

Determination of the structure and dynamics of the fuzzy coat of an amyloid fibril of IAPP using cryo-electron microscopy

Z. Faidon Brotzakis^{1,‡}, Thomas Löhr^{1,‡}, Steven Truong¹, Samuel E. Hoff²,
Massimiliano Bonomi² and Michele Vendruscolo^{1,*}

¹*Centre for Misfolding Diseases, Department of Chemistry,
University of Cambridge, Cambridge CB2 1EW, UK*

²*Institut Pasteur, Université Paris Cité, CNRS UMR 3528,
Department of Structural Biology and Chemistry, 75015 Paris, France*

[‡] *Authors have contributed equally*

* *Corresponding author: mv245@cam.ac.uk*

ABSTRACT

In recent years, major advances in cryo-electron microscopy (cryo-EM) have enabled the routine determination of complex biomolecular structures at atomic resolution. An open challenge for this approach, however, concerns large systems that exhibit continuous dynamics. To address this problem, we developed the metadynamic electron-microscopy metainference (MEMMI) method, which incorporates metadynamics, an enhanced conformational sampling approach, into the metainference method of integrative structural biology. MEMMI enables the simultaneous determination of the structure and dynamics of large heterogeneous systems by combining cryo-EM density maps with prior information through molecular dynamics, while at the same time modelling the different sources of error. To illustrate the method, we apply it to elucidate the dynamics of an amyloid fibril of the islet amyloid polypeptide (IAPP). The resulting conformational ensemble provides an accurate description of the structural variability of the disordered region of the amyloid fibril, known as fuzzy coat. The conformational ensemble also reveals that in nearly half of the structural core of this amyloid fibril the side-chains exhibit liquid-like dynamics despite the presence of the highly ordered network backbone of hydrogen bonds characteristic of the cross- β structure of amyloid fibrils.

Introduction

In the last several years, cryo-electron microscopy (cryo-EM) has been pushing the boundaries of structural biology in terms of structure resolution, system complexity, and macromolecular size^{1,2}. Imaging single particles by rapid cryo-cooling and vitrification^{3,4} enables structural studies under near-native environmental conditions while offering sample protection from beam radiation. Technical advances in electron detectors, computational algorithms accounting for beam-induced motion, and automation of data collection and image analysis⁵ have paved the way to a spectacular increase of high-resolution cryo-EM density maps. The Electron Microscopy Data Bank (EMDB) currently holds 15,800 single-particle cryo-EM density maps, which offer exquisitely detailed structural information about macromolecular systems of central importance in cell biology.

In standard cryo-EM structure determination, two-dimensional (2D) images of single particles are first classified in conformationally homogenous classes and then averaged in a computational image processing step, thereby leading to a substantial increase in the signal-to-noise ratio, and thus in structure resolution^{6,7}. However, the continuous dynamics of flexible regions are difficult to detect, therefore complicating the generation of homogenous classes of structures. The resulting low densities cannot be readily used to determine atomistic structures, and are thus often excluded from the final structural model. While methods such as the manifold embedding approach⁸ can determine multiple structures from cryo-EM density maps, to account for the conformational dynamics one should quantitatively and atomistically interpret the cryo-EM density maps as an envelope that corresponds to an averaged conformational ensemble of states with certain populations that interconvert with a characteristic timescale⁹. Such a viewpoint moves away from a single structure interpretation of the data and links the data to the statistical mechanics concept of free energy landscapes of conformational ensembles^{10,11}. Integrative structural ensemble-modelling methods incorporate experimental information into molecular simulations and enable the determination of structural ensembles that maximally conform to the experimental data with atomistic resolution¹²⁻²³. This technique has been applied using nuclear magnetic resonance (NMR) spectroscopy^{17,19,20,24-28}, fluorescence resonance energy transfer (FRET) microscopy²⁹, small-angle scattering techniques (SAXS/SANS)³⁰⁻³⁴, transition rate constants^{22,23} and cryo-EM data³⁵.

One of such methods, cryo-EM metainference (EMMI)³⁵, can accurately model a thermodynamic ensemble by combining prior information on the system, such as physico-chemical knowledge (e.g. a force field), with noisy (i.e. subject to systematic and random errors) and heterogeneous (i.e. encoding a conformational ensemble) experimental data, using cryo-EM density maps. EMMI has already been used in a series of complex macromolecular systems, including a CLP protease³⁶, microtubules³⁷, microtubule-tau complexes³⁸, ASCT2 transporter³⁹, and SARS-CoV-2 membrane receptor proteins⁴⁰, allowing access to the continuous dynamics of biomolecules with atomistic resolution. The quality of the EMMI structural ensembles, however, is closely related to the exhaustiveness of the conformational sampling, which requires a computational time that scales exponentially with the barriers delimiting individual structural states⁹.

To tackle this rare event problem, several enhanced sampling methods have been developed. Enhanced sampling molecular simulation methods can be classified as trajectory based⁴¹⁻⁴⁶ and collective-variable (CV) based⁴⁷⁻⁵¹. For detailed reviews, readers can refer to the recent literature^{52,53}. A particularly powerful CV-based enhanced sampling method, which is very efficient once appropriate CVs are chosen, is metadynamics^{51,54}. Metadynamics adds a history-dependent bias to the system as a function of microscopic degrees of freedom of the system known as collective variables. With this bias, the simulations can escape deep free energy minima and sample transitions between different states. The choice of the CVs is critical to achieve the desired speed-up of convergence⁵². Recent developments in identifying and automating the search for appropriate CVs have increased the efficiency of this method, thus providing a remedy to the conformational sampling problem⁵⁵⁻⁵⁹. While metainference has been combined already with metadynamics⁶⁰, this was not the case for EMMI (**Figure 1A**). Combining EMMI with enhanced sampling methods can lead to accurate and efficient determination of structural ensembles using the vast dataset of cryo-EM database, which can in turn provide atomistic insight to a range of biomolecular systems and processes.

In this work, we present the MEMMI method, which incorporates metadynamics in EMMI in order to accelerate the ability of EMMI to sample structural ensembles with slowly interconverting states. We illustrate the application of this approach to determine the structural ensemble of an amyloid fibril formed by the full-length (residues 1-37) islet amyloid polypeptide (IAPP), an aberrant assembly associated with the degeneration of pancreatic β -cells in type-2 diabetes (T2D). When functioning correctly, IAPP, together with insulin, contributes to glycaemic control. IAPP and insulin are synthesized and stored together in pancreatic β -cells, but when IAPP aggregates in the extracellular space of the islets of Langerhans, amyloid-induced apoptosis of β -cells may occur⁶¹. In 95% of T2D patients, IAPP is found as extracellular amyloid deposits⁶²⁻⁶⁴, which form through surface-catalysed secondary nucleation⁶⁵. IAPP fibrils represent a challenging system for structural biology studies, since no unique structure can be resolved in the low-density regions of the 12-residue long N-terminal tails, due to conformational heterogeneity and associated errors in the measurement. For this reason, although recent cryo-EM experiments determined various amyloid fibril structures of IAPP⁶⁶⁻⁶⁸, the structural heterogeneity in the disordered flanking regions, known as fuzzy coat, has so far proved impossible to resolve accurately. It would be desirable to acquire a better understanding of the conformational properties of the fuzzy coat, since this region is thought to play a central role in the interactions of amyloid fibrils with other cellular components such as RNA molecules and molecular chaperones^{69,70}. Moreover, the fuzzy coat is likely to be involved in cell membrane binding, potentially promoting the catalysis of aggregation and capturing amyloid precursors^{69,70}. Recent studies on the tau protein, which is implicated in numerous neurodegenerative diseases known as tauopathies, show that depending on pH conditions, the thick fuzzy coat changes the fibril properties, including mechanically stiffness, and repulsive and adhesive behaviours^{69,70}. Here, we detail the continuous dynamics of the fuzzy coat of an IAPP fibril and utilise a thermodynamical theory of melting to characterize the different regions of the fibril to gain insight into its mechanical properties.

The MEMMI method

The cryo-EM forward model. A cryo-EM density map resulting from class-averaging and 3D reconstruction is typically encoded as voxels on a grid, and the map is generally distributed in this form. For computational efficiency, and to enable differentiability and analysis of correlations between data points, the map can be converted to a Gaussian mixture model $\phi_D(\mathbf{x})$ (GMM) consisting of N_D Gaussian components

$$\phi_D(\mathbf{x}) = \sum_{i=1}^{N_D} \phi_{D,i}(\mathbf{x}) = \sum_{i=1}^{N_D} \omega_{D,i} G(\mathbf{x} | \mathbf{x}_{D,i}, \boldsymbol{\Sigma}_{D,i}) \quad (1)$$

where \mathbf{x} is a vector in Cartesian space, $\omega_{D,i}$ is the scaling factor of the i -th component of the data GMM and G is a normalized Gaussian function centered at $\mathbf{x}_{D,i}$ with covariance matrix $\boldsymbol{\Sigma}_{D,i}$. The agreement between models generated by molecular dynamics (MD) and the data GMM is calculated by the following overlap function $ov_{MD,i}$

$$ov_{MD,i} = \int \phi_M(\mathbf{x}) \phi_{D,i}(\mathbf{x}) d\mathbf{x} \quad (2)$$

where $\phi_M(\mathbf{x})$ corresponds to the model GMM obtained from molecular dynamics. To deal with the heterogeneity of the system, EMMI simulates many replicas, r , of the system. The overlap between model GMM and data GMM is estimated over the ensemble of replicas as an average overlap per GMM component $\overline{ov}_{MD,i}$. This forward model overlap can then be compared to the data GMM self-overlap $ov_{DD,i} = \int \phi_D(\mathbf{x}) \phi_{D,i}(\mathbf{x}) d\mathbf{x}$

Metainference. Metainference is a Bayesian approach for modelling statistical ensembles by combining prior information on a system with experimental data subject to noise or systematic errors¹⁷. This framework is particularly well suited to structural ensemble determination through molecular dynamics simulations, in which the *prior* (i.e. the force field) is updated with information from experimental methods, such as NMR spectroscopy, SAXS, or cryo-EM data. Metainference is designed to handle systematic errors (such as biases in the force field or forward model), random errors (due to noise in experimental data), and errors due to the limited sample size of the ensemble¹⁷. The model generation is governed by the metainference energy function, defined as $E_{MI} = -k_B T \log(p_{MI})$, in which k_B is the Boltzmann constant, T is the temperature, and p_{MI} is the metainference posterior probability

$$p_{MI}(\mathbf{X}, \sigma^{SEM}, \sigma^B | \mathbf{D}) = \prod_{r=1}^{N_R} p(X_r) \prod_{i=1}^{N_D} p(d_i | \mathbf{X}, \sigma_i^{SEM}, \sigma_{r,i}^B) p(\sigma_{r,i}) \quad (3)$$

