Electronic Supplementary Material (ESI) for Chemical Communications


# Supporting Information

# for

# Early amyloid β-protein aggregation precedes conformational change

**Bogdan Barz,[a] Olujide O. Olubiyi,[a] and Birgit Strodel[a,b,*]**

[a] *Forschungszentrum Jülich GmbH, Institute of Complex Systems: Structural Biochemistry (ICS-6), 52425 Jülich, Germany*
[b] *Institute of Theoretical and Computational Chemistry, Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany*

[*] Corresponding author: b.strodel@fz-juelich.de

# Materials and Methods

## 1. Computational details

We performed five independent all-atom molecular dynamics (MD) simulations of 20 Aβ42 monomers inserted in a cubic box with side length 35 nm resulting in a solute concentration of ∼0.8 mM. The initial Aβ configurations were selected from previous MD simulations of Aβ monomer in water[1] and solution NMR structures (PDB code 1Z0Q) in order to avoid bias towards a particular secondary structure. Each simulation was started with different initial velocities and, after equilibration, comprised a 200 ns production run resulting in a total of 1 $\mu$s simulation time. To investigate the impact of amino acids I41–A42 on the aggregation process we performed an additional aggregation simulation of 20 Aβ40 peptides for 200 ns and compared it with a similar simulation of Aβ42.

All MD simulations were performed with the Gromacs 4.5.5 parallel software package[2] using the all-atom OPLS/AA force field[3,4] and the GB/SA implicit solvent.[5] While the OPLS/AA force field might not be the best choice for folded proteins,[6,7] the situation is different for unfolded proteins.[1] The OPLS/AA-generated conformations for the Aβ40 and Aβ42 monomer best match experimental data.[8,9] A study by Sgourakis et al. using the combination of the OPLS-AA force field and TIP3P water model identified distinct Aβ40 versus Aβ42 structures consistent with NMR data.[8] More recently, Lemkul et al.[9] suggested OPLS/AA with the TIP4P water model as superior to AMBER03 or CHARMM22/CMAP for studying Aβ40, producing very similar results as GROMOS96 53A6 and GROMOS96 54A7 in terms of helical and β-strand content, calculated NMR shifts, and radii of gyration that agree well with experimental data. When studying the effect of different force fields on peptide aggregation, Nguyen et al.[10] showed that OPLS/AA explored the most diverse conformations for the aggregation of the $A\beta_{16-22}$ fragment in explicit water.

While MD simulations in explicit solvent are generally more accurate[7] and preferred in the case of small systems, for large systems as in this study they can be computationally overwhelming. Thus, in the current study we use the GB/SA implicit solvent.[5] Previous studies of full-length Aβ aggregation involving more than two monomers have not only been studied in implicit solvent, but also with a coarse-grained protein model.[11] Results from this study are discussed and compared to our findings in the main text. Due to the computational requirements for explicit solvent simulations, the use of an atomistic model with an implicit solvent model can thus be considered as an important step forward towards more detailed simulations of large-scale Aβ aggregation. Here, a legitimate question is whether the aggregation pathway and oligomer structures will reflect the same characteristics as one would observe in explicit solvent. Based on previous simulations[13,14] and the comparison of our Aβ oligomer structures to experimental observations we are confident that this question can be positively answered. The small changes in the secondary structure during Aβ aggregation observed in our study are supported by both experiment[12] and simulations using explicit solvent.[13] The replica exchange molecular dynamics (REMD) simulation of Aβ42 dimer formation using the OPLS/AA force field and the SPC/E water model produced collision cross sections for dimers in agreement with experimental values.[13] The dimer conformation is rather unstructured with only small amounts of β-sheet (∼8%), which is very similar to our observations. It has to be noted that this REMD simulation was started from completely helical Aβ42 and involved 64 200 ns replicas with temperatures ranging from 315 to 450 K. I.e., even though an enhanced sampling of the conformational space in explicit solvent was performed, Zhan et al. found very similar Aβ42 dimer structures as we do in our implicit solvent MD study. A study by Kent et al.[14] benchmarking implicit solvent simulation of the $A\beta_{10-35}$ fragment revealed a good

agreement between simulation and experimental results for OPLS/AA in combination with the GB/SA implicit solvent model. Especially the diffusion constant of $A\beta_{10-35}$ was best reproduced by OPLS/AA with GB/SA compared to other force fields, including CHARMM22 with TIP3P. These findings lend further support to our conclusion that our choice of force field and solvent model have at most minor influence on our results for the aggregation of Aβ42.

