A small molecule stabilises the disordered native state of the Alzheimer's Aβ peptide

Thomas Löhr[1], Kai Kohlhoff[2], Gabriella T. Heller[1,3], Carlo Camilloni[4] and Michele Vendruscolo[1*]

[1]Department of Chemistry, University of Cambridge, Cambridge, UK.
[2]Google Research, Mountain View, CA, USA.
[3]Department of Structural and Molecular Biology, University College London, London, UK.
[4]Dipartimento di Bioscienze, Università degli Studi di Milano, Milano, Italy.

*Corresponding author(s). E-mail(s): mv245@cam.ac.uk;
Contributing authors: tl465@cam.ac.uk; kohlhoff@google.com;
g.heller@ucl.ac.uk; carlo.camilloni@unimi.it;

Abstract

The stabilisation of native states of proteins is a powerful drug discovery strategy. It is still unclear, however, whether this approach can be applied to intrinsically disordered proteins. Here we report a small molecule that stabilises the native state of the Aβ42 peptide, an intrinsically disordered protein fragment associated with Alzheimer's disease. We show that this stabilisation takes place by a dynamic binding mechanism, in which both the small molecule and the Aβ42 peptide remain disordered. This disordered binding mechanism involves enthalpically favourable local π-stacking interactions coupled with entropically advantageous global effects. These results indicate that small molecules can stabilise disordered proteins in their native states through transient non-specific interactions that provide enthalpic gain while simultaneously increasing the conformational entropy of the proteins.

Keywords: Markov model, molecular dynamics, disordered proteins

1 Introduction

Drug development for Alzheimer's disease has been a tremendous challenge in the past decades[1]. This condition is characterised by the formation of protein aggregates, such as fibrillar forms of the amyloid-β 42 peptide (Aβ42)[2, 3]. This protein fragment is intrinsically disordered, i.e. it does not form a single stable folded structure as a monomer, but instead exists in a dynamic equilibrium of states with transient local structure and fast transitions[4–12]. Many drug development efforts focused on aggregation-prone proteins such as Aβ42 attempt to target the already-formed fibril and/or the structurally elusive oligomeric species[13–15]. Other attempts aimed to identify small molecules capable of stabilising monomeric Aβ42 into a well-structured conformation[16–18] or generally interfering with the interaction of disordered proteins to structured partners by binding to their interfacing regions[19]. Since the most populated state of disordered proteins is conformationally highly heterogeneous, it has also been suggested that it may be more convenient to identify small molecules capable of stabilising this disordered state[20, 21]. The idea is that since the free energy landscape of disordered proteins is 'inverted' when compared with the funnel concept of folded proteins, with the disordered state as the free energy minimum and the ordered states exhibiting relatively high free energies [22], small molecules stabilising this minimum would be easier to develop, as they would not have to restructure the topology of the free energy landscape itself.

Independently from the strategy pursued, however, it is extremely challenging to characterize the binding mode of small molecules to disordered protein on an atomistic level. While some experimental methods such as nuclear magnetic resonance spectroscopy can provide quantitative information, it is often not sufficient to clearly understand the interactions and kinetics underlying the binding[20].

Molecular dynamics is one of the tools that can provide the necessary spatial and temporal resolution to study the interaction between disordered proteins and small molecules[20]. Together with Bayesian restraints from experimental data, molecular dynamics simulations have been used to characterize the thermodynamics of these binding modes in the case of the oncoprotein c-Myc[23] and Aβ42[5]. In the former study, urea was used as a control molecule to assess the sequence-specificity of the drug. In the latter case of Aβ42, we studied the interaction with the small molecule 10074-G5, and showed it was able to inhibit Aβ42 aggregation by binding the disordered monomeric form of the peptide. The interaction was characterized both experimentally, using various biophysical techniques, and computationally, using restrained molecular dynamics simulations with enhanced sampling. While in both systems the binding mode was found to be highly dynamic, a quantitative study of the kinetics was not possible due to the use of time-dependent restraints and biases during the simulation. The microscopic kinetics in form of contact lifetimes and autocorrelations could be especially instructive to fully

understand the origin of entropic and enthalpic stabilisation in these extremely dynamic binding events (Fig. 1)[21].

A quantitative study of the kinetics of these interactions may allow a more targeted approach to the design of both drugs and better experiments to probe their binding modalities. However, even with atomistic computational approaches, gaining insight into the kinetics, i.e. transition rates, relaxation constants, autocorrelations, and state lifetimes can be challenging. This is because in contrast to folded systems, the definition of states for disordered proteins is not always clear: due to the generally shallow free energy landscape state transitions may be fast, but not always distinct[6]. New developments in the theory of dynamical systems now allow an optimal state decomposition and transition operator to be learned using deep neural networks, for example using the VAMPNet framework[24, 25]. To acquire kinetic information for a system one would traditionally use a Markov state model[26, 27]: One first finds a suitable low-dimensional embedding of the system coordinates, followed by using a clustering algorithm to define microstates. Transitions between these can then be counted to build up statistics and thus construct a transition matrix. This matrix can then be coarse-grained to obtain macroscopic kinetics[28, 29].

Koopman-operator[30, 31] based models such as VAMPNet combine these two steps into a single function that can be approximated by a neural network and also yield a probabilistic state assignment in lieu of a discrete one[24, 25]. The former feature has the advantage of both simplifying the model construction process, as the hyperparameter search over various dimensionality reduction and clustering techniques is replaced by a simplified search over neural network parameters, and allowing a more accurate model due to the use of a single arbitrarily non-linear function compared to two steps that are heavily restricted in terms of search space. Probabilistic state assignments are inherently well suited to disordered proteins, as the typically shallow free energy basins and low barriers can be encoded with some ambiguity. This constrained VAMPNet approach was recently utilized by us to determine the kinetic ensemble of the disordered Aβ42 monomer[6].

Here, we use this technique to build kinetic ensembles of Aβ42 with 10074-G5 and urea as a control molecule to expand on our previous thermodynamic ensembles[5]. We compare the transition rates, lifetimes, and state populations with the previous kinetic ensemble of the Aβ42 monomer[6], and further characterize the atomic-level protein-drug interactions.

2 Results

2.1 Molecular dynamics simulations and soft Markov state models

We performed two explicit-solvent molecular dynamics simulations of Aβ42 with one molecule of urea and one molecule of 10074-G5, respectively. Both simulations were performed in multiple rounds of 1,024 trajectories on the

Google Compute Engine as described previously[6]. As before, we used a soft Markov state model approach using the constrained VAMPNet framework[24] to construct kinetic ensembles. The major advantages of this method, compared to regular discrete Markov state models, are the soft state definitions and the use of a single function mapping directly from arbitrary system coordinates to a state assignment probability, allowing for more optimal models. To aid our analysis, we added our previous simulation of Aβ42 with no additional molecules to our dataset. We refer to it as the *apo* ensemble[6]. We compared all ensembles using a decomposition into two states. In addition to being easier to interpret, this approach allows for a direct comparison of the slowest timescales in contrast to higher state-count models.

2.2 Computational and experimental validation

Constructing a kinetic ensemble using the constrained VAMPNet approach requires choosing the number of states and the model lag time. The latter is a critical parameter that needs to be chosen such that the model can accurately resolve both long and short timescales. This can be done by plotting the dependence of the slowest relaxation timescales on the lag time (Extended Fig. A1). A stricter measure is the Chapman-Kolmogorov test, comparing multiple applications of the Koopman operator estimated at a certain lag time τ with a Koopman operator estimated at a multiple of this lag time $n\tau$ (Extended Fig. A2, Methods)[32]. To evaluate sampling convergence, we visualized the dependence of the mean relaxation timescales on the number of trajectories used to evaluate these timescales (Extended Fig. A5). With sufficient sampling of kinetics, we would expect the global timescales to be unchanged within error. Experimental validation was performed by comparing back-calculated chemical shifts to ones from experiment[33]. Because the small molecule 10074-G5 only has minor effects on the chemical shifts of Aβ42[23], well below the error of the forward model, we used the chemical shift dataset from the apo ensemble as a point of comparison (Extended Fig. A3).