Here, \mathbf{X} is a vector representing the atomic coordinates of the full ensemble, consisting of individual replicas X_r ; σ^{SEM} is the error incurred by the limited number of replicas in the ensemble; σ^B encodes the random and systematic errors in the prior, forward model, and experiment; and $\mathbf{D} = [d_i]$ is the experimental data. Note σ^{SEM} is calculated per data point (σ_i^{SEM}), while σ^B is computed per datapoint i and replica r as $\sigma_{i,r}^B$. In the present case, the likelihood $p(d_i | X, \sigma_i^{SEM}, \sigma_{r,i}^B)$ takes the form of a Gaussian function

$$p(d_i|\mathbf{X}, \sigma_i^{SEM}, \sigma_{r,i}^B) = \frac{1}{\sqrt{2\pi} \sqrt{(\sigma_{r,i}^B)^2 + (\sigma_i^{SEM})^2}} \exp\left[-\frac{1}{2} \frac{[ov_{DD,i} - \overline{ov}_{MD,i}]^2}{(\sigma_{r,i}^B)^2 + (\sigma_i^{SEM})^2}\right] \quad (4)$$

where $\overline{ov}_{MD,i}$ is the ensemble average of the overlap. The metainference energy function for multiple replicas then becomes

$$E_{MI}(\mathbf{X}, \sigma) = E_{MD}(\mathbf{X}) + \frac{k_B T}{2} \sum_{r,i}^{N_R, N_D} \frac{[ov_{DD,i} - \overline{ov}_{MD,i}]^2}{(\sigma_{r,i}^B)^2 + (\sigma_i^{SEM})^2} + E_\sigma \quad (5)$$

where E_σ represents the energy associated with the error $\sigma = (\sigma^B, \sigma^{SEM})$

$$E_\sigma = k_B T \sum_{r,i}^{N_R, N_D} -\log(\sigma_{r,i}^B) + \frac{1}{2} \log\left[(\sigma_{r,i}^B)^2 + (\sigma_i^{SEM})^2\right] \quad (6).$$

E_{MD} represents the molecular dynamics force field. While the space of conformations X_r is sampled by multi-replica molecular dynamics simulations, the error parameters for each datapoint $\sigma_{r,i}^B$ are sampled by a Monte Carlo sampling scheme at each time step. The error parameter related to the limited number of replicas used to estimate the forward model (σ^{SEM}) can be chosen as a constant or estimated on the fly by using a windowed average³².

Metadynamic cryo-EM metainference (MEMMI). To accelerate the sampling of the metainference ensemble, one can utilise an enhanced-sampling scheme such as metadynamics^{51,60}. In this case, we use parallel-bias metadynamics (PBMetaD)⁵⁵ with the multiple walkers scheme⁷¹. Here, V_{PB} is a time-dependent biasing potential acting on a set of N_{CV} collective variables $s(\mathbf{X})$, which in turn are functions of the system coordinates

$$V_{PB}(s_j(\mathbf{X}), t) = -k_B T \log\left\{\sum_{j=1}^{N_{CV}} \exp\left[\frac{V_G(s_j(\mathbf{X}), t)}{k_B T}\right]\right\} \quad (7)$$

In contrast to conventional metadynamics, in PBMetaD multiple one-dimensional bias potentials V_G are deposited rather than a single high-dimensional one. This alleviates the curse of dimensionality while still allowing an efficient exploration of phase space⁵⁵. Additionally, the use of multiple replicas through the multiple walkers scheme⁷¹ allows the sharing of the bias potential to drastically improve the sampling performance, while at the same time being a natural fit for the replica averaging approach of EMMI. Analogously to well-tempered metadynamics, these bias potentials $V_G(s_j(\mathbf{X}), t)$ eventually converge to the free energy $F(s_j(\mathbf{X}))$. The MEMMI energy function then becomes

$$E_{\text{MEMMI}}(\mathbf{X}, \sigma, t) = E_{\text{EMMI}}(\mathbf{X}, \sigma) + \sum_{r=1}^{N_R} V_{\text{PB}}[\mathbf{s}(\mathbf{X}_r), t] \quad (8)$$

While the PBMetaD bias potential is shared among replicas, each replica may still experience a varying potential depending on its location in phase space. Thus, the arithmetic average over the forward models (i.e., the overlap) no longer presents an unbiased estimate of the ensemble average. It therefore needs to be replaced with a weighted average, utilising the bias potential of each replica r to unbiased the ensemble at time t

$$f_i(\mathbf{X}, t) = \sum_{r=1}^{N_R} w(\mathbf{X}_r, t) f_i(\mathbf{X}_r, t) \quad (9)$$

where the unbiasing weight $w(\mathbf{X}_r, t)$ is defined as

$$w(\mathbf{X}_r, t) = \exp\left\{\frac{V_{\text{PB}}[\mathbf{s}(\mathbf{X}_r), t]}{k_B T}\right\} \left[\sum_{j=1}^{N_R} \exp\left\{\frac{V_{\text{PB}}[\mathbf{s}(\mathbf{X}_j), t]}{k_B T}\right\}\right]^{-1} \quad (10).$$

The unbiasing procedure used here is analogous to the standard umbrella-sampling technique by Torrie and Valleau⁴⁷. Now the ensemble average $\overline{ov}_{MD,i}(\mathbf{X})$ is given by

$$\overline{ov}_{MD,i}(\mathbf{X}) = \sum_{r=1}^{N_R} w(\mathbf{X}_r, t) ov_{MD,i}(\mathbf{X}_r) \quad (12)$$

The MEMMI energy E_{MEMMI} is thus equal to Eq. 5.

Initial fibril structure. We build the initial structure of the full fibril by starting from a deposited fibril structure (PDB:6Y1A), which only contains the core of the fibril, and extending each of the 16 peptide chains by adding the missing 12-residue N-terminal sequence with guidance from the cryo-EM density map EMD-10669. We use the macromolecular model-building program coot⁷².

Molecular dynamics setup and equilibration. We continue by creating a $12.34 \times 12.34 \times 12.34$ nm cubic simulation box, solvating with 58291 water molecules and neutralizing the net charge by adding 48 Cl⁻ ions. We use the CHARMM22*⁷³ force field and TIP3P⁷⁴ water models. We continue with an energy minimization followed by a 500 ps NPT equilibration at a temperature of 310 K and pressure of 1 atm, followed by an additional 2 ns NVT equilibration at 310 K. The molecular dynamics parameters are the same used previously⁴⁰.

MEMMI simulations. We first express the experimental voxel map data as a data GMM containing 10,000 Gaussians in total, resulting in a 0.975 correlation to the original voxel map EMD-10669⁶¹, using the gmmconvert utility⁷⁵. We continue by extracting 32 configurations from the previous NVT equilibration and initiate a MEMMI simulation, consisting of 32 replicas, resulting in an aggregate runtime of 5.49 μ s, using PLUMED.2.6.0-dev⁷⁶ and gromacs-2020.6⁷⁷. The simulation is performed in the NVT ensemble at 310 K using the same

MD parameters as in the equilibration step. Configurations are saved every 10 ps for post-processing. The cryo-EM restraint is updated every 2 MD steps, using neighbour lists to compute the overlaps between model and data GMMs, with a neighbour list cut-off of 0.01 and update frequency stride of 100 steps. The biasing collective variables $s = [s_i]$ in the simulation are shown in **Figure S1A** and the biasing scheme is PBMetaD⁵⁵ with the well-tempered⁷⁸ and multiple-walkers⁷¹ protocols. The hill height is set to 0.3 kJ/mol, with a deposition frequency of 200 steps and an adaptive Gaussians diffusion scheme⁷⁹. The biasing collective variables correspond to degrees of freedom of the left hand-side N-terminal. The respective degrees of freedom of the right hand-side N-terminal do not feel a metadynamics potential and therefore in the remaining text will be referred to as EMMI degrees of freedom and are also listed in **Figure S1**. As a post-processing step, we generate the final structural ensemble by resampling the generated configurations based on the converged unbiasing weights for each structure after an equilibration of 7 ns shown in **Figure 1A**. To establish convergence, we perform a clustering analysis on the structural ensemble based separately on the first and second half of each replica (taking into accounts the weights) using the GROMOS method⁸⁰, and with metric the root mean square deviation (RMSD) calculated on the C α (CA) atoms (**Figure S1B**). Time-traces of CVs as well as their time-dependent free energy profile are shown in Figs. S2-S4. For molecular visualizations and calculating the local correlation of the final structural-ensemble-generated cryo-EM map with the experimental cryo-EM map, we use Chimera and gmconvert^{81,75}. Except otherwise mentioned, all the structural analysis is performed on degrees of freedom of left side tail N-tail and chains 3 and 4. We do so in order to avoid the finite size effects on the outermost N-tails.

Structure and dynamics of IAPP amyloid fibrils

Acceleration of the conformational sampling. MEMMI accelerates the conformational sampling by biasing a set of microscopic degrees of freedom of the system, also known as collective variables (CVs, **Figure S1**). To demonstrate the performance of this approach, we visualize the time-trace of the collective variable χ_3^4 which corresponds to the disulfide bond dihedral of the 4th peptide in the eight-layer stack of the fibril. This peptide is thus representative of a buried monomer with little interaction with the fibril ends. Note that, due to C2 helical symmetry, there are two of these dihedrals, one corresponding to the one side and one to the other sides N-terminal tail (**Figure 1B**). The sampling of one side is accelerated by a biasing potential of Eq. 7 (MEMMI), while the other is not (EMMI). Compared to the dihedral, the biased disulfide χ_3^4 (MEMMI) shows an increased transition rate, and thus more efficient conformational sampling (**Figure 1C**). While both methods characterise the wells of the stable states (L, R), MEMMI is able to provide access to higher free energy transition state (TS) regions (**Figure 1D**). Monitoring a non-periodic CV, such as the number of contacts between N-tail three and four, shows diffusion along high-low contact regions in both MEMMI and EMMI cases but somewhat more frequent in the MEMMI case. The combination of diffusion shown in the traces of CVs and the free energy profiles as a function of simulation time of all collective variables biased by EMMI and MEMMI are shown in **Figures S2-S4**, and indicate that our simulations are well converged. Taken together, these results show that both

the MEMMI and EMMI simulations are converged in the low free energy regions, while MEMMI enables visiting high free energy regions.

