

An integrated machine learning approach delineates an entropic expansion mechanism for the binding of a small molecule to α -synuclein

Sneha Menon[†], Subinoy Adhikari[†], Jagannath Mondal^{*}

Tata Institute of Fundamental Research, Hyderabad, India

eLife Assessment

This study describes the application of machine learning and Markov state models to characterize the binding mechanism of alpha-Synuclein to the small molecule Fasudil. The results suggest that entropic expansion can explain such binding. However, the simulations and analyses in their present form are **inadequate**.

*For correspondence:

jmondal@tifrh.res.in

[†]These authors contributed equally to this work

Competing interest: The authors declare that no competing interests exist.

Funding: See page 25

Preprint posted
27 March 2024

Sent for Review
27 March 2024

Reviewed preprint posted
03 June 2024

Reviewed preprint revised
08 November 2024

Version of Record published
18 December 2024

Reviewing Editor: Rosana Collepardo, University of Cambridge, United Kingdom

© Copyright Menon, Adhikari et al. This article is distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use and redistribution provided that the original author and source are credited.

Abstract The mis-folding and aggregation of intrinsically disordered proteins (IDPs) such as α -synuclein (α S) underlie the pathogenesis of various neurodegenerative disorders. However, targeting α S with small molecules faces challenges due to the lack of defined ligand-binding pockets in its disordered structure. Here, we implement a deep artificial neural network-based machine learning approach, which is able to statistically distinguish the fuzzy ensemble of conformational substates of α S in neat water from those in aqueous fasudil (small molecule of interest) solution. In particular, the presence of fasudil in the solvent either modulates pre-existing states of α S or gives rise to new conformational states of α S, akin to an ensemble-expansion mechanism. The ensembles display strong conformation-dependence in residue-wise interaction with the small molecule. A thermodynamic analysis indicates that small-molecule modulates the structural repertoire of α S by tuning protein backbone entropy, however entropy of the water remains unperturbed. Together, this study sheds light on the intricate interplay between small molecules and IDPs, offering insights into entropic modulation and ensemble expansion as key biophysical mechanisms driving potential therapeutics.

Introduction

Intrinsically disordered proteins (IDPs) comprise a class of proteins that lack a unique, well-defined three-dimensional structure under physiological conditions. Defying the classical structure-function paradigm (*Wright and Dyson, 2015; Babu et al., 2011*), they are involved in crucial cellular processes such as regulation of cell cycle signalling and act as hubs in interaction networks mainly due to their promiscuous binding nature (*Csizmok et al., 2016; Wang et al., 2011; Dyson, 2016*). Emergent studies present disordered proteins as an important component of biomolecular condensates that play regulatory roles in several key cellular functions (*Alberti et al., 2019; Chong and Forman-Kay, 2016; Shin and Brangwynne, 2017; O'Flynn and Mittag, 2021*).

Considering the broad range of functionality of IDPs, their dysregulation and altered abundance can lead to a number of diseases including cancer, diabetes, cardiovascular and neurodegenerative disorders (*Wright and Dyson, 2015; Babu et al., 2011; Uversky et al., 2008*). In rational structure-based

