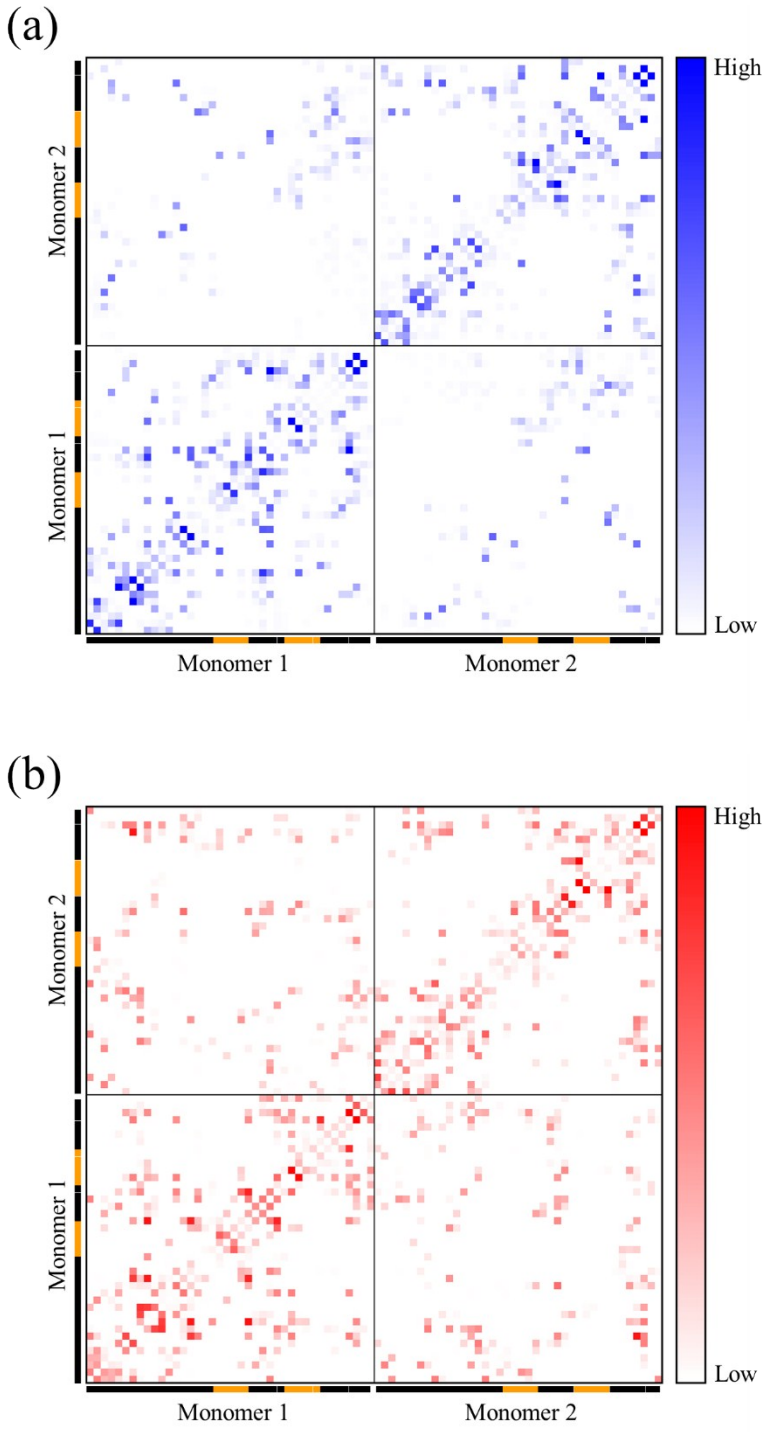

Figure S7. Sidechain contact maps for all residues of 2Aβ in the aqueous phase (a) and in the gas phase (b). The regions with high helical probability (sequence ^{19}FFAED^{23} and ^{30}AIIGL^{34}) are highlighted in orange.

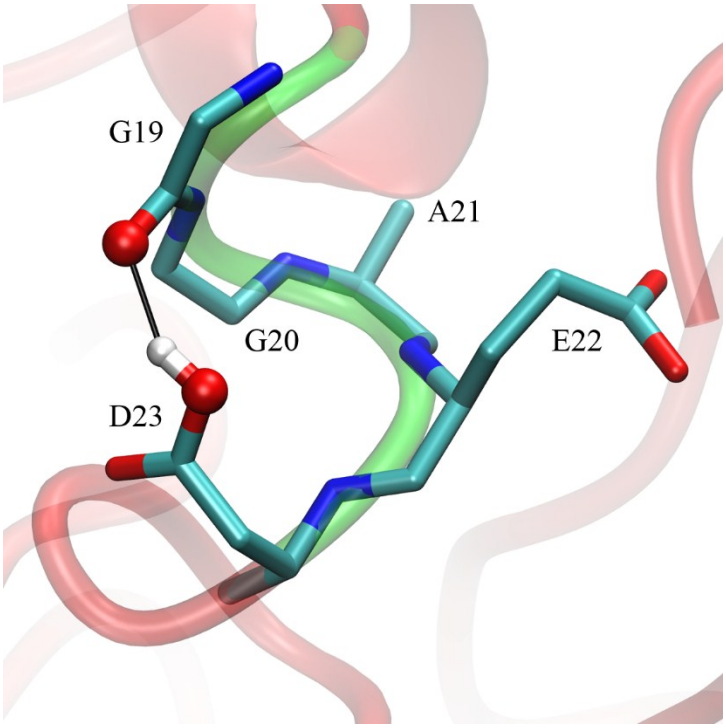

Figure S8. Selected local conformer of the FΔG mutant of 2Aβ in the gas phase. The ^{19}GGAED^{23} region is highlighted in transparent green. Carbon atoms are colored in cyan, hydrogen are white, oxygen are red, and nitrogen are blue. The hydrogen bonds is depicted as black continuous lines. Only hydrogen atom forming the interaction is shown.

