Multiscale simulations of polypeptide aggregation

Multilevel coarse graining of multiscale problems

A. Guljas, F. Musil, C. Clementi*, Department of Physics, Freie Universität Berlin*

In Short

- The process of amyloid fibrillation, i.e., the formation of long and ordered aggregates from peptides, is associated with several diseases, ranging from diabetes to Alzheimer's and Parkinson's.
- Amyloid formation is a multiscale process, with timescales ranging from nanoseconds for the monomer conformational changes to hours for the formation of experimentally observable fibrils.
- We develop accurate and efficient coarse-grained (CG) models that allow us to simulate the aggregation process over biologically relevant timescales.
- We combine our results with experimental analysis for the complete characterization of the fibrillation of the 37-residue human islet amyloid polypeptide (IAPP), related to type 2 diabetes.

Amyloid fibrils play a central role in many diseases such as Alzheimer's, Parkinson's, or Diabetes. In the last couple of decades, amyloid fibrillation has been intensely studied both experimentally and computationally. Despite recent advances, the understanding of this process is still lacking. For instance, it is not clear whether aggregation and amyloidosis are restricted to specific proteins or represent a generic protein state, which can be triggered by physiochemical and cellular conditions. Answers to these questions could, on one side, boost our ability to identify and possibly disrupt disease-related amyloidogenesis, or, on the other side, allow us to engineer controlled aggregation for the design of "smart" biomaterials.

The main challenges in the complete characterization of amyloid formation are presented by the gap in time- and length-scale between the experimental and simulation studies. Amyloid fibrillation is an intrinsically multiscale process, spanning from the fast ($\sim$ ns) and localized dynamics of the disordered peptides to the order/disorder transition of the monomers, and the slow ($\sim$ hr) aggregation of multiple peptides in the formation of the fibril (see Figure 1C). Experiments such as Fluorescence resonance energy transfer (FRET), can time-resolve the kinetics of donor-acceptor distances (between 2-9 nm) at $\sim$10 μs resolution. While atomistic molecular dynamics (MD) is extremely valuable in elucidating nano- to micro-second processes in sub-nm detail, the simulation of fiber formation remains inaccessible even on the most powerful special-purpose hardware.

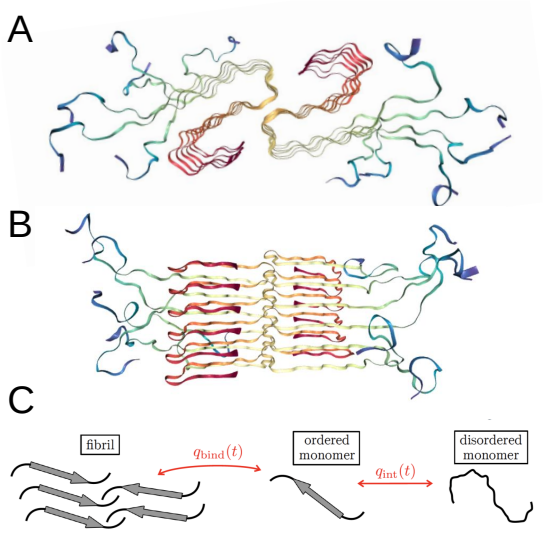

Figure 1: *Snapshots from simulations of a periodically repeated fibril of the human islet amyloid polypeptide (IAPP), comprised of 12 IAPP. (A) View orthogonal to the stacking direction. (B) View along the stacking direction. The ordered core of the fibril is easily distinguished from the disordered corona. (C) The docking of a peptide to an already existing fibril is in the simplest scenario modeled by two reaction coordinates $q_{int}(t)$ and $q_{bind}(t)$, which describe the intra-peptide and the inter-peptide ordering, respectively.*

In this project, we investigate the fibrillation of the Human Islet Amyloid Polypeptide (IAPP) with a combination of theoretical/computational/experimental techniques. In particular, the development of efficient and reliable coarse-grained (CG) models of the fibrillation process is key to simulating timescales longer than ms, thus bridging the gap between experiment and atomistic simulation, for a complete and multiscale characterization of IAPP amyloidogenesis.