During the MD simulations the system was maintained at 300 K using the leap-frog stochastic dynamics integrator with a time step of 4 fs and a time constant for temperature coupling of 2 ps. Electrostatic and van der Waals interactions were cut off at 1.2 nm.

## 2. Transition network analysis

To derive transition networks we first defined the aggregation state as a number with three digits, $N1|N2|N3$, where each digit corresponds to a structural feature or to an oligomeric state. $N1$ represents the oligomeric size, identified using a cutoff distance of 0.5 nm between any two atoms belonging to different peptides. $N2$ is the average number of hydrogen-bonds between individual chains from the oligomer. Hydrogen-bonds are defined based on distance and angle cutoffs of 0.35 nm and 30°, respectively. $N3$ is the average number of amino acids in β-strand conformation per peptide in the oligomer. An amino acid is defined to be in a β-strand conformation if the dihedral angles ϕ and ψ of the backbone are contained in the polygon with vertices $(-180,180)$, $(-180,126)$, $(-162,126)$, $(-162,108)$, $(-144,108)$, $(-144,90)$, $(-50,90)$, $(-50,180)$.[15] To calculate the transition matrix that includes all pairwise transitions between aggregation states we first identified all the aggregation states and the number of transitions between states along the 1 $\mu$s trajectory using a lag time of 20 ps. Using these states and transitions we built an $N \times N$ matrix, where $N$ is the number of states encountered. Each element of the matrix represents the population of a particular transition between two states. From the transition matrix we have derived a new matrix that preserves the maximum flow using the minimum-cut algorithm.[16–19] The maximum flow transition matrix was converted into a transition network using the software Visone[20] and the minimum stress algorithm in combination with the link routing procedure. In the transition network plots, the nodes represent aggregation states, the area of each node is proportional to the population of the state, and the color of the node indicates the oligomer size ($N1$). The thickness of network edges corresponds to the number of transitions between two states.

## 3. Structural analysis

Inter-molecular contact maps were calculated for pair-wise amino acids (Cα atoms) of different proteins using a 0.75 nm cutoff. All oligomer interfaces were considered from frames corresponding to every 10 ns during the last 100 ns of the 5 simulations. The contact map containing the difference between the Aβ42 and Aβ40 contact probabilities from Fig. S2 was based on one MD simulation of 200 ns for each alloform.

# References

[1] O. O. Olubiyi and B. Strodel, *J. Phys. Chem. B*, 2012, **116**, 3280–3291.

[2] B. Hess, C. Kutzner, D. van der Spoel and E. Lindahl, *J. Chem. Theory Comput.*, 2008, **4**, 435–447.

[3] W. L. Jorgensen, D. S. Maxwell and J. Tirado-Rives, *J. Am. Chem. Soc.*, 1996, **118**, 11225–11236.

[4] G. A. Kaminski, R. A. Friesner, J. Tirado-Rives and W. L. Jorgensen, *J. Phys. Chem. B*, 2001, **105**, 6474–6487.

[5] D. Qiu, P. S. Shenkin, F. P. Hollinger and W. C. Still, *J. Phys. Chem. A*, 1997, **101**, 3005–3014.

[6] D. Matthes and B. L. d. Groot, *Biophys. J.*, 2009, **97**, 599–608.

[7] K. A. Beauchamp, Y.-S. Lin, R. Das and V. S. Pande, *J. Chem. Theory Comp.*, 2012, **8**, 1409–1414.

[8] N. G. Sgourakis, Y. Yan, S. A. McCallum, C. Wang and A. E. Garcia, *J. Mol. Biol.*, 2007, **368**, 1448–1457.