2.3 10074-G5 has minor impact on ensemble-averaged structural properties of Aβ42

To evaluate the influence of 10074-G5 and urea on the structural conformations of Aβ42, we calculated state-averaged contact maps and secondary structure content for each state of all ensembles (Extended Fig. A4a-c). In all cases we find a state decomposition into a more extended state with few inter-residue contacts, and a slightly more compact form with a higher number of local backbone interactions. We will refer to these as the compact and extended states, respectively. The addition of a small molecule has little effect on the formation of contacts and other structural motifs. This finding is consistent with our recent experimental thermodynamic and kinetic characterization of this interaction, and the absence of strong chemical shift perturbations in the holo ensemble[5].

2.4 10074-G5 and urea decelerate the formation of more compact states

Compared to the previously published kinetic ensemble of the apo form of Aβ42, the kinetic ensembles in the presence of both urea and 10074-G5 show a deceleration of more compact state formation (Fig. 2). Notably, the transition from the more compact form to the more extended state is unaffected. This change is also mirrored in the state populations, which exhibit a strong shift towards the extended state. We note that even though there are strong changes in the state populations, the ensemble-averaged contact maps are very similar (Extended Fig. A4a-c). This is likely due to the high sensitivity of the VAMPNet method to minor changes in free energy barrier regions. These will have a significant effect on the kinetics and thus state populations, but not on the ensemble averaged structure due to the relatively low thermodynamic weight[34]. While the lifetimes of the extended states increase, the ones for the more compact form are unchanged within model error. We can thus conclude that within our model, the small molecule has a strong effect on the contact formation rates, but no influence on the contact dissociation rates.

2.5 Small molecules shift the system to more entropically stable states by short-lived local interactions

To evaluate the impact of 10074-G5 on the conformational space of Aβ42, we calculated the Ramachandran and state entropy for all ensembles, as well as the autocorrelation of sidechain χ_1 dihedral angles (Fig. 3). The Ramachandran entropy can indicate relative flexibility of the backbone, thus revealing potential regions of dynamic changes as a result of interactions between the peptide and small molecule[5]. Resolving this change in the entropy over residues (Fig. 3a) indicates strong increases in the relatively hydrophobic C-terminal region of Aβ42. This entropy increase is confirmed globally by the sum of the entropies over all residues (Fig. 3b). As an alternative metric, we also calculated the entropy in the state assignments (Fig. 3c), this can be thought of as indicating the overall ambiguity in the state definition. Again, we find a relatively strong increase in the conformational entropy of Aβ42 for the ensemble with 10074-G5, and only minor increases for urea. These results are in agreement with our previous observations from simulations of the equilibrium ensembles in that the presence of 10074-G5 increases the conformations available to Aβ42, via the entropic expansion mechanism[5, 21].

To better understand the impact of the small molecule on local kinetics we calculated the autocorrelation of the sidechain χ_1 dihedral angles (Fig. 3). We see an increase in the autocorrelation, specifically for aromatic residues and M35, indicating a slowing of side chain rotations. This suggests that despite an increase in the backbone entropy, the peptide is able to visit many locally stable states, resulting in local enthalpic stabilisation.

2.6 Interactions of 10074-G5 with Aβ42 are dominated by π-stacking and other electrostatic effects

To better understand the origin of the global and local effects of 10074-G5 on the ensemble we analysed the interactions on a residue and atomistic level (Fig. 4). While the probability of forming a contact between the small molecule and a residue shows certain mild preferences (Fig. 4a), these become more evident when looking at the lifetimes of these contacts (Fig. 4b). Here, the longest contacts are formed by π-stacking with certain aromatic residues (F4, Y10, F19, F20) and by interactions with M35. This result also explains the reduction in side-chain rotations for these residues (Fig. 3d). On an atomistic level the π-π interactions exhibit some anisotropy (Extended Fig. A6). The importance of the nitro- and benzofurazan fragments is also highlighted. Finally, we also investigated the conditionality of π-π interactions, i.e., if we see an interaction between the molecule and residue i, what is the probability of also observing an interaction with residue j (Fig. 4e-g)? The probabilities here are uniformly low but indicate a slight preference (13 %) for a triple π-stack involving the terminal aromatic ring of 10074-G5 and residues F19 and F20 of Aβ42. The importance of π-π stacking interactions was also noted in a computational study on the interactions of small molecules with α-synuclein[35].

These results indicate that this disordered binding mechanism operates on two levels whereby local enthalpically favourable interactions coupled with global entropically advantageous effects. The local interactions are predominantly of electrostatic nature and result in a reduction of sidechain rotations on specific residues. At the same time, these interactions also allow the exploration of more backbone conformations, thus resulting in a net entropy increase for Aβ42. This influence expands into the global kinetics of the system, significantly slowing the formation of local structure.

3 Discussion

The results outlined above present a possible example of the previously proposed entropic expansion mechanism for the binding of small molecules to disordered proteins[21, 36]. This mechanism is distinct from the entropic collapse and folding-on-binding mechanisms[37, 38]. The concept of disordered binding is difficult to probe, as the tools suitable to detecting small changes in the conformational ensemble of disordered proteins are still in their infancy[20]. Nuclear magnetic resonance experiments can provide information, but it should usually be interpreted in a structural framework, necessitating molecular simulations with ensemble-averaged restraints[39], or re-weighting approaches[40]. This constraint causes issues whenever we are also interested in kinetics, as by enhancing the sampling we modify the natural dynamics of the system. Nevertheless, an approach to incorporate ensemble-averaged experimental measurements into Koopman models has recently been proposed[41]. Neither is it generally possible to use enhanced sampling methods to study

kinetics without having some *a priori* knowledge of the system states. A framework allowing the incorporation of experimental data into a kinetic model and also allowing the use of enhanced sampling methods such as metadynamics[42], without prior knowledge of states, would make the study of these systems easier and more accurate.

As we have shown, a kinetic model is crucial to fully explain the nature of these binding interactions. This is in part due to the ability to use the slowest timescales of the system to reliably define metastable states, something that is notoriously difficult for disordered proteins without access to the time dimension. This clustering alone is already sensitive enough to reveal differences between systems that are nearly invisible when comparing ensemble-averaged results and more conventional clustering methods[5]. Increases in local autocorrelation and global state transitions might be seen as indicators of both local enthalpic stabilisation and global entropic expansion. The former result hints at the possibility of designing small molecules that exhibit high specificity, as the global entropic stabilisation effect may be due to transient, local, enthalpically-favourable interactions[21]. The two level global entropy – local enthalpy effect becomes especially visible when looking at the timescales: The slowest state transitions of the protein are on the order of microseconds, while the local, enthalpically-favourable π-π interaction lifetimes are no longer than tens of nanoseconds.

The observed binding mechanism also identify π-π stacking interactions as a major driving force. Similar effects have been observed for the binding of another small molecule, fasudil, and α-synuclein, which is also intrinsically disordered[35]. We note that while that study proposed a 'shuttling model' mechanism to explain the diffusion of the small molecule on the α-synuclein surface, here we demonstrate the stabilisation of a native state of a disordered protein by a disordered binding mechanism. The π-π stacking phenomenon also plays a major role in liquid-liquid phase separation[43], suggesting a possible link between the effect of these small molecules and the hypothesized state of some proteins in a crowded environment. For molecular simulations, the force field may present a barrier in studying π-π interactions in more detail. This is because these interactions are not explicitly part of the potential, but only approximated with a combination of electrostatic and hydrophobic terms[44]. Polarizable force fields may offer a computationally affordable alternative that could more accurately model this type of binding[45].

Looking forward, it may become possible to pursue a drug discovery strategy for disordered proteins based on the stabilisation of their native states through the disordered binding mechanism that we have described here. This strategy would extend an approach to disordered proteins that has already proven successful for folded proteins[46], and would have the advantage of maintaining the proteins in their native functional states.