Conformational heterogeneity of the fuzzy coat. A structure of the IAPP fibril core (residues 13-37) has been previously published (PDB:6Y1A) using the cryo-EM data used also in the present work (EMD-10669). Here, we determine a structural ensemble of the whole IAPP fibril (residues 1-37). We model the system as a stack of eight peptides per side (**Figure 2A**). While the core of the fibril maintains a parallel β -sheet structure, the flanking region (residues 1-12) exhibits a large conformational heterogeneity (**Figure 2B**). While we find the cores residues 12-37 to be largely in a β -sheet conformation, we also note significant heterogeneity for residues 23-24, 32-34 and 37. As shown in **Figure 2B**, residues 23-24 interact with TYR37 and maintain mostly a coil structure, while residues 32-34 interact with the fuzzy coat and maintain mostly a coiled structure. We also detected a small population of α -helical conformations in the region of residues 5-9.

Correlation between experimental and calculated cryo-EM maps. We estimate the correlation of the MEMMI structural ensemble with the experimental cryo-EM map (**Figure 3**). We find that using a structural ensemble IAPP (residues 1-37) correlates better with the experimental cryo-EM map than a single structure (PDB:6Y1A) (**Figure 3A,B**). The coefficient of correlation of the structural ensemble to the experimental electron density map is on average 0.92. Furthermore, an important feature of MEMMI is its ability to estimate the error in the experimental electron density map (**Figures 3C and S5**). We find that the relative error per Gaussian data point is on average 0.09, where the relative error is the error of each Gaussian data point with respect to the total overlap between all data GMM and the i^{th} component of the data GMM³⁵. The low-density, high-error volume around residues S34, N35, and T36 can likely be attributed to the lack of MES/NaOH buffer in our MEMMI simulations, which is present in the experimental setup. Both MEMMI and EMMI exhibit good correlation (about 0.83) in the respective N-tail region with the cryo-EM map (**Figure 3D**).

Comparison of the dynamical properties of the fibril core and fuzzy coat. The conformational properties of the fuzzy coat and core region have been shown to be relevant modulating the properties of amyloid fibrils, including their ability to interact with various cellular components^{69,70}. To investigate this phenomenon in the case of the present IAPP amyloid fibril, we take inspiration from a thermodynamic theory of melting and characterize the residue-dependent Lindemann parameter Δ_L (**Figure 4**), which encodes information on solid-like and liquid-like behaviour⁸². At the backbone-level, we find that the fibril core (23-30) is solid-like ($\Delta_L < 0.15$), while the flanking region (1-12) is on the verge of a liquid-like behaviour ($\Delta_L \geq 0.15$). The Lindemann parameters of the side-chains indicate more mobility and are liquid-like outside the region of residues 20-32. These results reveal that about half of the structural core of this amyloid fibril remains rather disordered at the side-chain level, a phenomenon observed also for folded native states⁸². This is in contrast to monomeric disordered proteins such as amyloid- β , which generally exhibit Lindemann parameters $\Delta_L \geq 1.0$ (**Figure S7**).

Residue-specific solubility of the amyloid fibril. To investigate whether or not the surface of the amyloid fibril is soluble, we calculate the solubility per residue, using the structure-corrected CamSol solubility score⁸³ (**Figure 5A,B**). We find three insoluble regions: (i) residues 5-7 (ATC) in the fuzzy coat, (ii) residues 26-29 (ILSS) at the ends of the fibril fragment, and (iii) residues 14-18 (NFLVH) on the fibril surface, stretched along the helical axis. The latter result is consistent with experimental evidence of residues N14, H18, and S20 being implicated in IAPP aggregation^{63,84} via secondary nucleation. We note that the solvent-exposed and aggregation-related residues 5-7 and 18-20 might present attractive targets for structure-based drug discovery.

Conclusions

We have presented the MEMMI method for the simultaneous determination of the structure and dynamics of large and conformationally heterogeneous biomolecular structures from cryo-EM measurements. To illustrate the type of information that can be extracted from this type of approach, we have reported a structural ensemble of an amyloid fibril formed by IAPP. The analysis of the structural ensembles has revealed the conformational and the aggregation-related solubility properties of the fuzzy coat of the amyloid fibril, and that many of the side-chains in the structural core of the amyloid fibril exhibit liquid-like behaviour. Since this phenomenon has also been observed for native states of proteins, these results reveal a similarity in the structural behaviour of proteins upon folding in their native and amyloid states.

Data availability. Input files for the simulations can be found at PLUMED-NEST ([plumID:22.023](https://zenodo.org/record/6518554#.YnLwby8Rqhw)) The ensemble can be found at <https://zenodo.org/record/6518554#.YnLwby8Rqhw>

Acknowledgements. Z.F.B. would like to acknowledge the Federation of European Biochemical Societies (FEBS) for financial support (LTF). M.B. and S.E.H. acknowledge the support of the French Agence Nationale de la Recherche (ANR), under grant ANR-20-CE45-0002 (project EMMI).

References

1. Kühlbrandt, W. The resolution revolution. *Science* **2014**, 343 (6178), 1443-1444.
2. Bai, X.C., McMullan, G. and Scheres, S. H. W. How cryo-EM is revolutionizing structural biology. **2014**, *Trends Bioch. Sci.* 40 (1), 49-57.
3. Dubochet, J. and McDowell, W. Vitrification of pure water for electron microscopy. *J. Microsc.* **1981**, 124 (3), 3-4.
4. Dubochet, J., Lepault, J., Freeman, J., Berriman, J. A. and Homo, J. C. Electron microscopy of frozen water. *J. Microsc.* **1982**, 128 (3), 219-237
5. Renaud, J. P. et al. Cryo-EM in drug discovery: Achievements, limitations and prospects. *Nature Rev. Drug. Disc.* **2018**, 17 (7), 471-492.
6. van Heel, M. and Frank, J. Use of **multivariate** statistics in analysing the images of biological macromolecules. *Ultramicroscopy*, **1981**, 6 (2), 187-194
7. Frank, J., Shimkin, B. and Dowse, H. Spider-A modular software system for electron image processing. *Ultramicroscopy*, **1981**, 6 (1), 343-358
8. Dashti, A. et al. Trajectories of the ribosome as a Brownian nanomachine. *Proc. Nat. Acad. Sci. U.S.A.* **2014**, 111 (49), 17492-17497
9. Bonomi, M., and Vendruscolo, M. Determination of protein structural ensembles using cryo-electron microscopy. *Curr. Op. Struct. Biol.* **2019**, 56, 37-45.
10. Bonomi, M., Heller, G. T., Camilloni, C. and Vendruscolo, M. Principles of protein structural ensemble determination. *Curr. Op. Struct. Biol.* **2017**, 42, 106-116.
11. Fraser, J. S., Lindorff-Larsen, K. and Bonomi, M. What will computational modeling approaches have to say in the era of atomistic cryo-EM data. *J. Chem. Inf. Mod.* **2020**, 60 (5), 2410-2412
12. Pittera, J. W. and Chodera, J. D. On the use of experimental observations to bias simulated ensembles *J. Chem. Theory Comput.* **2012**, 8 (10), 3445-3451
13. Cavalli, A., Camilloni, C. and Vendruscolo, M. Molecular dynamics simulations with replica-averaged structural restraints generate structural ensembles according to the maximum entropy principle. *J. Chem Phys.* **2013**, 138 (9), 094112
14. Roux, B. and Weare, J. On the statistical equivalence of restrained-ensemble simulations with the maximum entropy method. *J. Chem Phys.* **2013**, 138 (8), 084107
15. Boomsma, W. Ferkingho-Borg, J and Lindorff-Larsen, K. Combining experiments and simulations using the maximum entropy principle. *PLoS Comput. Biol.* **2014**, 10 (2), e1003406
16. Hummer, G. and Kofinger, J. Bayesian ensemble refinement by replica simulations and reweighting. *J. Chem. Phys.* **2015**, 143 (23), 243150
17. Bonomi, M., Camilloni, C., Cavalli, A. and Vendruscolo, M. MetaInference: A Bayesian inference method for heterogeneous systems. *Sci. Adv.* **2016**, 2 (1), e1501177
18. Marrinelli, F. and Faraldo, J. D. Ensemble-biased metadynamics. A molecular simulation method to sample experimental distributions. *Biophys. J.* **2015**, 108 (12), 2779-2782
19. Olsson, S., Wu, H., Paul, F., Clementi, C., Noe, F. Combining experimental and simulation data of molecular processes via augmented Markov models. *Proc. Nat. Acad. Sci. U.S.A.* **2017**, 114 (31), 8265-8270
20. Cesari, A., Reißer, S and Bussi, G. Using the maximum entropy principle to combine simulations and solution experiments. *Computation* **2018**, 6 (1), 15.