drug design, exploration of the target proteins is a crucial step based on which the interactions of small molecules are optimized to have stabilising interactions with a structured ligand-binding pocket (Bronowska, 2011; Klebe, 2015) and is typically applied to folded proteins. IDPs, however, are considered *undruggable* since they exist as an ensemble of conformations that precludes the usage of these traditional drug-design strategies (Metallo, 2010). Nevertheless, a few monomeric disordered proteins such as the oncogenic transcription factor c-Myc (Follis et al., 2008), Abeta42 (Ono et al., 2012; Zhu et al., 2013; Attanasio et al., 2013; Ehrnhoefer et al., 2008; Heller et al., 2020), Tau (Akoury et al., 2013), PTP1B (Krishnan et al., 2014), and osteopontin Kurzbach et al., 2014 have been successfully targeted by small molecules (Biesaga et al., 2021). The IDP of this article's interest, alpha-synuclein (α S) has remained an important drug-target (Fields et al., 2019; Tóth et al., 2014; Tatenhorst et al., 2016; Cao et al., 2022) owing to its long-rooted association with neurodegenerative diseases. Often, the early aggregation stages of α S into small oligomers is implicated in the causation of Parkinson's disease (Winner et al., 2011; Emin et al., 2022; Stephens et al., 2023; Ubbiali et al., 2022) and a potential therapeutic strategy has been the stabilization of α S in soluble monomeric form. In particular, the small molecule fasudil (focus of the present investigation) has been shown to interact with α S and retard its aggregation (Tatenhorst et al., 2016). The anti-aggregative effect of fasudil on α S assembly was tested both in vivo and in vitro and it was found to significantly reduce α S aggregation in both scenarios. Characterisation of the effect of small molecule binding on the ensemble properties of IDPs and the molecular details of the stabilizing interactions can provide the basis for selectivity and can be effectively calibrated to induce the desired effect to alleviate disease conditions. In particular, latest effort by Robustelli and coworkers (Robustelli et al., 2022) had produced an unprecedentedly long-time scale (1500 μ s) all-atom molecular dynamics (MD) simulation trajectory of a small molecule drug, fasudil with monomeric α S, in which the predicted protein-fasudil interactions were found to be in good agreement with previously reported NMR chemical shift data. However, comparison of the global and local structural features of the α S ensemble in neat water and that in the presence of fasudil (Robustelli et al., 2022; see Figure 1—figure supplements 1–6) did not indicate a significant difference that is a customary signature of the dynamic IDP ensemble. Characterization of the peculiar nature of IDPs such as α S using the conventional experimental and computational approaches have proven to be challenging. IDPs can be best described by an ensemble of conformations potentially sampled by the protein along with their associated statistical/thermodynamic weights. While biophysical methods can provide an ensemble averaged conformation of an IDP, computer simulations can extensively sample the conformational ensembles of IDPs at atomic resolutions. However, biophysical methods lack the potential to provide an ensemble view and there are inherent limitations of force fields used in computer simulations to model IDPs in agreement with experiments. Integrative approaches have been exploited in studying IDPs as well as small-molecule binding to IDPs (Bonomi et al., 2016; Bottaro and Lindorff-Larsen, 2018; Robustelli et al., 2018; Gomes et al., 2020; Stelzl et al., 2022). In recent years, Markov State Models (MSMs) have been exploited to elucidate a states-and-rates view of biomolecular dynamics i.e. finding the kinetically relevant states and the transition rates between them (Prinz et al., 2011; Bowman et al., 2009; Husic and Pande, 2018). The first step in the process of building a kinetic model / MSM of any complex biomolecular ensemble is to identify a suitable metric or collective variable (CV) that efficiently captures its conformational space.

CVs such as torsion angles, distances, chemical shifts and so forth, can characterize the complex molecular motion in a high number of dimensions. However, this presents a challenge to effectively analyze trajectories generated from MD simulations. In order to address this, dimension reduction is performed to obtain a set of slow variables that can be further subjected to clustering into kinetically discrete metastable states. Since the high-dimensional input features describing the heterogeneous conformational landscape of IDPs can have non-linear relationships, non-linear dimension reduction techniques are most suitable. Recent advances in the field of machine learning have shown better results in capturing such non-linear relationships (Adhikari and Mondal, 2023). In this study, we employed a deep neural network, specifically β -variational autoencoder (Kingma and Welling, 2013; Rezende et al., 2014) to perform dimension reduction and further utilized this machine-learned latent dimension for building MSM in order to better understand the conformational landscape of IDPs in presence and in absence of a small molecule. In the current work, to elucidate the small-molecule mediated modulatory effect on α S ensemble, we harnessed deep-learning based β -Variational AutoEncoder (β -VAE)