Springer Nature 2021 LATEX template

4 Methods

4.1 Details of the simulations

All simulations were performed on the Google Compute Engine with `n1-highcpu-8` preemptible virtual machine instances, equipped with eight Intel Haswell CPU cores and 7.2 GB of RAM. Molecular dynamics simulations were performed with GROMACS 2018.1[47], with 1,024 starting structures sampled from the previously performed apo simulation[6] using the Koopman model weights. Each conformation was placed in the center of a rhombic dodecahedron box with a volume of 358 nm³, and the corresponding small molecule was placed in the corner of the box. The force field parameters for urea were taken from the CHARMM22* force field[48] and the ones for 10074-G5 were computed using the Force Field Toolkit (FFTK)[49] and Gaussian 09[50], as described previously[5]. The systems were then solvated using between 11,698 (11,707) and 11,740 (11,749) water molecules. Both systems were minimized using the steepest descent method to a maximum target force of 1,000 kJ/-mol/nm. Both systems were subsequently equilibrated, first over a time range of 500 ps in the NVT ensemble using the Bussi thermostat[51] and then over another 500 ps in the NPT ensemble using Berendsen pressure coupling[52]. In both equilibrations position restraints were placed on all heavy atoms. All production simulations were performed using 2 fs time steps in the NVT ensemble using the CHARMM22*[48] force field and TIP3P water model[53] at 278 K and LINCS constraints[54] on all bonds. Electrostatic interactions were modelled using the Particle-Mesh-Ewald approach[55] with a short-range cutoff of 1.2 nm. All simulations used periodic boundary conditions. We again used the fluctuation-amplification of specific traits (FAST) approach[56] to adaptive sampling, with clustering performed through time-lagged independent component analysis (TICA)[57, 58] using a lag time of 5 ns and Cα distances fed to the k-means clustering algorithm[59] to yield 128 clusters. 1,024 new structures were then sampled from these clusters based on maximizing the deviation to the mean Cα distance matrix for each cluster and maximizing the sampling of the existing clusters, using a balance parameter of $\alpha = 1.0$, with all amino acids weighted equally. This approach was performed once for each ensemble, however we also chose to perform 32 additional long-trajectory simulations for the 10074-G5 ensemble, yielding a total of 2,079 trajectories for the latter, and 2,048 trajectories for the urea ensemble. The total simulated times were 306 μs and 279 μs for the 10074-G5 and urea ensembles, respectively. The shortest and longest trajectories for 10074-G5 (urea) were 21 (24) ns and 1,134 (196) ns. All trajectories were subsampled to 250 ps timesteps for further analysis.

4.2 Details of the neural networks

State decomposition and kinetic model construction was performed using the constrained VAMPNet approach[24, 25], using the same method described

previously[6]. We again chose flattened inter-residue nearest-neighbour heavy-atom distance matrices as inputs, resulting in 780 input dimensions. We used the self-normalizing neural network architecture[60] with scaled-exponential linear units, normal LeCun weight initialization[61] and alpha dropout. We chose an output dimension of 2, thus yielding a soft two-state assignment. The datasets were prepared by first creating a test dataset by randomly sampling 10 % of the frames. In the case of 10074-G5 we excluded all frames in which the closest distance between the small molecule and peptide was higher than 0.5 nm. We then created 20 randomized 9:1 train-validation splits to allow a model error estimate. Training was performed by using three trials for each train-validation split and picking the best performing model based on the VAMP2 score[31] of the test set. We implemented the model using Keras 2.2.4[62] with the Tensorflow 2.1.0[63] backend. We chose the following model hyperparameters based on two successive coarser and finer grid searches: A network lag time of 5 ns, a layer width of 512 nodes, a depth of 2 layers, an L2 regularization strength of 10^{-7} and a dropout of 0.05. Training was performed in 10,000 frame pairs using the Adam minimizer[64] with a learning rate of 0.05, $\beta_2 = 0.99$ and epsilon of 10^{-4}, and an early stopping criterion of a minimum validation score improvement of 10^{-3} over the last five epochs. For the constrained part of the model, we reduced the learning rate by a factor of 0.02. We used a single Google Compute Engine instance with 12 Intel Haswell cores, 78 GB of RAM, and an NVidia Tesla V100 GPU.

4.3 Details of the kinetic analysis

After training, VAMPNet yields a state assignment vector $\chi(\mathbf{x}_t)$ for each frame $\mathbf{x}_t$ of the ensemble. Based on this vector, we can calculate state averages $\langle A_i \rangle$ for any observable $A(\mathbf{x}_t)$:

$$\langle A_i \rangle = \left(\sum_{t=1}^{T} \chi_i(\mathbf{x}_t) \right)^{-1} \sum_{t=1}^{T} \chi_i(\mathbf{x}_t) A(\mathbf{x}_t) \tag{1}$$

Here, i is the corresponding state and the sum runs over all time steps. To calculate an ensemble average $\langle A \rangle$, one first calculates a weight w_t for each frame using the model equilibrium distribution π:

$$w_t = \frac{\langle \chi(\mathbf{x}_t) \mid \pi \rangle}{\sum_{t=1}^{T} \langle \chi(\mathbf{x}_t) \mid \pi \rangle}, \tag{2}$$

which leads to the ensemble average

$$\langle A \rangle = \sum_{t=1}^{T} w_t A(\mathbf{x}_t). \tag{3}$$

Because each trained model will classify the states in an arbitrary order, we need to sort the state assignment vectors based on state similarity. We did this by comparing the state-averaged contact maps using root-mean-square

deviation as a metric, and grouping states based on the lowest value. Any deviations are thus accounted for in the overall model error.

4.4 Model validation

The Koopman matrix $\mathbf{K}(\tau)$ is given directly by the neural network model, along with the equilibrium distribution π. We validated our models using the Chapman-Kolmogorov test:

$$\mathbf{K}(n\tau) \approx \mathbf{K}^n(\tau) \tag{4}$$

where τ is the model lag time, and $n\tau$ is a low integer-multiple of the lag time. The model should therefore behave the same way whether we estimate it at a longer lag time or repeatedly apply the transfer operator. We first estimate a suitable lag time τ by plotting the relaxation timescales over the chosen lag time. The lag time τ should be chosen to be as small as possible, but large enough to not have any impact on the longer relaxation timescales, which represent the slowest motions of the system. The temporal resolution of the model is thus given by this lag time. The relaxation timescales t_i are calculated from the eigenvalues λ_i of the Koopman matrix $\mathbf{K}(\tau)$ as follows:

$$t_i = \frac{-\tau}{\log |\lambda_i|} \tag{5}$$

We can similarly compute the state lifetimes $\bar{t}_i$ from the diagonal elements of the Koopman matrix $\mathbf{K}(\tau)_{ii}$ using:

$$\bar{t}_i = \frac{-\tau}{\log \mathbf{K}(\tau)_{ii}} \tag{6}$$

4.5 Experimental validation

We backcalculated the nuclear magnetic resonance chemical shifts using the CamShift algorithm[33] as implemented in PLUMED 2.4.1[65, 66]. We again used the same ensemble averaging procedure described above.

4.6 Errors

Errors are calculated over all trained neural network models. To obtain a more meaningful estimate, we only consider frames that were part of the bootstrap training sample of the corresponding model, i.e., one of the 20 models described above. The reported averages are the mean, and the errors the 95th percentiles over all 20 models, unless reported otherwise.

Declarations

4.7 Code availability

Analysis code and notebooks are available from https://github.com/vendruscolo-lab/ab42-g5-ensemble and https://zenodo.org/record/5659241.