21. Bottaro S., Bengtsen, T. and Lindorff-Larsen, K. Integrating molecular simulation and experimental data. A Bayesian/maximum entropy reweighting approach. *Methods. Mol. Biol.* **2020**, 2112, 219-240
22. Brotzakis, Z. F., Vendruscolo., M and Bolhuis, P. G. A method of incorporating rate constants in molecular dynamics simulations. *Proc. Natl. Acad. Sci. U.S.A.* **2021**, 118, e1012423118
23. Bolhuis, P.G., Brotzakis Z. F. and Vendruscolo, M. A maximum caliber approach for path ensembles *Eur. Phys. J. B.* **2021**, 94 (188), 1-21
24. Camilloni, C. and Vendruscolo. M. A tensor free method for the structural and dynamical refinement of proteins using residual dipolar couplings. *J. Phys. Chem. B.* **2015**, 119 (3), 653-661
25. Bottaro, S. and Lindorff-Larsen, K. Biophysical experiments and molecular simulations. *Science*, **2015**, 361, 355
26. Heller, G.T., et al. Small-molecule sequestration of amyloid- β as a drug discovery strategy for Alzheimer's disease. *Sci. Adv.* **2020**, 6 (45), eabb592
27. Löhr, T., Kohlhiff, K., Heller, G.T., Camilloni, M and Vendruscolo, M. A kinetic ensemble of Alzheimer's A β peptide. *Nat. Comp. Sci.* **2021**, 1 (1), 71-78
28. Löhr, T., Jussupow, A. and Camilloni, C. Metadynamic metainference: Convergence towards force field independent structural ensembles of a disordered peptide. *J. Chem. Phys.* **2017**, 146, 165102
29. Bonomi, M. et al. Determining protein complex structures based on a Bayesian model of in vivo Förster Resonance Energy Transfer (FRET) data. *Mol. Cell. Proteomics.* **2014**,
30. Pollack, L. Time resolved folding and RNA folding. *Biopolymers.* **2011**, 95 (8), 543-549.
31. Hermann, M. R. and Hub, J. S. SAXS restrained ensemble simulations of intrinsically disordered proteins with commitment to the principle of maximum entropy. *J. Chem. Theory Comput.* **2019**, 15 (9), 5103-5115
32. Bengtsen, T. et al. Structure and dynamics of a nanodisc by integrating NMR, SAXS and SANS experiments with molecular dynamics simulations. *eLife*, **2020**, 9, e56518
33. Paisonni, C., Jussupow, A. and Camilloni, C. Determination of protein structural ensembles by hybrid resolution SAXS restrained molecular dynamics. *J. Chem. Theory Comput.* **2020**, 16 (4), 2825-2834
34. Jussupow, A. et al. The dynamics of linear polyubiquitin. *Sci. Adv.* **2020**, 6 (42), eabc3786
35. Bonomi, M., Pellarin, R. and Vendruscolo, M., Simultaneous determination of protein structure and dynamics using cryo-electron microscopy. *Biophys. J.* **2018**, 114 (7), 1604-1613.
36. Vahidi, S. et. al. Reversible inhibition of the ClpP-protease via an N-terminal conformational switch. *Proc. Natl. Acad. Sci. U.S.A* **2018**, 115 (28), E6447-E6456
37. Eshun-Wilson, L. et. al. Effects of alpha-tubulin acetylation on microtubule structure and stability. *Proc. Nat. Acad. Sci. U.S.A.* **2019**, 116 (21), 10366-10371
38. Brotzakis, Z. F., et al. A structural ensemble of a tau-microtubule complex reveals regulatory tau phosphorylation and acetylation mechanisms. *ACS Cent. Sci.* **2021**, 7 (12), 1986-1995

39. Garibsingh, R-AA et. al. Rational Design of ASCT2 inhibitors using an integrated experimental-computational approach. *Proc. Natl. Acad. Sci. U.S.A* **2021**, 118 (37), e2104093118
40. Brotzakis, Z. F., Löhr, T. and Vendruscolo, M. Determination of intermediate state structures in the opening pathway of SARS-CoV-2 spike using cryo-electron microscopy. *Chem Sci* **2021**, 12, 9168-9175.
41. Dellago, C., Bolhuis, P. G., Csajka, F. S. and Chandler, D. Transition path sampling and the calculation of rate constants. *J. Chem. Phys.* **1998**, 108 (5),1964
42. van Erp, T.S., Moroni, D. and Bolhuis, P.G. A novel path sampling method for the calculation of rate constants. *J. Chem. Phys.* **2003**, 118 (17), 7762
43. Allen, R.J., Warren P. B. and ten Wolde, P. R. Sampling rare switching events in biochemical networks. *Phys. Rev. Lett.* **2005**, 94, 018104
44. Faradjian, A.K. and Elber, R. Computing timescales from reaction coordinates by milestoning. *J. Chem. Phys.* **2004**, 120, 10880
45. Brotzakis, Z. F. Approximating free energy and committer landscapes in standard transition path sampling using virtual interface exchange. *J. Chem. Phys.* **2019**, 151, 174111
46. Chong, L. T., Saglam, A. S. and Zuckerman, D.M. Path sampling strategies for simulating rare events in biomolecular systems. *Curr. Opin. Struct. Biol.* **2017**, 43, 88-94
47. Torrie, J. and Valleau, G. M. Non physical sampling distribution in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **1977**, 23, 187-199
48. Grubmuller, H. Predicting slow structural transitions in macromolecular systems: Conformational flooding. *Phys. Rev. E.* **1995**, 52 (3), 2893-2906
49. Wang, F. and Landau. Efficient, multiple-range random walk algorithm to calculate the density of states. *Phys. Rev. Lett.* **2001**, 86 (10),2050-2053
50. Darve, E. and Pohorille, A. Calculating free energies using average force. *J. Chem. Phys.* **2001**, 115 (20), 9169
51. Laio, A. and Parrinello, M. Escaping free energy minima. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, 99 (20), 12562-12566
52. Valsson, O., Tiwary, P. and Parrinello, M. Enhancing rare fluctuations: rare events and metadynamics from a conceptual viewpoint. *Ann. Rev. Phys. Chem.* **2016**, 67, 159-184
53. Bolhuis, P.G and Swenson, D. W. Transition path sampling as Markov chain Monte Carlo of trajectories: Recent algorithms, software, applications, and future outlook. *Adv. Theory Simul.* **2021**, 4 (4),1
54. Bonomi, M. et al. Promoting transparency and reproducibility in enhanced molecular simulations. *Nature Methods.***2019**, 16, 670-673
55. Pfaendtner J. and Bonomi, M. Efficient sampling of high dimensional free energy landscapes with parallel bias metadynamics. *J. Chem. Theory Comput.* **2015**, 11, 5062
56. Tiwary, P. and Berne, B. J. Spectral gap optimization of order parameters for sampling complex molecular systems. *Proc. Natl. Acad. Sci. U.S.A.* **2016**, 113 (11), 2839-2844
57. Sultan, M. M. and Pande, V. J. Accelerating metadynamics by using kinetically selected collective variables. *J. Chem. Theory Comput.* **2017**, 13 (6), 2440-2447
58. Marcelo, J. et al. Reweighted autoencoder variational bayes for enhanced sampling (RAVE). *J. Chem. Phys.* **2018**, 149, 072301

59. Brotzakis, Z. F. and Parrinello, M. Enhanced sampling of protein conformational transition via dynamically optimized collective variables. *J. Chem. Theory Comput.* **2018**, 15 (2),1393-1398
60. Bonomi, M. Camilloni, C. and Vendruscolo, M. Metadynamic metainference: enhanced sampling of the metainference ensemble using metadynamics. *Sci. Rep.* **2016**, 6, 31232
61. Pilkington, E.H. et. al. Effects of protein corona on IAPP amyloid aggregation, fibril remodeling and toxicity. *Sci. Rep.* **2017**, 7, 2455
62. Zhang, J., Tan, J., Pei, R. and Ye, S. Acidic environment significantly alters aggregation pathway of human islet amyloid polypeptide at negative lipid membrane. *Langmuir.* **2020**, 36, 1530-1537
63. Santos, J. et. al. pH-Dependent Aggregation in intrinsically disordered proteins is determined by charge and lipophilicity. *Cells* **2020**, 9, (1), 145
64. Nguyen PH, et. al. Amyloid oligomers: A joint experimental/computational perspective on Alzheimer's disease, Parkinson's disease, type II diabetes, and amyotrophic lateral sclerosis. *Chem Rev.* **2021**,121 (4), 2545-2647
65. Rodriguez Camargo, D. C. et al. Surface-catalyzed secondary nucleation dominates the generation of toxic IAPP aggregates., *Front. Mol. Biosci.* **2021**, 8, 757425.
66. Cao, Q., et al. Cryo-EM structure and inhibitor design of human IAPP (amylin) fibrils. *Nat. Struct. Mol. Biol.* **2020**, 27 (7), 653-659.
67. Cao, Q, et al. Cryo-EM structures of hIAPP fibrils seeded by patient-extracted fibrils reveal new polymorphs and conserved fibril cores. *Nat. Struct. Mol. Biol.* **2021**, 28 (9), 724-730.
68. Röder, C. R. et al. Cryo-EM structure of islet amyloid polypeptide fibrils reveals similarities with amyloid- β fibrils. *Nat. Struct. Mol. Biol.* **2020**, 27 (7), 660-667
69. Ulamec, S. M., Brockwell, D. J., Radford S. E. Looking beyond the vore: The role of flanking regions in the aggregation of amyloidogenic peptides and proteins. *Front Neurosci.* **2020**, 14, 611285.
70. Wegmann, S., Medalsy, I. D., Mandelkow, E. and Müller, J. The fuzzy coat of pathological human tau fibrils is a two-layered polyelectrolyte brush. *Proc. Natl. Acad. Sci.* **2013**, 110, (4), E313-E321
71. Raiteri, P., Laio, A., Gervasio, F. L., Micheletti, C. and Parrinello, M. Efficient reconstruction of complex free energy landscapes by multiple walkers metadynamics. *J. Phys. Chem. B.* **2006**, 110, 3533– 3539.
72. Emsley, P. Bernhard, L., Scott, W.G. and Cowtan, K. Features and development of coot. *Acta. Crystallogr. D Biol. Crystallogr.* **2010**, 66 (4), 486-501
73. Piana, S., Lindorff-Larsen, K. and Shaw, D.E. How robust are protein-folding simulations with respect to force field parameterization? *Biophys J.* **2011**,100 (9), L47
74. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. and Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, 79 (2), 926-935.
75. Kawabata, T. Multiple Subunit Fitting into a low-resolution density map of a macromolecular complex using a gaussian mixture model. *Bioph. J.* **2008** 95, 4643-4658.