and Markov State Models (MSM) to dissect α S ensemble in water and in aqueous fasudil solution. A projection of the atomistically simulated conformation ensemble of α S in its apo state and in presence of Fasudil, along their respective latent space describes the conformational landscape, which is further refined by developing individual MSMs. Our result reveals that, compared to the macrostates identified in water, the presence of fasudil led to an increase in the number of metastable states. Characterisation of small-molecule interaction with the protein indicates that fasudil has differential interactions with each of these macrostates. These macrostates exhibit clear distinctions as evident from their C_{α} contact maps which was further analysed using a convolutional VAE. An entropic perspective at the global (whole protein) and local (residue) level reveals that fasudil binding can potentially trap the protein and the resultant states may be disordered or ordered in nature. The effect is mainly manifested at the backbone level with a decrease in entropy at the fasudil binding hotspots in some states. The observations reported in this study demonstrate how the binding of a small molecule, fasudil, modulates the conformational properties of the disordered protein, α S, manifested as a change in the backbone entropy of the fasudil binding hotspots.

Results and discussion

Modulation of free energy landscape of α S in presence of fasudil

Simulations of α S monomer in the presence of the small molecule fasudil and 50 mM NaCl have been reported in a previous study by *Robustelli et al., 2022*. This study showed no large-scale differences between the bound and unbound states of α S. In our study, in order to compare the α S-fasudil ensemble with an apo ensemble at same salt concentration, we spawned multiple all-atom MD simulations of α S in neat water and 50 mM NaCl from multiple starting conformations to generate a cumulative ensemble of $\sim 62 \mu$ s.

In a clear departure from the classical view of ligand binding to a folded globular protein, the visual change in α S ensemble due to the presence of small molecule is not so strikingly apparent. In order to understand the underlying ensemble modulatory effect of the small-molecule binding events, the complex and fuzzy conformational ensemble of α S (see *Figure 1a*) needs to be delineated using a suitable spatial and temporal decomposition into its key sub-ensembles or metastable states. This prompted us to analyse the two ensembles using the framework of MSM. As a suitable input feature to build the MSM, we estimated the set of inter-residue C_{α} pairwise distances, as this feature largely incorporates the conformational space of α S. In a standard MSM analysis, this high-dimensional input feature requires further processing such as dimensionality reduction, before discretisation using clustering for the construction of an MSM.

We started our investigation by searching for an optimized representation of high-dimensional and heterogeneous ensemble of monomeric α S conformation in presence of fasudil. As the datasets, particularly in the context of IDPs, are more complex and large in size, the need for effective dimensionality reduction techniques becomes more conspicuous. Dimension reduction using principal component analysis, independent component analysis (*Hyvärinen, 2013*), singular value decomposition (*Klema and Laub, 1980*) and linear discriminant analysis (*Fisher, 1936*) linearly transform the high dimensional data into a lower dimensional manifold. However, non-linear methods such as kernel PCA (*Schölkopf et al., 2005*), multidimensional scaling (*Kruskal and Wish, 1978*), isomap (*Tenenbaum et al., 2000*), fastmap (*Faloutsos and Lin, 1995*), locally linear embedding (*Roweis and Saul, 2000*) etc *Sumithra and Surendran, 2015* have outperformed such linear transformation methods and have been used for free energy calculations (*Das et al., 2006*) and determining timescales to capture slow conformational changes (*Schwantes and Pande, 2015; Wehmeyer and Noé, 2018*). In the present work, we opted to draw inspiration from artificial deep neural network based framework to employ model-agnostic and mostly unsupervised non-linear approach to derive optimized non-linear latent feature space that can be employed for statistical state-space decomposition of IDP such as α S.

Over the recent years, unsupervised machine learning algorithms such as autoencoders have been effective as a dimension reduction tool, owing to non-linear activations of the neurons that allows them to capture complex relationships and non-linear patterns in the high dimensional data (*Hinton and Salakhutdinov, 2006; Adhikari and Mondal, 2023*). This high dimensional input is represented into a lower dimensional latent space, which is then used to reconstruct the input. However, the latent space is 'deterministic', that is it provides only a fixed mapping of the input to latent space, which might