4.8 Data availability

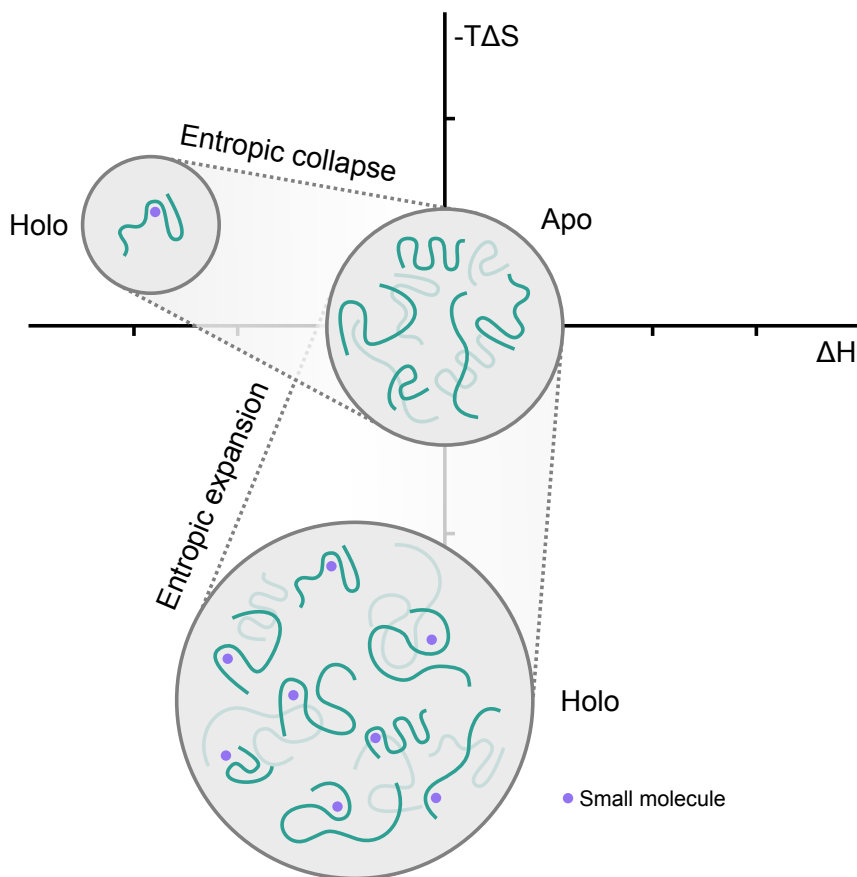Subsampled trajectory and intermediate data, as well as the trained neural network weights and analysis notebooks, are available from https://zenodo.org/record/5659241.

Fig. 1 Illustration of two different native state stabilisation mechanisms of disordered proteins. The interaction with a small molecule can result in a reduction or increase of conformational space of the protein, thus resulting in a positive or negative entropic contribution to the binding free energy. A loss of entropic native state stabilisation will often be compensated with a stronger enthalpic binding affinity, while an increase in entropy often requires more dynamic and thus weaker binding.
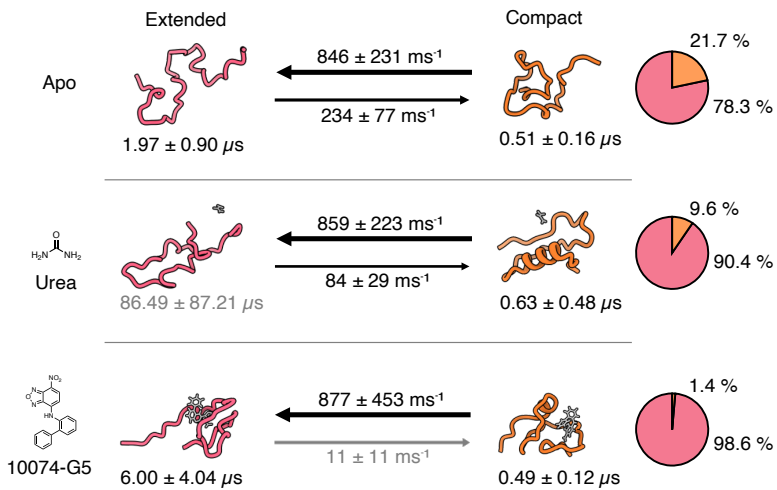
Fig. 2 Impact of small molecules on the state transition rates, state lifetimes, and populations. The arrows indicate the mean state transition rates, the number below the representative structures is the mean state lifetime, the pie charts show the mean state populations. Errors are the standard deviations of the bootstrap sample of the mean over all 20 models.
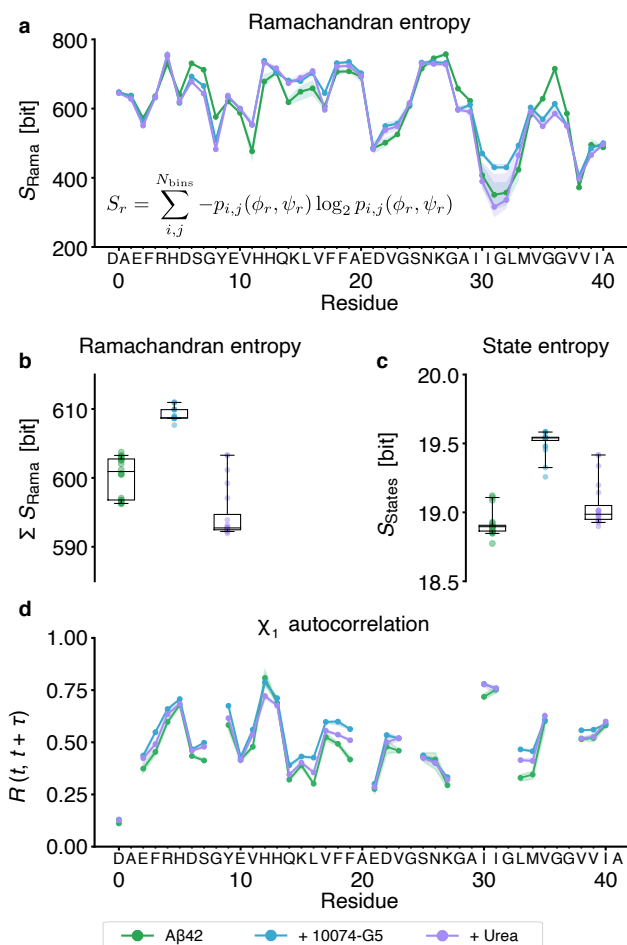
Fig. 3 Effect of small molecules on conformational and state entropy of Aβ42, showing that 10074-G5 increases the conformational entropy of the peptide. **a** Ramachandran entropy, i.e. information entropy over the distribution of φ and ψ backbone dihedral angle conformations, using 100 bins. **b** Sum of the Ramachandran entropies over all residues for all ensembles. **c** State entropy, i.e., the population-weighted mean of the information entropy of each set of state assignments. More ambiguity in the state assignments leads to a correspondingly higher state entropy. **d** Autocorrelation of all sidechain χ_1 dihedral angles with a lag time of $\tau = 5$ ns. Shaded areas in **a** and **d** indicate the 95th percentiles of the bootstrap sample of the mean over all 20 models. Whiskers and boxes in **b** and **c** indicate the 95th percentiles and quartiles of the bootstrap sample of the mean over all 20 models, respectively.
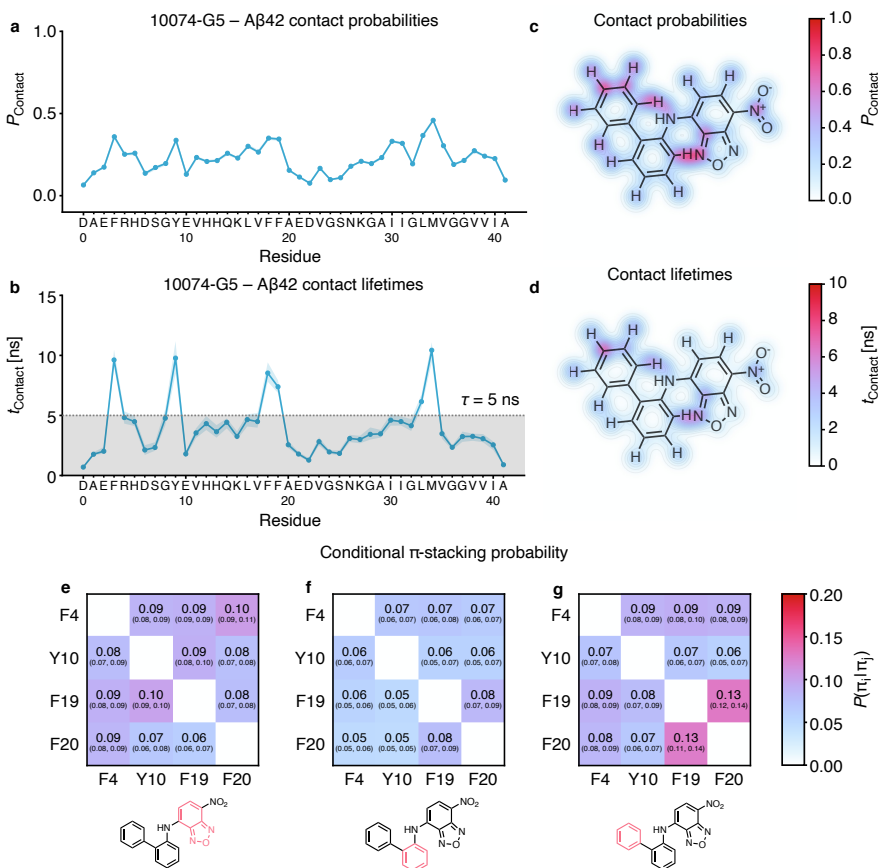
Fig. 4 Residue- and atomic level interactions of 10074-G5 with Aβ42 showing regions on the small molecule responsible for binding. **a** Contact probabilities of 10074-G5 and Aβ42 with a cut-off of 0.45 nm. **b** Lifetimes of these contacts, estimated using a Markov state model for each contact formation with a lag time of $\tau = 5$ ns, indicated with grey shading. Coloured shaded areas in **a** and **b** indicate the 95th percentiles of the bootstrap sample of the mean over all 20 models. **c, d** Contact probabilities and lifetimes of each atom of 10074-G5 with any residue of Aβ42. **e, f, g** Conditional probability of forming a π-stacking interaction, given an existing π-stacking interaction for all aromatic groups in the small molecule. Tuples indicate the 95th percentiles of the bootstrap sample of the mean over all 20 models.