[9] S. R. Gerben, J. A. Lemkul, A. M. Brown and D. R. Bevan, *J. Biomol. Struct. Dyn.*, 2013, **0**, 1–16.

[10] P. H. Nguyen, M. S. Li and P. Derreumaux, *Phys. Chem. Chem Phys.*, 2011, **13**, 9778–9788.

[11] B. Urbanc, M. Betnel, L. Cruz, G. Bitan and D. B. Teplow, *J. Am. Chem. Soc.*, 2010, **132**, 4266–4280.

[12] M. Ahmed, J. Davis, D. Aucoin, T. Sato, S. Ahuja, S. Aimoto, J. I. Elliott, W. E. Van Nostrand and S. O. Smith, *Nat. Struct. Mol. Biol.*, 2010, **17**, 561–567.

[13] T. Zhang, J. Zhang, P. Derreumaux and Y. Mu, *J. Phys. Chem. B*, 2013, **117**, 3993–4002.

[14] A. Kent, A. K. Jha, J. E. Fitzgerald and K. F. Freed, *J. Phys. Chem. B*, 2008, **112**, 6175–6186.

[15] D. K. Klimov and D. Thirumalai, *Structure*, 2003, **11**, 295–307.

[16] L. R. Ford and D. R. Fulkerson, *Canad. J. Math.*, 1956, **8**, 399–404.

[17] S. V. Krivov and M. Karplus, *J. Chem. Phys.*, 2002, **117**, 10894–10903.

[18] S. V. Krivov and M. Karplus, *Proc. Natl. Acad. Sci. USA*, 2004, **101**, 14766–14770.

[19] R. E. Gomory and T. C. Hu, *J. Soc. Ind. Appl. Math.*, 1961, **9**, 551–570.

[20] U. Brandes and D. Wagner, *Graph Drawing Software*, 2003, p. 321340.