76. Tribello, G. A., Bonomi, M. Branduardi, D., Camilloni, C., Bussi, G. PLUMED 2: New feathers for an old bird. *Comp. Phys. Comm.* **2014**, 185 (2), 604-613.
77. Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B. and Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*. **2015**, 1 (2) 19–25
78. Barducci, A., Bussi, G. and Parrinello, M. Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **2008** 100, 020603
79. Branduardi, D. Bussi, G. and Parrinello, M. Metadynamics with adaptive Gaussians. *J. Chem. Theory Comput.* **2012**, 8 (7), 2247-2254
80. Daura, X., van Gunsteren, W. F., Jaun, B., Mark, A. E., Gademann, K. and Seebach, D., Peptide folding: When simulation meets experiment. *Angew. Chem., Int. Ed.* **1999**, 38 (1/2), 236–240
81. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E., UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comp. Chem.* **2004**, 25 (13), 1605-1612
82. Lindorff-Larsen, K., et. al. Simultaneous determination of protein structure and dynamics. *Nature*. **2005**, 433, 128-132
83. Sormanni, P., Aprile, F. A. and Vendruscolo, M. The CamSol method of rational design of protein mutants with enhanced solubility. *J Mol Biol.* **2015**, 427 (2), 478-90.
84. Cao, P et. al. Sensitivity of amyloid formation by human islet amyloid polypeptide to mutations at residue *Mol Biol.* **2012**, 421 (2-3), 282-95.

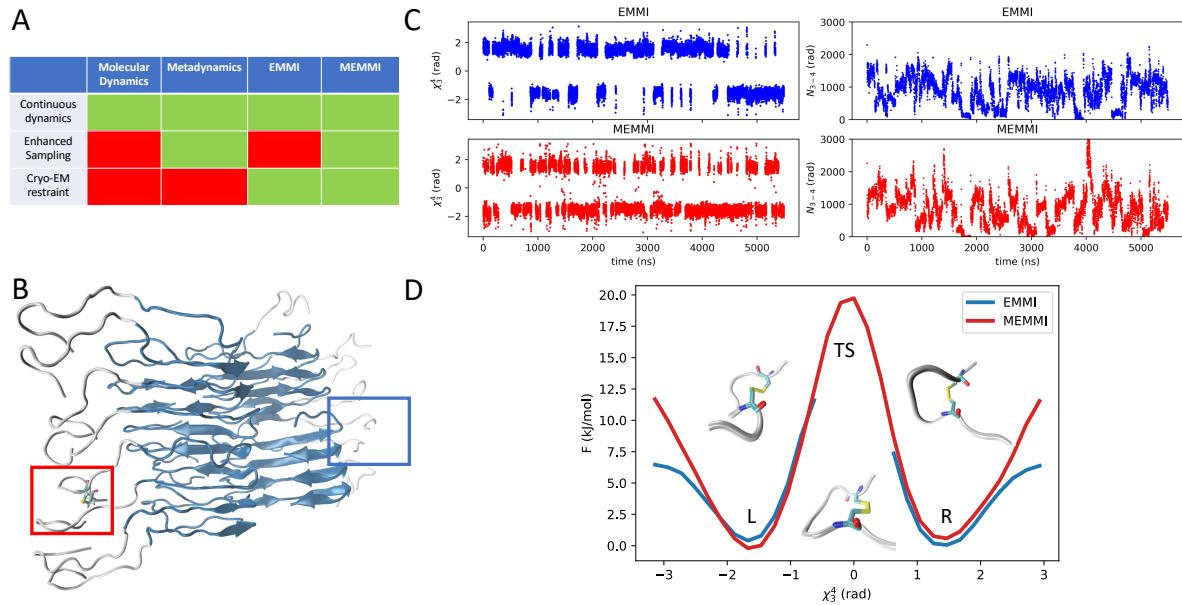


Figure 1. Acceleration of the conformational sampling using MEMMI. (A) Summary of the capability of the different approaches discussed in this work (molecular dynamics, metadynamics, EMMI and MEMMI) to include continuous dynamics, enhanced sampling and cryo-EM restraints. (B) A representative configuration of an IAPP amyloid fibril with a biased disulfide bond (red) and an unbiased one (blue). (C) Time trace of the χ_3^4 and N_{3-4} CVs with EMMI and MEMMI. (D) Free energy profile as a function of χ_3^4 with the corresponding structural states of the disulfide bond: MEMMI (red) EMMI (blue).

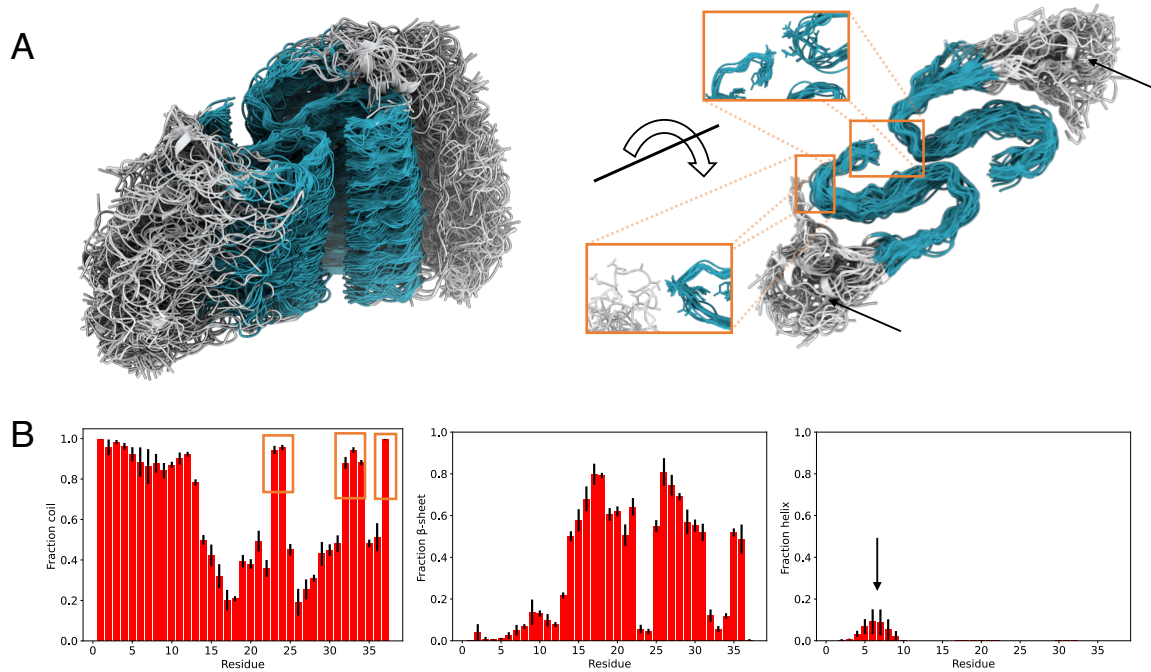


Figure 2. A structural ensemble of a IAPP amyloid fibril. (A) Core fibril residues (13-37) are shown in cyan, while the N-terminal tail residues (1-12) are shown in grey (fuzzy coat). This representation of the structural ensemble was generated by extracting 50 conformations from the final structural ensemble. Close-ups of interfaces are shown from 10 conformations. Examples of individual conformations are shown in **Figure S6. (B)** Structural analysis reporting on the fraction of coil, β -sheet and α -helix formed per residue in the ensemble obtained from MEMMI. Error bars are calculated as standard deviation between the first and second simulation half.

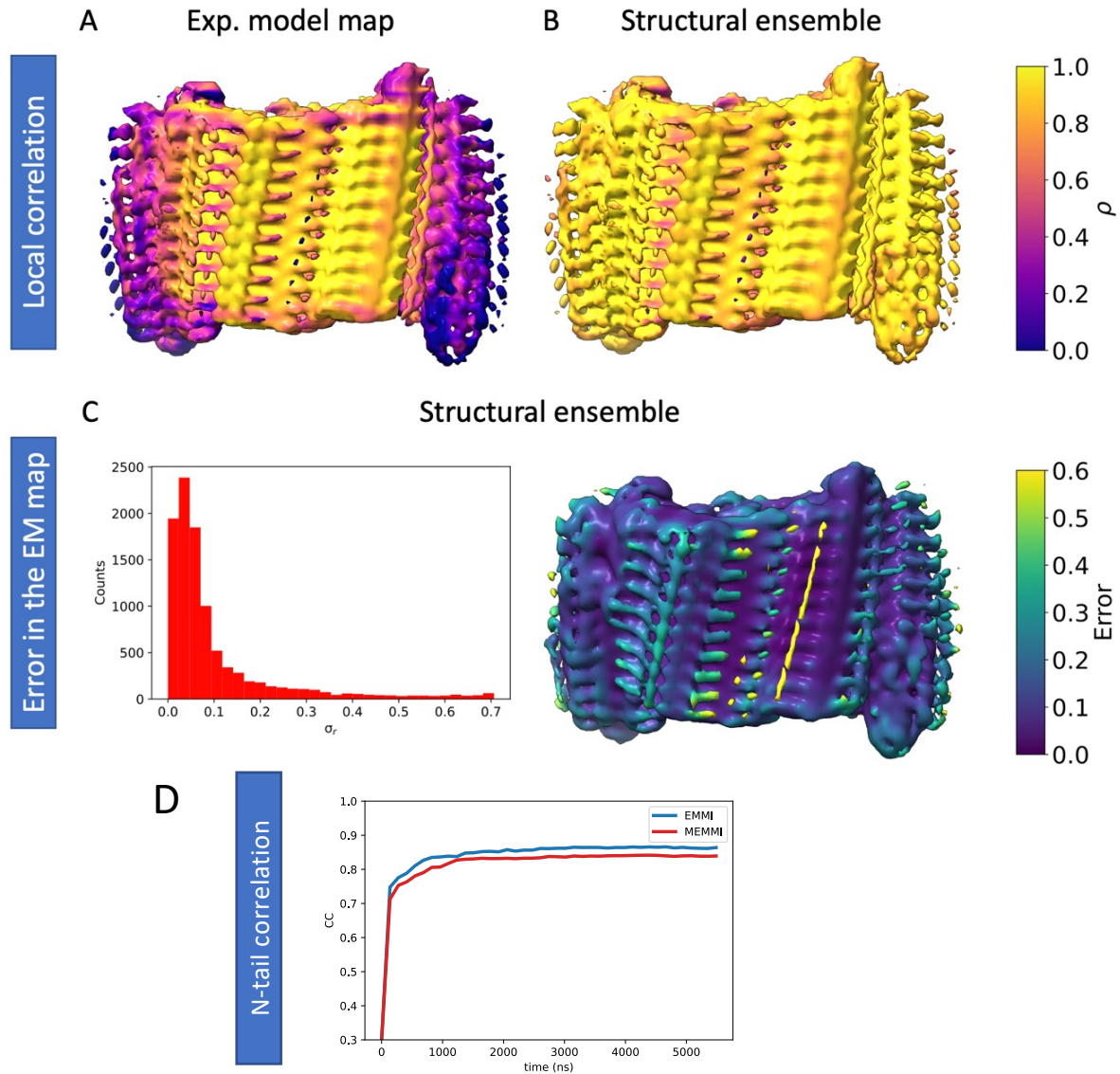


Figure 3. Local correlation and error in the data from MEMMI. (A, B) Assessment of the local correlation between the cryo-EM map (EMD-EM-10669) with cryo-EM maps generated by a previously determined single structure (PDB:6Y1A) (A), and the MEMMI structural ensemble (B). (C) Histogram of the error in the GMM data as obtained by the structural ensemble (left). Error in the data, projected in the EMD-EM-10669 (right). (D) Time-dependent correlation of the left/right side N-tail structural ensemble (MEMMI/EMMI) with the corresponding region in the cryo-EM map.