References

[1] Cummings, J., Lee, G., Zhong, K., Fonseca, J., Taghva, K.: Alzheimer's disease drug development pipeline: 2021 **7**(1), 12179. https://doi.org/10.1002/trc2.12179. Accessed 2021-09-27

[2] Hampel, H., Hardy, J., Blennow, K., Chen, C., Perry, G., Kim, S.H., Villemagne, V.L., Aisen, P., Vendruscolo, M., Iwatsubo, T., Masters, C.L., Cho, M., Lannfelt, L., Cummings, J.L., Vergallo, A.: The

Amyloid-β Pathway in Alzheimer's Disease, 1–23. https://doi.org/10.1038/s41380-021-01249-0. Accessed 2021-10-25

[3] Hardy, J.A., Higgins, G.A.: Alzheimer's disease: The amyloid cascade hypothesis **256**(5054), 184–185 1566067. https://doi.org/10.1126/science.1566067. Accessed 2019-11-26

[4] Ball, K.A., Phillips, A.H., Nerenberg, P.S., Fawzi, N.L., Wemmer, D.E., Head-Gordon, T.: Homogeneous and Heterogeneous Tertiary Structure Ensembles of Amyloid-β Peptides **50**(35), 7612–7628. https://doi.org/10.1021/bi200732x. Accessed 2019-10-10

[5] Heller, G.T., Aprile, F.A., Michaels, T.C.T., Limbocker, R., Perni, M., Ruggeri, F.S., Mannini, B., Löhr, T., Bonomi, M., Camilloni, C., Simone, A.D., Felli, I.C., Pierattelli, R., Knowles, T.P.J., Dobson, C.M., Vendruscolo, M.: Small-molecule sequestration of amyloid-β as a drug discovery strategy for Alzheimer's disease **6**(45), 5924 33148639. https://doi.org/10.1126/sciadv.abb5924. Accessed 2020-11-21

[6] Löhr, T., Kohlhoff, K., Heller, G.T., Camilloni, C., Vendruscolo, M.: A kinetic ensemble of the Alzheimer's Aβ peptide **1**(1), 71–78. https://doi.org/10.1038/s43588-020-00003-w. Accessed 2021-01-25

[7] Meng, F., Bellaiche, M.M.J., Kim, J.-Y., Zerze, G.H., Best, R.B., Chung, H.S.: Highly Disordered Amyloid-β Monomer Probed by Single-Molecule FRET and MD Simulation **114**(4), 870–884. https://doi.org/10.1016/j.bpj.2017.12.025. Accessed 2019-10-10

[8] Nasica-Labouze, J., Nguyen, P.H., Sterpone, F., Berthoumieu, O., Buchete, N.-V., Coté, S., De Simone, A., Doig, A.J., Faller, P., Garcia, A., Laio, A., Li, M.S., Melchionna, S., Mousseau, N., Mu, Y., Paravastu, A., Pasquali, S., Rosenman, D.J., Strodel, B., Tarus, B., Viles, J.H., Zhang, T., Wang, C., Derreumaux, P.: Amyloid β Protein and Alzheimer's Disease: When Computer Simulations Complement Experimental Studies **115**(9), 3518–3563. https://doi.org/10.1021/cr500638n. Accessed 2019-10-10

[9] Paul, A., Samantray, S., Anteghini, M., Khaled, M., Strodel, B.: Thermodynamics and kinetics of the amyloid-β peptide revealed by Markov state models based on MD data in agreement with experiment **12**(19), 6652–6669. https://doi.org/10.1039/D0SC04657D. Accessed 2021-09-27

[10] Roche, J., Shen, Y., Lee, J.H., Ying, J., Bax, A.: Monomeric Aβ 1–40 and Aβ 1–42 Peptides in Solution Adopt Very Similar Ramachandran Map Distributions That Closely Resemble Random Coil **55**(5), 762–775. https://doi.org/10.1021/acs.biochem.5b01259. Accessed 2019-10-10

[11] Rosenman, D.J., Connors, C.R., Chen, W., Wang, C., García, A.E.: Aβ Monomers Transiently Sample Oligomer and Fibril-Like Configurations: Ensemble Characterization Using a Combined MD/NMR Approach **425**(18), 3338–3359. https://doi.org/10.1016/j.jmb.2013.06.021. Accessed 2019-10-10

[12] Sgourakis, N.G., Merced-Serrano, M., Boutsidis, C., Drineas, P., Du, Z., Wang, C., Garcia, A.E.: Atomic-Level Characterization of the Ensemble of the Aβ(1–42) Monomer in Water Using Unbiased Molecular Dynamics Simulations and Spectral Algorithms **405**(2), 570–583. https://doi.org/10.1016/j.jmb.2010.10.015. Accessed 2020-02-10

[13] Habchi, J., Arosio, P., Perni, M., Costa, A.R., Yagi-Utsumi, M., Joshi, P., Chia, S., Cohen, S.I.A., Müller, M.B.D., Linse, S., Nollen, E.A.A., Dobson, C.M., Knowles, T.P.J., Vendruscolo, M.: An anticancer drug suppresses the primary nucleation reaction that initiates the production of the toxic Aβ42 aggregates linked with Alzheimer's disease **2**(2), 1501244. https://doi.org/10.1126/sciadv.1501244. Accessed 2019-10-10

[14] Habchi, J., Chia, S., Limbocker, R., Mannini, B., Ahn, M., Perni, M., Hansson, O., Arosio, P., Kumita, J.R., Challa, P.K., Cohen, S.I.A., Linse, S., Dobson, C.M., Knowles, T.P.J., Vendruscolo, M.: Systematic development of small molecules to inhibit specific microscopic steps of Aβ42 aggregation in Alzheimer's disease **114**(2), 200–208. https://doi.org/10.1073/pnas.1615613114. Accessed 2019-10-10

[15] Lieblein, T., Zangl, R., Martin, J., Hoffmann, J., Hutchison, M.J., Stark, T., Stirnal, E., Schrader, T., Schwalbe, H., Morgner, N.: Structural rearrangement of amyloid-β upon inhibitor binding suppresses formation of Alzheimer's disease related oligomers **9**, 59306. https://doi.org/10.7554/eLife.59306. Accessed 2021-09-27