# Figures

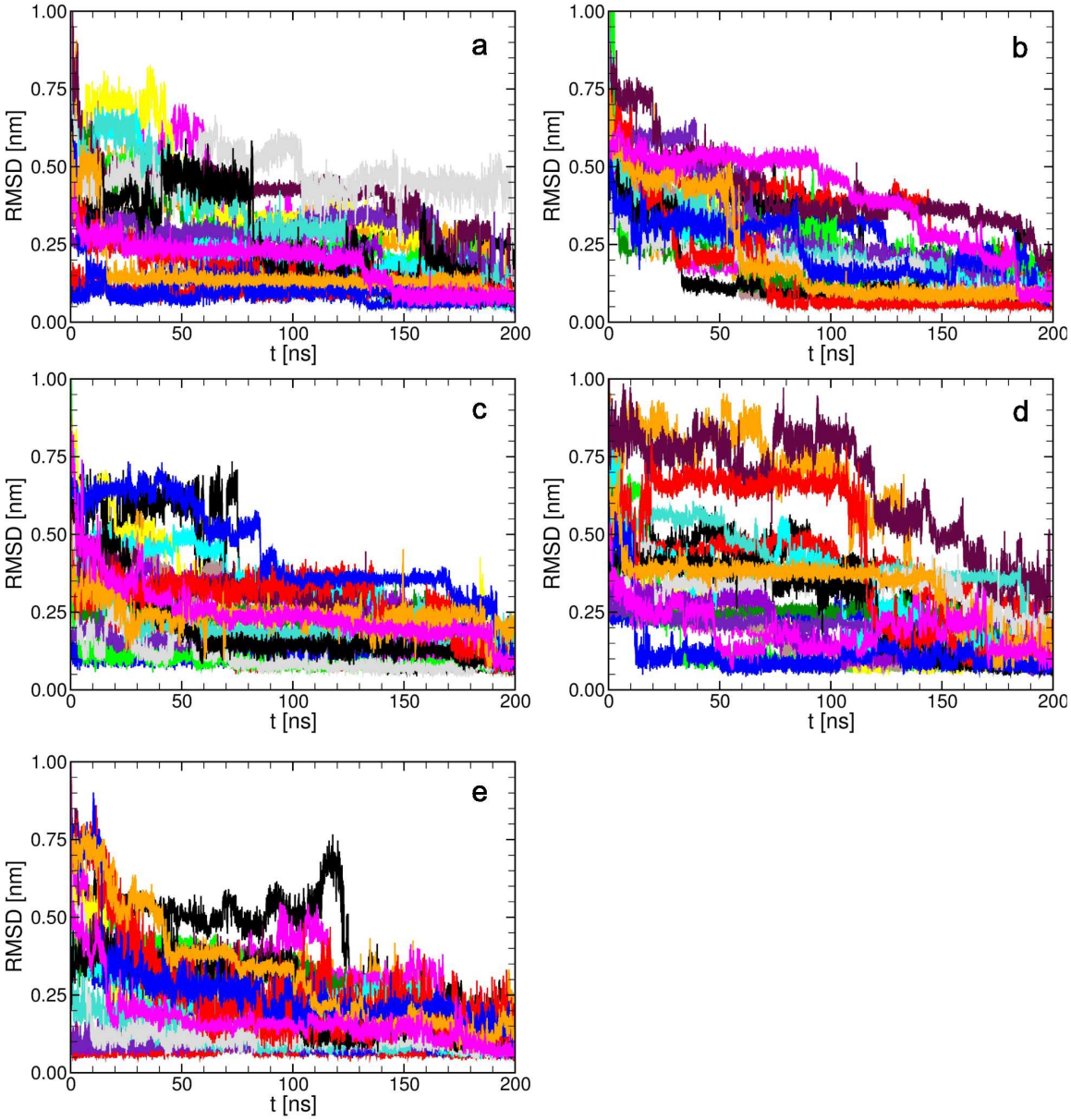


**Figure S1** RMSD of individual monomers (carbon alpha atoms) with respect to their final structures during each of the five trajectories. a), b), c), d) and e) correspond to trajectories 1, 2, 3, 4, and 5, respectively. Different colors correspond to different monomers.

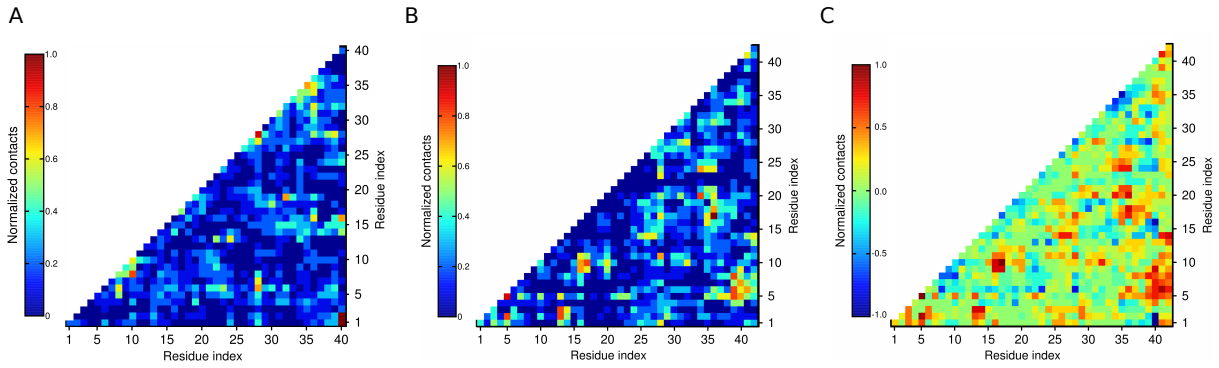


**Figure S1** Inter-molecular contact maps for (a) Aβ40 and (b) Aβ42 during the last 100 ns of 200 ns simulations per peptide. Color coding corresponds to the normalized number of contacts. (c) Map showing the difference between inter-molecular contacts of Aβ42 and Aβ40. Color coding corresponds to the difference (Aβ42−Aβ40) in the normalized number of contacts of the two alloforms, i.e., positive numbers reflect increased contacts in Aβ42 relative to Aβ40 while for contacts with negative numbers the interactions are more pronounced in Aβ40.