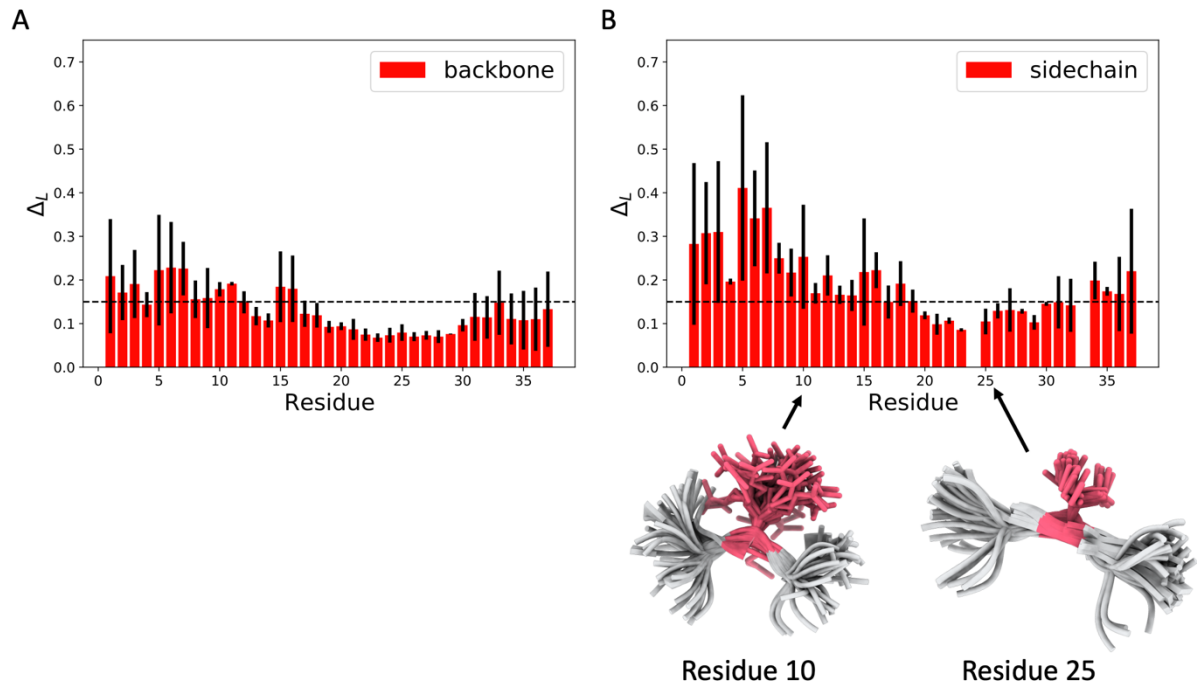


Figure 4. Analysis of the liquid-like and solid-like behaviours of the backbone and side-chains. (A,B) Per-residue Lindemann parameter (Δ_L) for backbone (A) and side-chain (B) atoms. The dashed line at $\Delta_L = 0.15$ marks the threshold value of the Lindemann parameter that distinguishes the solid-like and liquid-like behaviours. The behaviour of the side-chains is solid-like only in the region of residues 20-32, which is about half of the fibril core (residue 13-37). Error bars are calculated as standard deviation between the first and second simulation half.

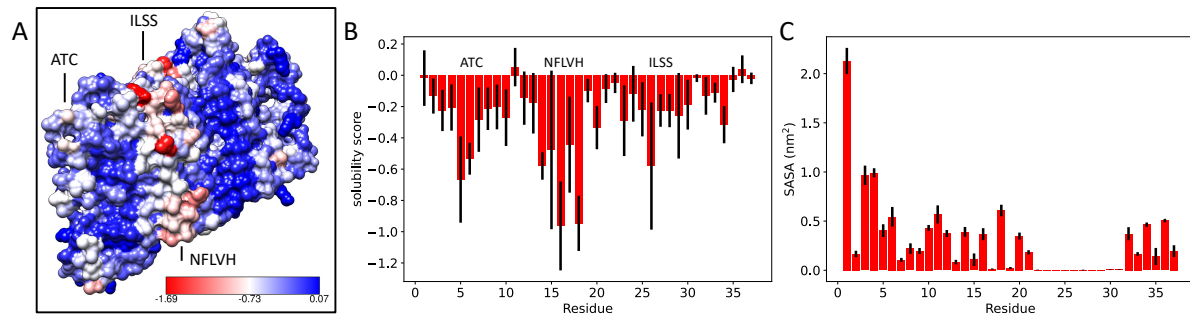


Figure 5. Analysis of the solubility of the fibril surface from the MEMMI ensemble. (A) Representative structure with CamSol solubility scores indicated on the surface (right). **(B)** Per-residue CamSol solubility score. **(C)** Per-residue solvent-accessible surface area. Error bars are calculated as standard deviation between the first and second simulation half.

Supplementary Information

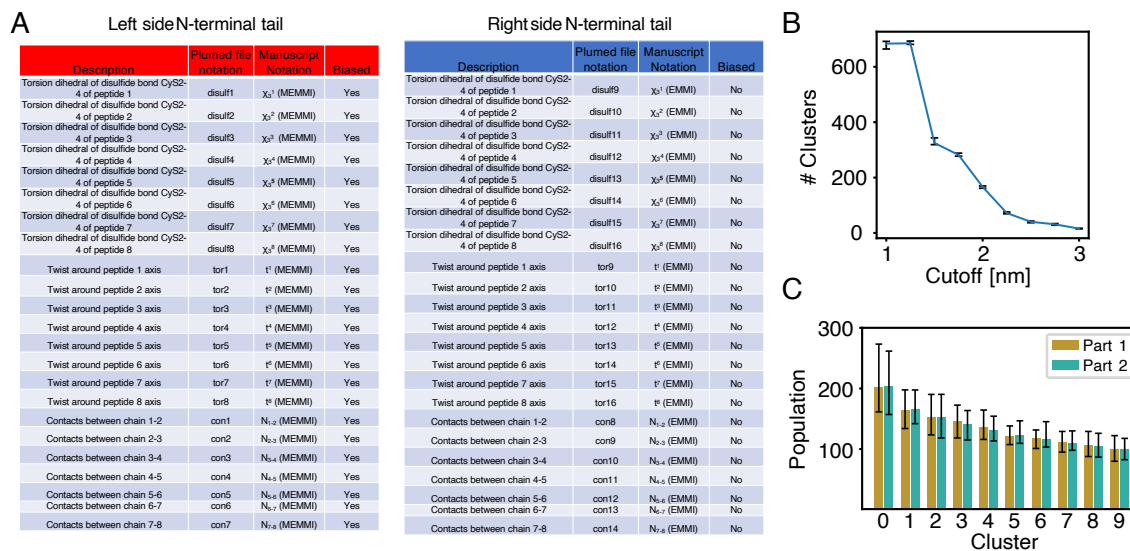
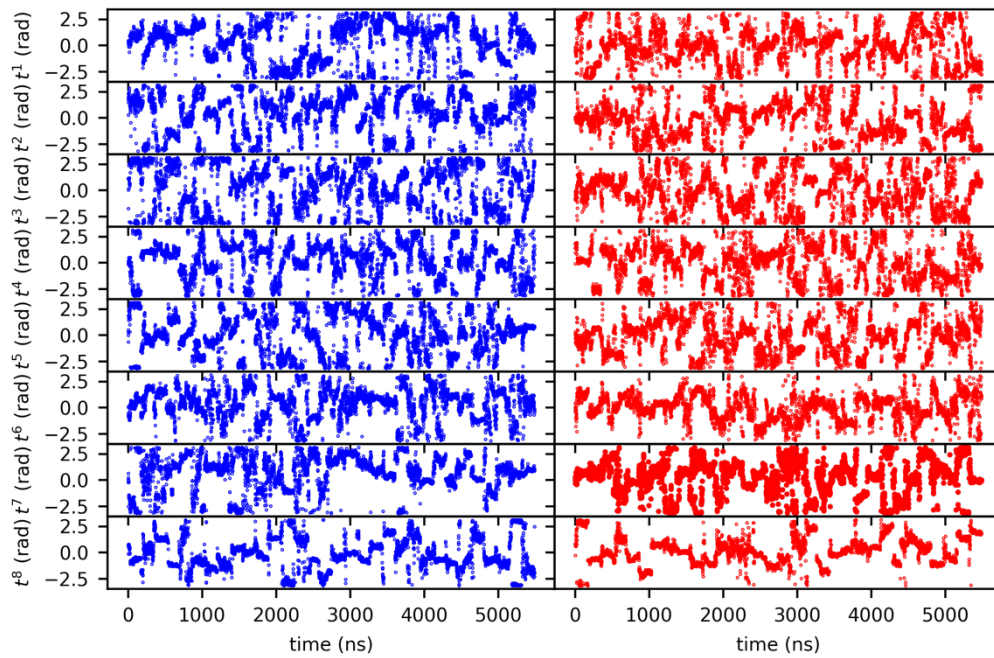
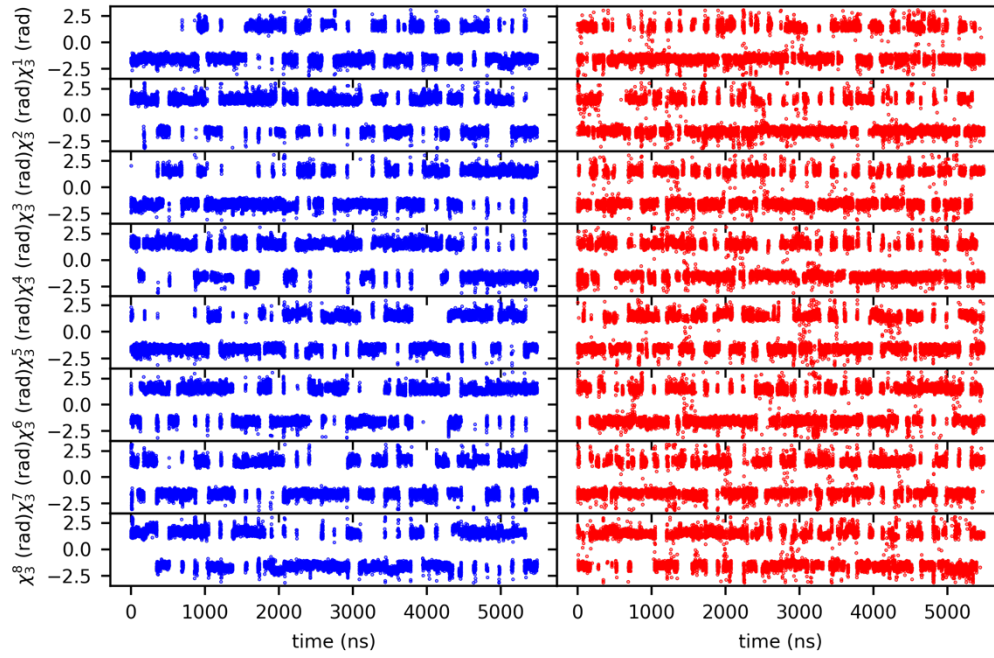