[16] Sinha, S., Lopes, D.H.J., Du, Z., Pang, E.S., Shanmugam, A., Lomakin, A., Talbiersky, P., Tennstaedt, A., McDaniel, K., Bakshi, R., Kuo, P.-Y., Ehrmann, M., Benedek, G.B., Loo, J.A., Klärner, F.-G., Schrader, T., Wang, C., Bitan, G.: Lysine-Specific Molecular Tweezers Are Broad-Spectrum Inhibitors of Assembly and Toxicity of Amyloid Proteins **133**(42), 16958–16969. https://doi.org/10.1021/ja206279b. Accessed 2021-09-14

[17] Zhu, M., De Simone, A., Schenk, D., Toth, G., Dobson, C.M., Vendruscolo, M.: Identification of small-molecule binding pockets in the soluble monomeric form of the Aβ42 peptide **139**(3), 035101. https://doi.org/10.1063/1.4811831. Accessed 2019-10-10

[18] Tóth, G., Gardai, S.J., Zago, W., Bertoncini, C.W., Cremades, N., Roy, S.L., Tambe, M.A., Rochet, J.-C., Galvagnion, C., Skibinski, G.,

Finkbeiner, S., Bova, M., Regnstrom, K., Chiou, S.-S., Johnston, J., Callaway, K., Anderson, J.P., Jobling, M.F., Buell, A.K., Yednock, T.A., Knowles, T.P.J., Vendruscolo, M., Christodoulou, J., Dobson, C.M., Schenk, D., McConlogue, L.: Targeting the Intrinsically Disordered Structural Ensemble of α-Synuclein by Small Molecules as a Potential Therapeutic Strategy for Parkinson's Disease **9**(2), 87133. https://doi.org/10.1371/journal.pone.0087133. Accessed 2019-10-10

[19] Xu, Y., Shi, J., Yamamoto, N., Moss, J.A., Vogt, P.K., Janda, K.D.: A credit-card library approach for disrupting protein–protein interactions **14**(8), 2660–2673. https://doi.org/10.1016/j.bmc.2005.11.052. Accessed 2021-09-14

[20] Heller, G.T., Aprile, F.A., Vendruscolo, M.: Methods of probing the interactions between small molecules and disordered proteins **74**(17), 3225–3243. https://doi.org/10.1007/s00018-017-2563-4. Accessed 2019-10-10

[21] Heller, G.T., Sormanni, P., Vendruscolo, M.: Targeting disordered proteins with small molecules using entropy **40**(9), 491–496. https://doi.org/10.1016/j.tibs.2015.07.004. Accessed 2019-10-10

[22] Granata, D., Baftizadeh, F., Habchi, J., Galvagnion, C., De Simone, A., Camilloni, C., Laio, A., Vendruscolo, M.: The inverted free energy landscape of an intrinsically disordered peptide by simulations and experiments **5**(1), 15449. https://doi.org/10.1038/srep15449. Accessed 2019-10-10

[23] Heller, G.T., Aprile, F.A., Bonomi, M., Camilloni, C., De Simone, A., Vendruscolo, M.: Sequence Specificity in the Entropy-Driven Binding of a Small Molecule and a Disordered Peptide **429**(18), 2772–2779. https://doi.org/10.1016/j.jmb.2017.07.016. Accessed 2019-10-10

[24] Mardt, A., Pasquali, L., Noé, F., Wu, H.: Deep learning Markov and Koopman models with physical constraints. Proceedings of Machine Learning Research, vol. 107, pp. 451–475. PMLR. http://proceedings.mlr.press/v107/mardt20a.html

[25] Mardt, A., Pasquali, L., Wu, H., Noé, F.: VAMPnets for deep learning of molecular kinetics **9**(1), 5. https://doi.org/10.1038/s41467-017-02388-1. Accessed 2019-10-10

[26] Chodera, J.D., Noé, F.: Markov state models of biomolecular conformational dynamics **25**, 135–144. https://doi.org/10.1016/j.sbi.2014.04.002. Accessed 2019-10-10

Springer Nature 2021 LATEX template

[27] Husic, B.E., Pande, V.S.: Markov State Models: From an Art to a Science **140**(7), 2386–2396. https://doi.org/10.1021/jacs.7b12191. Accessed 2019-10-10

[28] Noé, F., Wu, H., Prinz, J.-H., Plattner, N.: Projected and hidden Markov models for calculating kinetics and metastable states of complex molecules **139**(18), 184114. https://doi.org/10.1063/1.4828816. Accessed 2019-10-10

[29] Rabiner, L.R., Juang, B.H.: An Introduction to Hidden Markov Models, 13

[30] Klus, S., Nüske, F., Koltai, P., Wu, H., Kevrekidis, I., Schütte, C., Noé, F.: Data-Driven Model Reduction and Transfer Operator Approximation **28**(3), 985–1010. https://doi.org/10.1007/s00332-017-9437-7. Accessed 2019-10-10

[31] Wu, H., Noé, F.: Variational Approach for Learning Markov Processes from Time Series Data **30**(1), 23–66. https://doi.org/10.1007/s00332-019-09567-y. Accessed 2020-10-22

[32] Bowman, G.R., Pande, V.S., Noé, F. (eds.): An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation. Advances in Experimental Medicine and Biology, vol. 797. Springer Netherlands. https://doi.org/10.1007/978-94-007-7606-7. http://link.springer.com/10.1007/978-94-007-7606-7 Accessed 2019-11-26

[33] Kohlhoff, K.J., Robustelli, P., Cavalli, A., Salvatella, X., Vendruscolo, M.: Fast and Accurate Predictions of Protein NMR Chemical Shifts from Interatomic Distances **131**(39), 13894–13895. https://doi.org/10.1021/ja903772t. Accessed 2019-10-10

[34] Suárez, E., Wiewiora, R.P., Wehmeyer, C., Noé, F., Chodera, J.D., Zuckerman, D.M.: What Markov State Models Can and Cannot Do: Correlation versus Path-Based Observables in Protein-Folding Models **17**(5), 3119–3133. https://doi.org/10.1021/acs.jctc.0c01154. Accessed 2021-09-27

[35] Robustelli, P., Ibanez-de-Opakua, A., Campbell-Bezat, C., Giordanetto, F., Becker, S., Zweckstetter, M., Pan, A.C., Shaw, D.E.: Molecular basis of small-molecule binding to α-synuclein, 2021–0122426549. https://doi.org/10.1101/2021.01.22.426549. Accessed 2021-09-27

[36] Heller, G.T., Bonomi, M., Vendruscolo, M.: Structural Ensemble Modulation upon Small-Molecule Binding to Disordered Proteins **430**(16), 2288–2292. https://doi.org/10.1016/j.jmb.2018.03.015. Accessed 2021-09-29

[37] Lee, H., Mok, K.H., Muhandiram, R., Park, K.-H., Suk, J.-E., Kim, D.-H., Chang, J., Sung, Y.C., Choi, K.Y., Han, K.-H.: Local Structural Elements in the Mostly Unstructured Transcriptional Activation Domain of Human p53 * **275**(38), 29426–29432 10884388. https://doi.org/10.1074/jbc.M003107200. Accessed 2021-09-27

[38] Robustelli, P., Piana, S., Shaw, D.E.: Mechanism of Coupled Folding-upon-Binding of an Intrinsically Disordered Protein **142**(25), 11092–11101. https://doi.org/10.1021/jacs.0c03217. Accessed 2021-09-09

[39] Bonomi, M., Heller, G.T., Camilloni, C., Vendruscolo, M.: Principles of protein structural ensemble determination **42**, 106–116. https://doi.org/10.1016/j.sbi.2016.12.004. Accessed 2019-10-10

[40] Rangan, R., Bonomi, M., Heller, G.T., Cesari, A., Bussi, G., Vendruscolo, M.: Determination of Structural Ensembles of Proteins: Restraining vs Reweighting **14**(12), 6632–6641. https://doi.org/10.1021/acs.jctc.8b00738. Accessed 2019-10-10