Recently the IAPP amyloid structure was determined by cryo-electron microscopy 1. Our collaborators (R. Netz and C. Schütte) have performed and analyzed initial short (ns) atomistic MD simulations of a periodically repeated fibril and confirmed that it forms a stable fibril core in simulation (see Figure 1A-B) 2. Our experimental collaborators (B. Koksch) have established IAPP analogs containing FRET labels at different positions and are able to study the amyloid aggregation kinetics at different positions along the sequence. In addition, they have

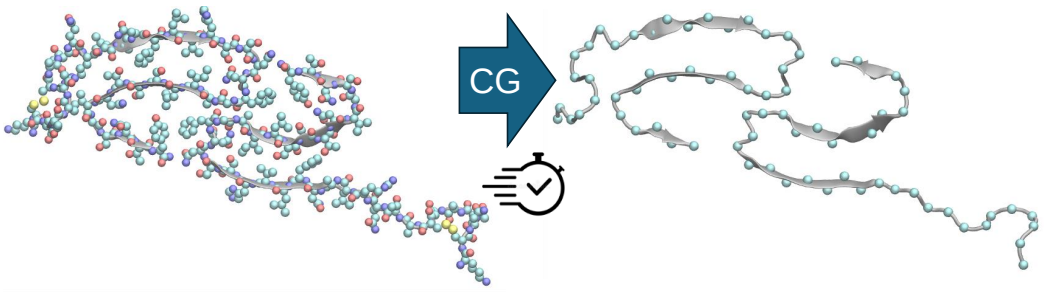

Figure 2: *The coarse-graining of the peptides from atomistic (left) to C_α (right) resolution allows us to simulate the aggregation process over biologically relevant timescales.*

access to Transmission Electron Cryomicroscopy, which allows visualization of IAPP fibril morphology and pre-fibrillar species.

In the past few years, we have shown that machine-learned CG models can reproduce quite accurately the thermodynamics of single proteins in solution as obtained from reference atomistic models 3. In a different project (supported by DFG SFB 1114-A04) we are building on our successful CGNet 4 and CGSchnet 5 to design CG models transferable in sequence space. In the present project, we take a complementary direction and characterize the dynamics of the aggregation of the IAPP amyloid fibrils, over long time- and length-scales. Instead of designing a transferable CG model able to simulate an arbitrary single protein in solution, here we focus on a specific *multi-protein* system, composed of multiple copies of the same sequence (IAPP) aggregating into fibrils (see Figure 2). We have performed many ($\sim$ 100) relatively short ($\sim$ 10–100 ns) atomistic simulations for systems composed of different copies of the IAPP sequence, from 1 to 12, using different substructures of the experimentally determined amyloid assembly 1. These simulations are added to our atomistic dataset of diverse short sequences to "learn" long-range interactions, that are essential for protein-protein association. Instead of transferability in sequence space here we seek "structural transferability" of the CG IAPP multi-protein model. We train with up to 12 copies of IAPP but want to use the model to predict the aggregation of many more copies of this protein, into longer fiber-like structures. The training of these models as well as the long CG simulations to observe the formation of the fibers require significant computational resources and is only possible on high-performance computing clusters. In particular, the training and simulation of large neural network models require the high-capacity GPUs of HLRN. The reduced dimensionality of the CG model and the renormalization of the fastest atomic vibrations allow to simulate the aggregation process for different numbers of peptides and thermodynamic conditions, that can be experimentally verified. We believe that the results of this project allow us to address open questions in the formation of amyloids, such as the role of hydrophobic interactions during aggregation.

WWW

https://www.mi.fu-berlin.de/en/sfb1114/reasearch/projects/b03

More Information

[1] C. Röder, T. Kupreichyk, L. Gremer, L. U. Schäfer, K. R. Pothula, R. B. Ravelli, D. Willbold, W. Hoyer, and G. F. Schröder. *Nat. Struct. Mol. Biol.* **27**(7), 660–667 (2020). doi: 10.1038/s41594-020-0442-4

[2] A. Bittracher, J. Moschner, B. Koksch, R. Netz, and C. Schütte. *Eur. Phys. J. B*, **94** 195 (2021). doi:10.1140/epjb/s10051-021-00200-0

[3] A. E. P. Durumeric, N. E. Charron, C. Templeton, F. Musil, K. Bonneau, A. S. Pasos-Trejo, Y. Chen, A. Kelkar, F. Noé, and C. Clementi. *Curr. Opin. Struct. Biol.*, **79**, 102533, (2023). doi: 10.1016/j.sbi.2023.102533

[4] J. Wang, S. Olsson, C. Wehmeyer, A. Pérez, N. E. Charron, G. de Fabritiis, F. Noé, and C. Clementi. *ACS Cent. Sci.* **5**(5), 755–767 (2019). doi:10.1021/acscentsci.8b00913

[5] B. E. Husic, N. E. Charron, D. Lemm, J. Wang, A. Pérez, A. Krämer, Y. Chen, S. Olsson, G. de Fabritiis, F. Noé, and C. Clementi. *J. Chem. Phys.*, **153**, 194101, (2020). doi: 10.1063/5.0026133

Project Partners

Christof Schütte, ZIB; Roland Netz, Freie Universität Berlin; Beate Koksch, Freie Universität Berlin

Funding

DFG SFB 1114 Project B03

DFG Subject Area

201-02 Biophysics