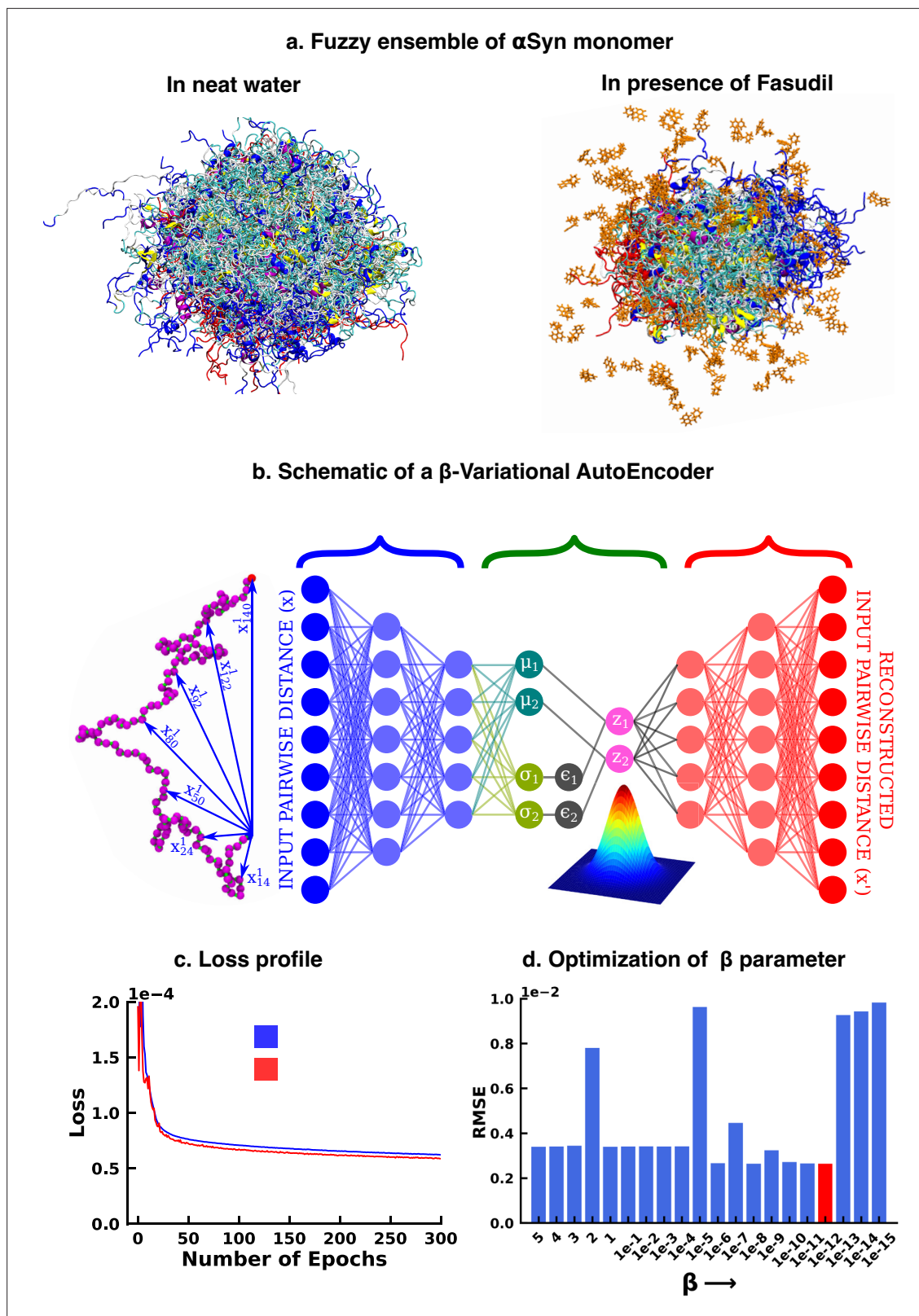


Figure 1. Deep neural network illustrating dimension reduction through β -Variational AutoEncoder. (a) Conformational ensemble view of α S and α S-Fasudil ensemble, (b) A schematic of β -Variational Autoencoder (β -VAE), (c) Training and validation loss for β -VAE and (d) RMSE as a function of the β parameter. The red-annotated β parameter was used for the investigation.

Figure 1 continued on next page

Figure 1 continued

The online version of this article includes the following figure supplement(s) for figure 1:

Figure supplement 1. Distribution of the radius of gyration (Rg) of the α S and α S-Fasudil ensembles.