[41] Mardt, A., Noé, F.: Progress in Deep Markov State Modeling: Coarse Graining and Experimental Data Restraints. http://arxiv.org/abs/2108.01927 Accessed 2021-09-21

[42] Laio, A., Parrinello, M.: Escaping free-energy minima **99**(20), 12562–12566. https://doi.org/10.1073/pnas.202427399. Accessed 2019-10-10

[43] Vernon, R.M., Chong, P.A., Tsang, B., Kim, T.H., Bah, A., Farber, P., Lin, H., Forman-Kay, J.D.: Pi-Pi contacts are an overlooked protein feature relevant to phase separation **7**, 31486. https://doi.org/10.7554/eLife.31486. Accessed 2021-09-27

[44] Paton, R.S., Goodman, J.M.: Hydrogen Bonding and π-Stacking: How Reliable are Force Fields? a Critical Evaluation of Force Field Descriptions of Nonbonded Interactions **49**(4), 944–955. https://doi.org/10.1021/ci900009f. Accessed 2021-09-27

[45] Baker, C.M.: Polarizable force fields for molecular dynamics simulations of biomolecules **5**(2), 241–254. https://doi.org/10.1002/wcms.1215. Accessed 2019-10-10

[46] Burton, A., Castaño, A., Bruno, M., Riley, S., Schumacher, J., Sultan, M.B., Tai, S.S., Judge, D.P., Patel, J.K., Kelly, J.W.: Drug Discovery and Development in Rare Diseases: Taking a Closer Look at the Tafamidis Story **15**, 1225–1243. https://doi.org/10.2147/DDDT.S289772. Accessed 2021-10-25

[47] Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess,

B., Lindahl, E.: GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers **1−2**, 19–25. https://doi.org/10.1016/j.softx.2015.06.001. Accessed 2019-10-10

[48] MacKerell, A.D., Bashford, D., Bellott, M., Dunbrack, R.L., Evanseck, J.D., Field, M.J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F.T.K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D.T., Prodhom, B., Reiher, W.E., Roux, B., Schlenkrich, M., Smith, J.C., Stote, R., Straub, J., Watanabe, M., Wiórkiewicz-Kuczera, J., Yin, D., Karplus, M.: All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins [†] **102**(18), 3586–3616. https://doi.org/10.1021/jp973084f. Accessed 2019-10-10

[49] Mayne, C.G., Saam, J., Schulten, K., Tajkhorshid, E., Gumbart, J.C.: Rapid parameterization of small molecules using the force field toolkit **34**(32), 2757–2770. https://doi.org/10.1002/jcc.23422. Accessed 2019-10-10

[50] Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., Scalmani, G., Barone, V., Petersson, G.A., Nakatsuji, H., Li, X., Caricato, M., Marenich, A.V., Bloino, J., Janesko, B.G., Gomperts, R., Mennucci, B., Hratchian, H.P., Ortiz, J.V., Izmaylov, A.F., Sonnenberg, J.L., Williams-Young, D., Ding, F., Lipparini, F., Egidi, F., Goings, J., Peng, B., Petrone, A., Henderson, T., Ranasinghe, D., Zakrzewski, V.G., Gao, J., Rega, N., Zheng, G., Liang, W., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Throssell, K., Montgomery, J.A. Jr., Peralta, J.E., Ogliaro, F., Bearpark, M.J., Heyd, J.J., Brothers, E.N., Kudin, K.N., Staroverov, V.N., Keith, T.A., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A.P., Burant, J.C., Iyengar, S.S., Tomasi, J., Cossi, M., Millam, J.M., Klene, M., Adamo, C., Cammi, R., Ochterski, J.W., Martin, R.L., Morokuma, K., Farkas, O., Foresman, J.B., Fox, D.J.: Gaussian 16 Revision A.03

[51] Bussi, G., Donadio, D., Parrinello, M.: Canonical sampling through velocity-rescaling **126**(1), 014101 arXiv:0803.4060. https://doi.org/10.1063/1.2408420. Accessed 2019-10-10

[52] Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A., Haak, J.R.: Molecular dynamics with coupling to an external bath **81**(8), 3684–3690. https://doi.org/10.1063/1.448118. Accessed 2019-10-10

[53] Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., Klein, M.L.: Comparison of simple potential functions for simulating liquid water **79**(2), 926–935. https://doi.org/10.1063/1.445869. Accessed 2019-10-10

[54] Hess, B.: P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation **4**(1), 116–122. https://doi.org/10.1021/ct700200b. Accessed 2019-10-10

[55] Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H., Pedersen, L.G.: A smooth particle mesh Ewald method **103**(19), 8577–8593. https://doi.org/10.1063/1.470117. Accessed 2019-10-10

[56] Zimmerman, M.I., Bowman, G.R.: FAST Conformational Searches by Balancing Exploration/Exploitation Trade-Offs **11**(12), 5747–5757. https://doi.org/10.1021/acs.jctc.5b00737. Accessed 2019-10-10

[57] Pérez-Hernández, G., Paul, F., Giorgino, T., De Fabritiis, G., Noé, F.: Identification of slow molecular order parameters for Markov model construction **139**(1), 015102. https://doi.org/10.1063/1.4811489. Accessed 2019-10-10

[58] Schwantes, C.R., Pande, V.S.: Improvements in Markov State Model Construction Reveal Many Non-Native Interactions in the Folding of NTL9 **9**(4), 2000–2009. https://doi.org/10.1021/ct300878a. Accessed 2019-10-10

[59] Arthur, D., Vassilvitskii, S.: K-means++: The Advantages of Careful Seeding. In: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms. SODA '07, pp. 1027–1035. Society for Industrial and Applied Mathematics. http://dl.acm.org/citation.cfm?id=1283383.1283494 Accessed 2019-12-03

[60] Klambauer, G., Unterthiner, T., Mayr, A., Hochreiter, S.: Self-Normalizing Neural Networks. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (eds.) Advances in Neural Information Processing Systems 30, pp. 971–980. Curran Associates, Inc. http://papers.nips.cc/paper/6698-self-normalizing-neural-networks.pdf Accessed 2020-10-22

[61] LeCun, Y.A., Bottou, L., Orr, G.B., Müller, K.-R.: Efficient BackProp. In: Montavon, G., Orr, G.B., Müller, K.-R. (eds.) Neural Networks: Tricks of the Trade: Second Edition. Lecture Notes in Computer Science, pp. 9–48. Springer. https://doi.org/10.1007/978-3-642-35289-8_3. https://doi.org/10.1007/978-3-642-35289-8_3 Accessed 2021-11-05

[62] Chollet, F.: Keras

[63] Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D.G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., Wicke, M., Yu, Y., Zheng, X.: TensorFlow: A System

for Large-Scale Machine Learning, pp. 265–283. https://www.usenix.org/conference/osdi16/technical-sessions/presentation/abadi Accessed 2019-11-10

[64] Kingma, D.P., Ba, J.: Adam: A Method for Stochastic Optimization. http://arxiv.org/abs/1412.6980 Accessed 2019-10-10

[65] PLUMED consortium: Promoting transparency and reproducibility in enhanced molecular simulations **16**(8), 670–673 31363226. https://doi.org/10.1038/s41592-019-0506-8

[66] Tribello, G.A., Bonomi, M., Branduardi, D., Camilloni, C., Bussi, G.: PLUMED 2: New feathers for an old bird **185**(2), 604–613 arXiv:1310.0980. https://doi.org/10.1016/j.cpc.2013.09.018. Accessed 2019-10-10

[67] Kabsch, W., Sander, C.: Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features **22**(12), 2577–2637. https://doi.org/10.1002/bip.360221211. Accessed 2019-10-10
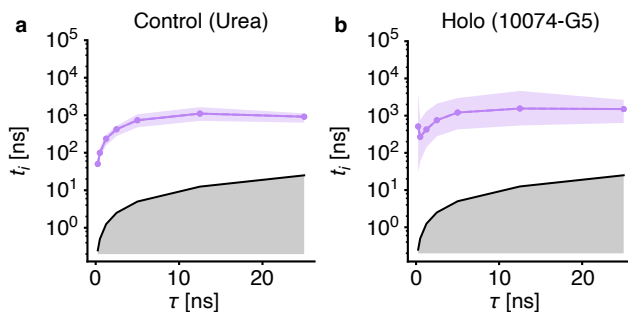
Appendix A   Extended Data