Figure S1. Metadynamics and clustering information. (A) Summary of all biased collective variables of the left side N-terminal tail and the respective unbiased ones on the right side N-terminal tail used in the analysis. (B) Dependence of number of clusters on the cut-off value used in the GROMOS clustering algorithm, using root-mean-square deviations of C α atoms in tails 3 and 4 of the left (biased) side. (C) Populations of the top 10 clusters for the two last 40 % chunks of the simulation. Error bars for (B) and (C) show the 95th percentiles over 20 separate clustering runs using 5000 frames sampled based on metadynamics weights from each part of the simulation.



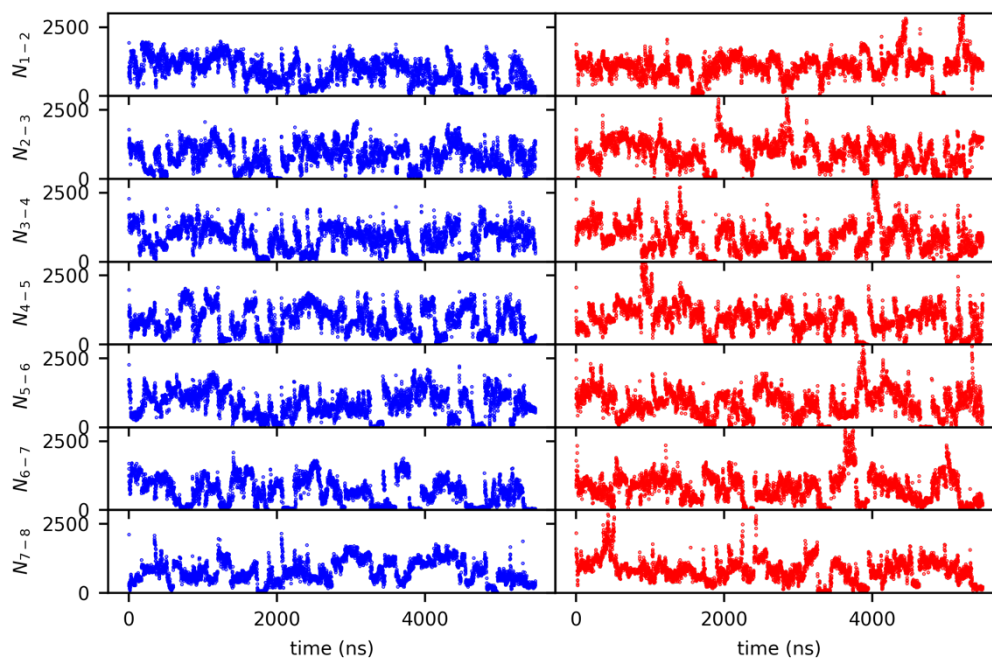


Figure S2. Assessment of the convergence of the simulations MEMMI-EMMI. Time evolution profiles for all unbiased (blue) and respective biased CVs (red).

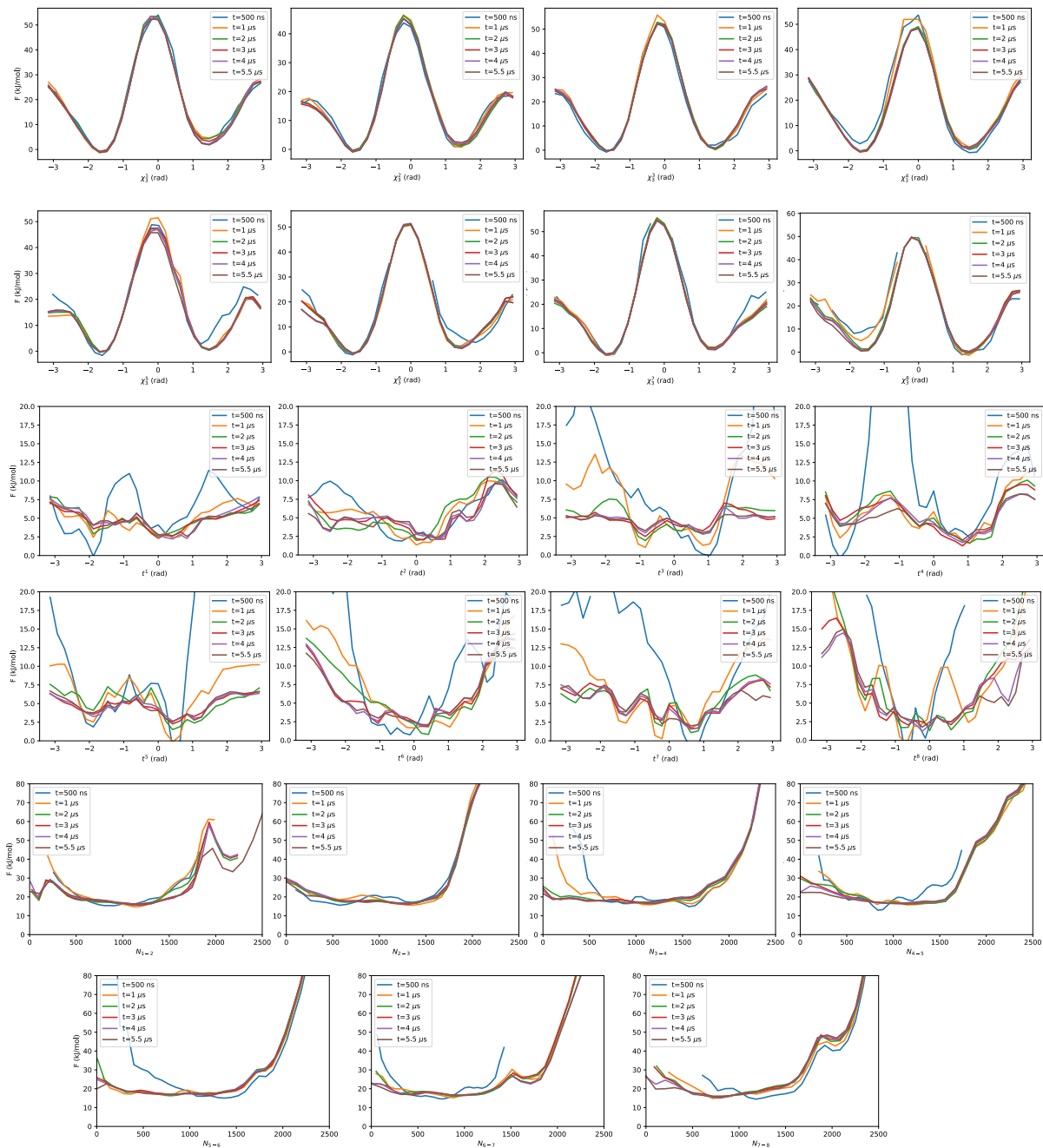


Figure S3. Assessment of the convergence of the MEMMI simulation. Free energy profiles for all biased collective variables for subsequent $1 \mu\text{s}$ increments of simulated time.