Tables

Table S1. Secondary structure propensities of 2Aβ averaged from conformers in aqueous MD simulations. The numbers in the bracket are standard deviations.

	Helix (%)	β (%)	Turn (%)	Coil (%)
$E1_s$	6.8 (3.7)	16.3 (2.3)	46.3 (4.3)	30.6 (3.0)
$E2_s$	7.0 (4.0)	6.8 (3.0)	51.7 (4.7)	34.5 (3.1)
$E3_s$	7.7 (2.6)	13.6 (4.5)	47.1 (4.6)	31.6 (2.8)

Table S2. Secondary structure propensities of 2Aβ averaged from conformers in gaseous MD simulations. The numbers in the bracket are standard deviations.

	Helix (%)	β (%)	Turn (%)	Coil (%)
$E1a_g$	15.0 (5.8)	7.6 (1.8)	55.4 (5.9)	22.0 (2.5)
$E1b_g$	16.6 (3.5)	9.8 (2.4)	48.2 (4.6)	25.4 (3.5)
$E2_g$	26.9 (4.2)	0.1 (0.4)	45.5 (4.4)	27.5 (2.0)
$E3_g$	23.8 (3.3)	1.1 (1.2)	49.6 (3.7)	25.5 (0.2)

Table S3. The local helix propensity by residue for the wild-type and mutant 2Aβ.

Wild-type	F19	F20	A21	E22	D23
	0.39	0.56	0.59	0.58	0.33
FΔA	A19	A20	A21	E22	D23
	0.47	0.52	0.55	0.52	0.14
FΔG	G19	G20	A21	E22	D23
	0.24	0.26	0.34	0.33	0.10

References

1. H. Sticht, P. Bayer, D. Willbold, S. Dames, C. Hilbich, K. Beyreuther, R. W. Frank and P. Rosch, *Eur J Biochem*, 1995, **233**, 293.
2. S. H. Chong and S. Ham, *Proc Natl Acad Sci USA*, 2012, **109**, 7636.
3. J. C. Gordon, J. B. Myers, T. Folta, V. Shoja, L. S. Heath and A. Onufriev, *Nucleic Acids Res*, 2005, **33**, W368.
4. V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg and C. Simmerling, *Proteins*, 2006, **65**, 712.
5. K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror and D. E. Shaw, *Proteins*, 2010, **78**, 1950.
6. W. L. Jorgensen, J. Chandrasekhar and J. D. Madura, *J Chem Phys*, 1983, **79**, 926.
7. T. Darden, D. York and L. Pedersen, *J Chem Phys*, 1993, **98**, 10089.
8. J.-P. Ryckaert, G. Ciccotti and H. Berendsen, *J Comput Phys*, 1977, **23**, 327.
9. S. Nose, *Mol Phys*, 1984, **52**, 255.
10. W. G. Hoover, *Phys Rev A*, 1985, **31**, 1695.
11. H. J. C. Berendsen, J. P. M. Postma, W. F. Vangunsteren, A. Dinola and J. R. Haak, *J Chem Phys*, 1984, **81**, 3684.
12. D.A. Case, R.M. Betz, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, C. Lin, T. Luchko, R. Luo, B. Madej, D. Mermelstein, K.M. Merz, G. Monard, H. Nguyen, H.T. Nguyen, I. Omelyan, A. Onufriev, D.R. Roe, A. Roitberg, C. Sagui, C.L. Simmerling, W.M. Botello-Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu, L. Xiao and P.A. Kollman (2016), AMBER 2016, University of California, San Francisco.
13. M. Ester, H.-P. Kriegel, J. Sander and X. Xu, *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)* 1996, 226.
14. R. Marchese, R. Grandori, P. Carloni and S. Raugei, *J Am Soc Mass Spectr*, 2012, **23**, 1903.
15. J. Y. Li, G. Rossetti, J. Dreyer, S. Raugei, E. Ippoliti, B. Luscher and P. Carloni, *Plos Comput Biol*, 2014, **10**, e1003838.
16. J. Y. Li, W. P. Lyu, G. Rossetti, A. Konijnenberg, A. Natalello, E. Ippoliti, M. Orozco, F. Sobott, R. Grandori and P. Carloni, *J Phys Chem Lett*, 2017, **8**, 1105.
17. J. M. Wang, P. Cieplak and P. A. Kollman, *J Comput Chem*, 2000, **21**, 1049.
18. W. R. P. Scott, P. H. Hunenberger, I. G. Tironi, A. E. Mark, S. R. Billeter, J. Fennen, A. E. Torda, T. Huber, P. Kruger and W. F. van Gunsteren, *J Phys Chem A*, 1999, **103**, 3596.
19. W. L. Jorgensen and J. Tirado-Rives, *Proc Natl Acad Sci USA*, 2005, **102**, 6665.
20. Y. Berezovskaya, M. Porrini and P. E. Barran, *Int J Mass Spectrom*, 2013, **345**, 8.
21. C. Larriba and C. J. Hogan, *J Comput Phys*, 2013, **251**, 344.
22. J. Zacharias and E. W. Knapp, *J Chem Inf Model*, 2014, **54**, 2166.
23. D. J. Abraham and A. J. Leo, *Proteins*, 1987, **2**, 130.
24. J. S. Viloria, M. F. Allega, M. Lambrughi and E. Papaleo, *Sci Rep*, 2017, **7,** 2838.