Fig. A1 Relaxation timescale as a function of model lag time for **a** control (urea) and **b** holo (10074-G5) ensembles. Gray shaded areas indicate timescales the Koopman model can no longer resolve. Coloured shaded areas indicate 95th percentiles of the sample mean over all 20 models.
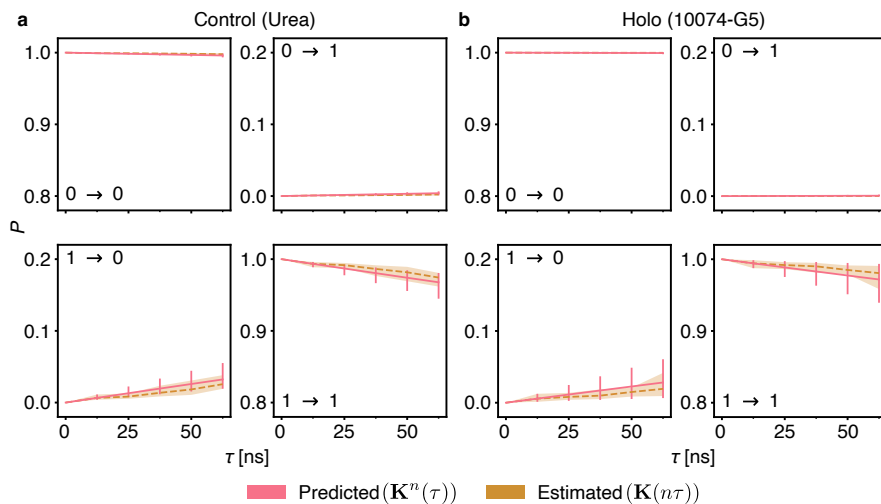

Fig. A2 Chapman-Kolmogorov test for **a** control (urea) and **b** holo (10074-G5) ensembles. Each panel indicates the transition probability for one matrix entry for successive applications (predicted, red) and estimations (estimated, orange) of the Koopman matrix. Shaded areas and error bars indicate 95th percentiles of the mean over all 20 models.
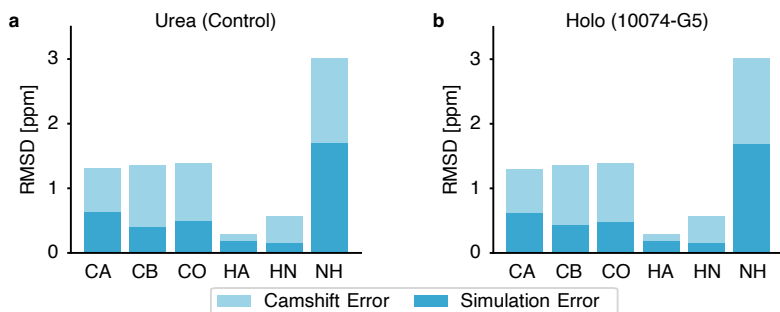
Fig. A3 Root-mean-square deviations between experimentally determined NMR chemical shifts[5] and those back-calculated using CamShift[33] for **a** control (urea) and **b** holo (10074-G5) ensembles. Light shaded areas indicate the intrinsic error of the CamShift predictor.
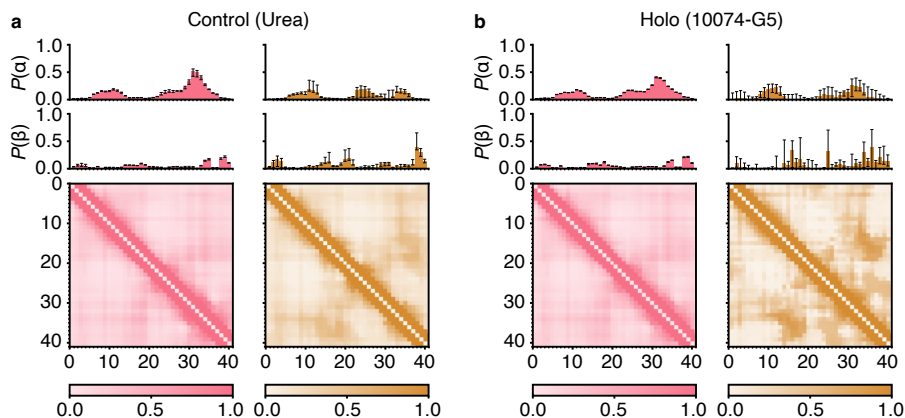

Fig. A4 Structural properties of the **a** control (urea) and **b** holo (10074-G5) ensembles. Top panels indicate α-helical and β-sheet contents over all residues as calculated using DSSP[67]. Bottom panels show heavy-atom contact probability maps with a cut-off of 0.8 nm. Error bars indicate 95th percentiles of the bootstrap sample of the mean over all 20 models.
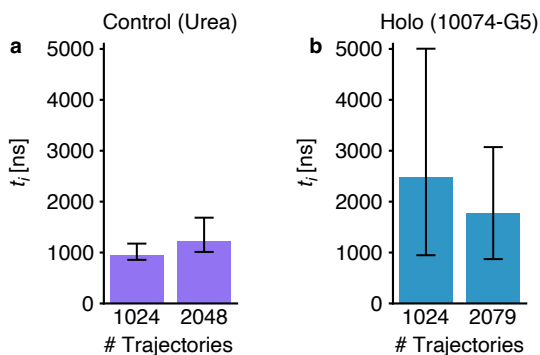
Fig. A5 Dependence of the relaxation timescales on the number of trajectories used to build the model. Errorbars indicate 95th percentiles of the bootstrap sample of the mean over the first 5 models.
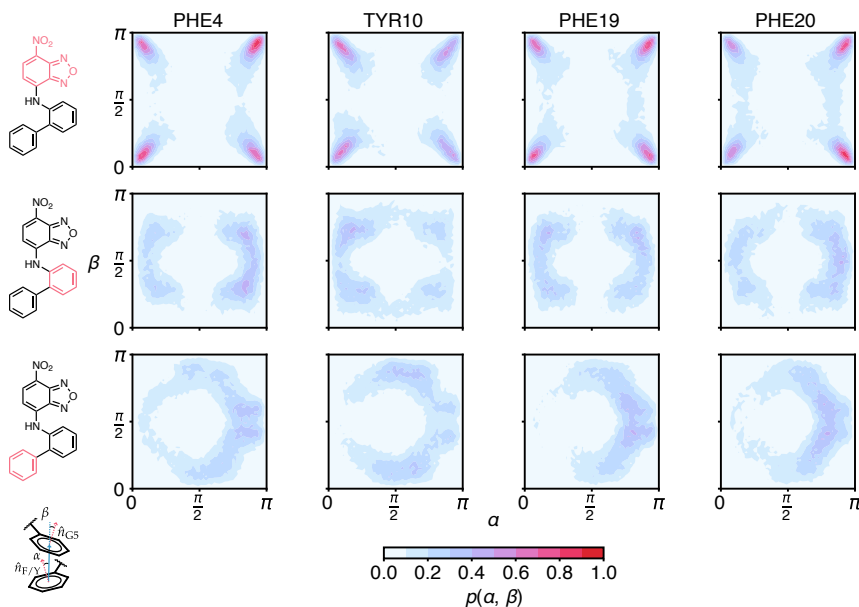

Fig. A6 Anisotropy of π-π molecule — aromatic side chain stacking interactions for the holo (10074-G5) ensemble. α is the stacking angle between the inter-aromatic distance vector and the aromatic side chain normal vector, while β is the angle inter-aromatic distance vector and the normal vector of the small molecule aromatic system[35]. Distributions show the density under the condition that the distance between both groups is below 0.6 nm.