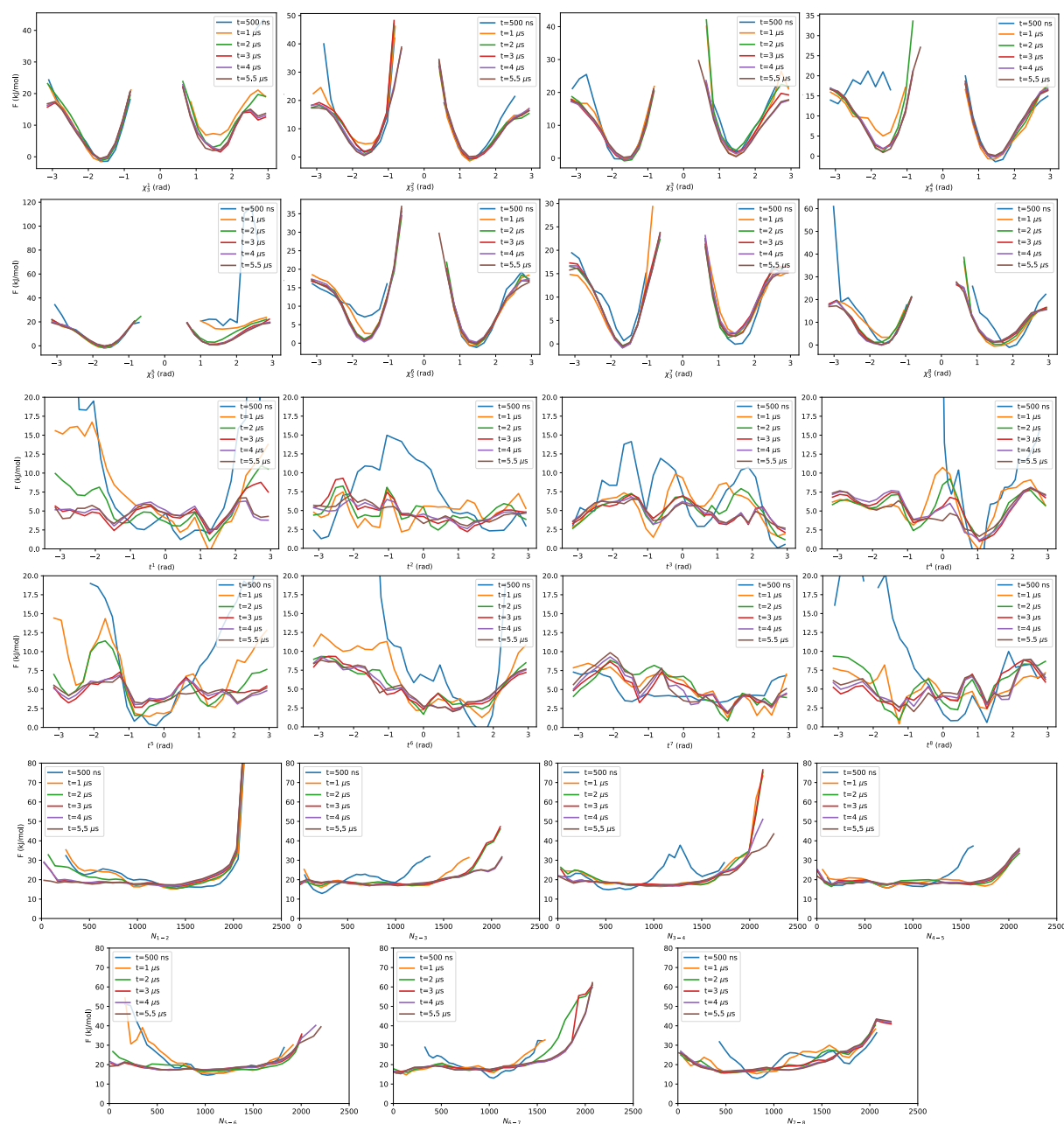


Figure S4. Assessment of the convergence of the EMMI simulation. Free energy profiles for all unbiased collective variables for subsequent 1 μ s increments of simulated time.

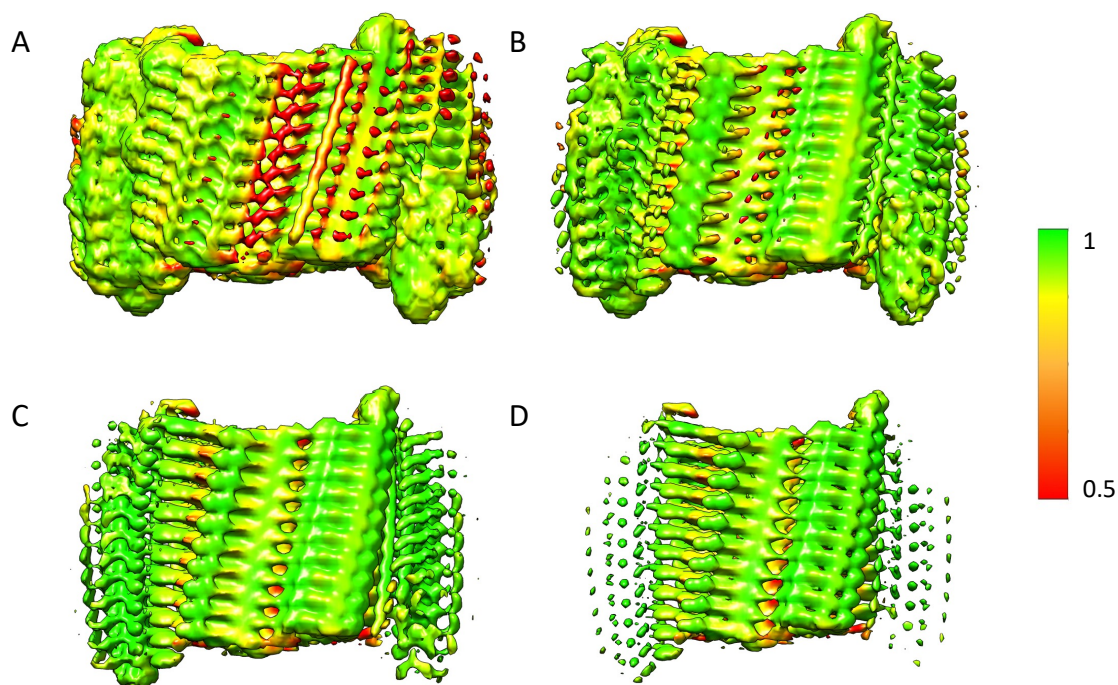


Figure S5. Assessment of the local correlation between experimental and calculated cryo-EM map. Local correlation of the cryo-EM map (EMD-EM-10669) with a map generated from the MEMMI ensemble of an IAPP amyloid fibril as a function of increasing strength of density (decreasing electron density thresholds): 1σ (A), 2σ (B), 3σ (C), and 4σ (D).

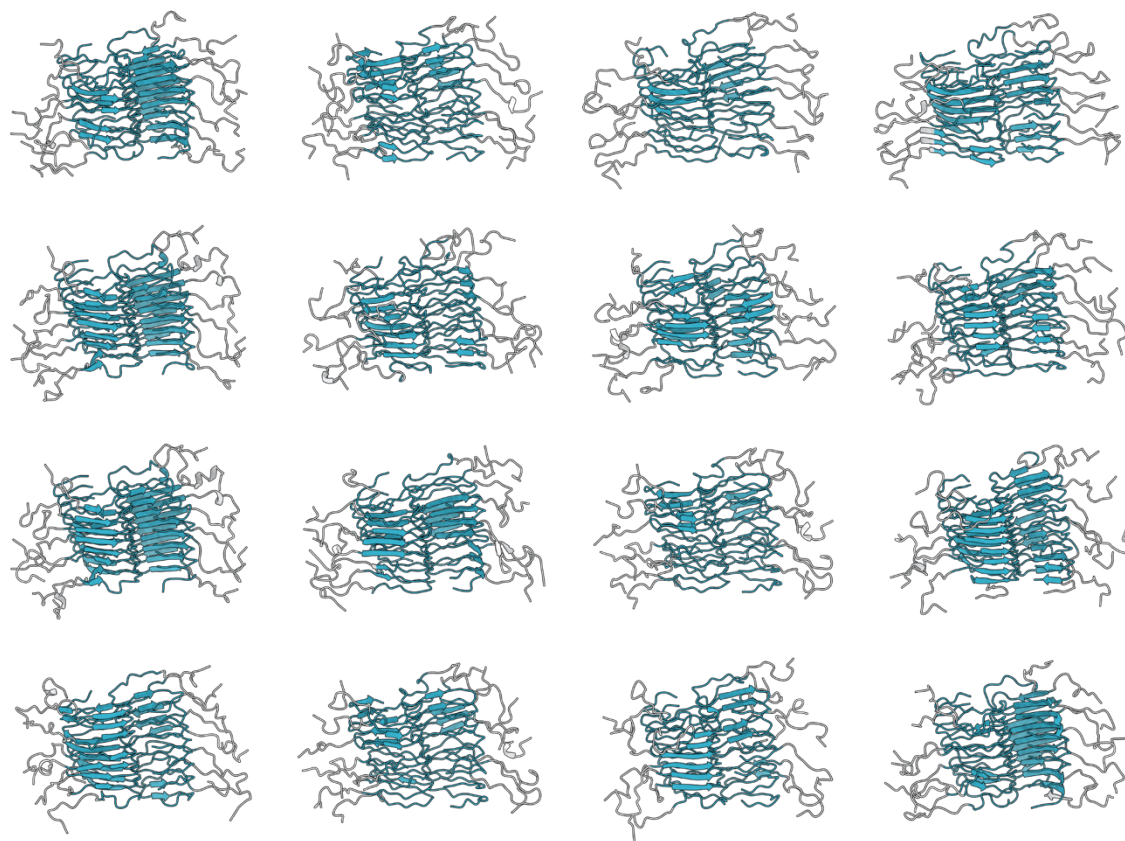


Figure S6. Individual conformations of the final structural ensemble. Sample of 16 structures from the ensemble shown in **Figure 2**.

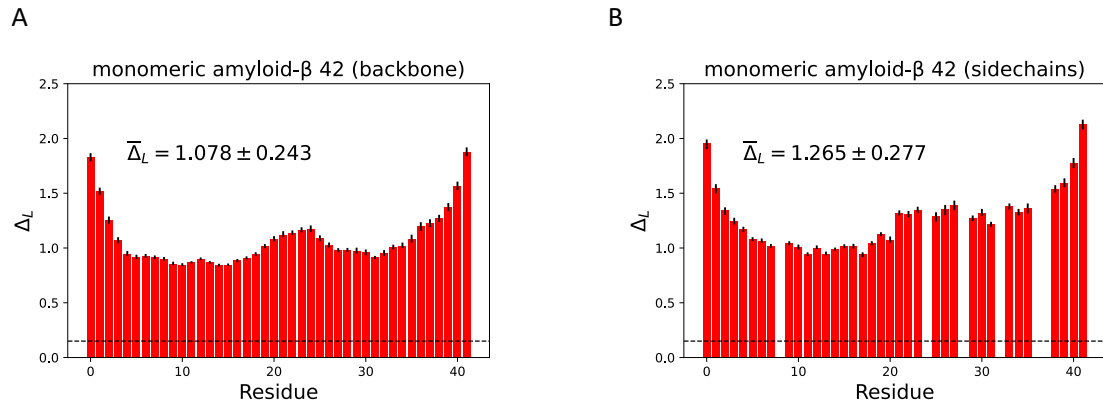


Figure S7. Lindemann parameters for the disordered monomeric amyloid-β 42 peptide computed from a previously published ensemble²⁷. Lindemann parameters calculated for the backbone (A) and side chains (B) with the liquid-solid transition and residue-mean indicated. Error bars indicate the 95th percentile of the mean of a bootstrap sample over all 5119 trajectories in the ensemble.