Figure supplement 2. Distribution of end-to-end distance for the α S and α S-Fasudil ensembles.

Figure supplement 3. Distribution of SASA for α S and α S-Fasudil ensembles.

Figure supplement 4. Residue wise percentage secondary structure of helical nature in the α S and α S-Fasudil ensembles.

Figure supplement 5. Residue wise percentage secondary structure of sheet nature in the α S and α S-Fasudil ensembles.

Figure supplement 6. Residue wise percentage secondary structure of coil nature in the α S and α S-Fasudil ensembles.

Figure supplement 7. Each blue diamond label represents the RMSD (nm) w.r.t the starting structure of aSyn-fasudil simulation.

Figure supplement 8. Fraction of helix content in the starting structures of the α S-Fasudil and α S simulations.

Figure supplement 9. Fraction of β -sheet content in the starting structures of the α S-Fasudil and α S simulations.

Figure supplement 10. Fraction of coil content in the starting structures of the α S-Fasudil and α S simulations.

Figure supplement 11. Total SASA of the starting structures of the α S-Fasudil and α S simulations.

Figure supplement 12. Residue wise SASA of the starting structure of the α S-Fasudil (red square) and α S (blue squares) simulations.

not account for variations or uncertainty present in the data. Consequently, this could generate an inaccurate result for a point that does not exist in the latent space. Thus, when projecting new, similar data onto the latent space, a deterministic autoencoder will map similar inputs to nearly identical points in the latent space, potentially lacking the diversity and capturing only a single mode of the data distribution. In contrast, the latent space of a Variational Autoencoder (VAE) is 'probabilistic' in nature, thereby avoiding this limitation of autoencoders. The technical details have been documented in the Method section. However, here we provide a brief rationale for the choice of the β -VAE model.

Variational autoencoder (VAE) (Kingma and Welling, 2013; Rezende et al., 2014; Figure 1b) is also an unsupervised machine learning algorithm, which although is closely related to autoencoder, it is inextricably linked with variational bayesian methods. A VAE consists of an encoder network and a decoder network, much like a traditional autoencoder. The encoder takes input data and maps it to a probabilistic distribution in the latent space, while the decoder reconstructs the input data from samples drawn from this distribution. The probabilistic nature of the latent space is a result of applying variational inference, which aims to estimate the actual posterior using an approximate distribution, parameterized by a mean vector and a variance vector. This estimation is achieved by minimizing the Kullback-Leibler (KL) divergence (also known as relative entropy or information gain) between the two distributions. As a result, instead of producing a single point in the latent space for each input, the encoder outputs a distribution. The latent variable is then sampled from this distribution, which is then used by the decoder to reconstruct the input. This makes the latent space continuous, thereby improving its ability to interpolate novel, unseen data. The VAE model can be further improved by enhancing the probability of generating an actual data while keeping the distance between the true and the posterior distribution under a small threshold. This results in the weighting of the KL divergence term in the loss function. The factor determining this weight is represented by β and hence the name β -VAE, which is used in this study. The implementation of the β -VAE is made publicly available in our group's [Github page](#) (copy archived at [JMLab-tifrh, 2024](#)). The loss function of β -VAE is given as,

$$\text{Loss}_{VAE} = \text{Reconstruction Loss} + \beta \times \text{KL Divergence} \quad (1)$$

The β -VAE model was trained over a number of β values ranging between 5 and 1×10^{-15} . A β value of 1×10^{-12} was chosen as this value gave us the minimum RMSE between the original and the reconstructed data (see Figure 1d). This model was further used to project the α S data in its apo state for further analysis.

We have also computed the VAMP2 score after dimension reduction of C α pairwise distance using a linear dimension reduction method namely time-lagged independent component analysis (tICA) and compared it with the score obtained from the latent dimension of β -VAE. We found that the VAMP2 score obtained from β -VAE was consistently higher than tICA derived dimensions for the α S-Fasudil ensemble, thereby indicating that β -VAE would help identifying relatively slower CVs (see

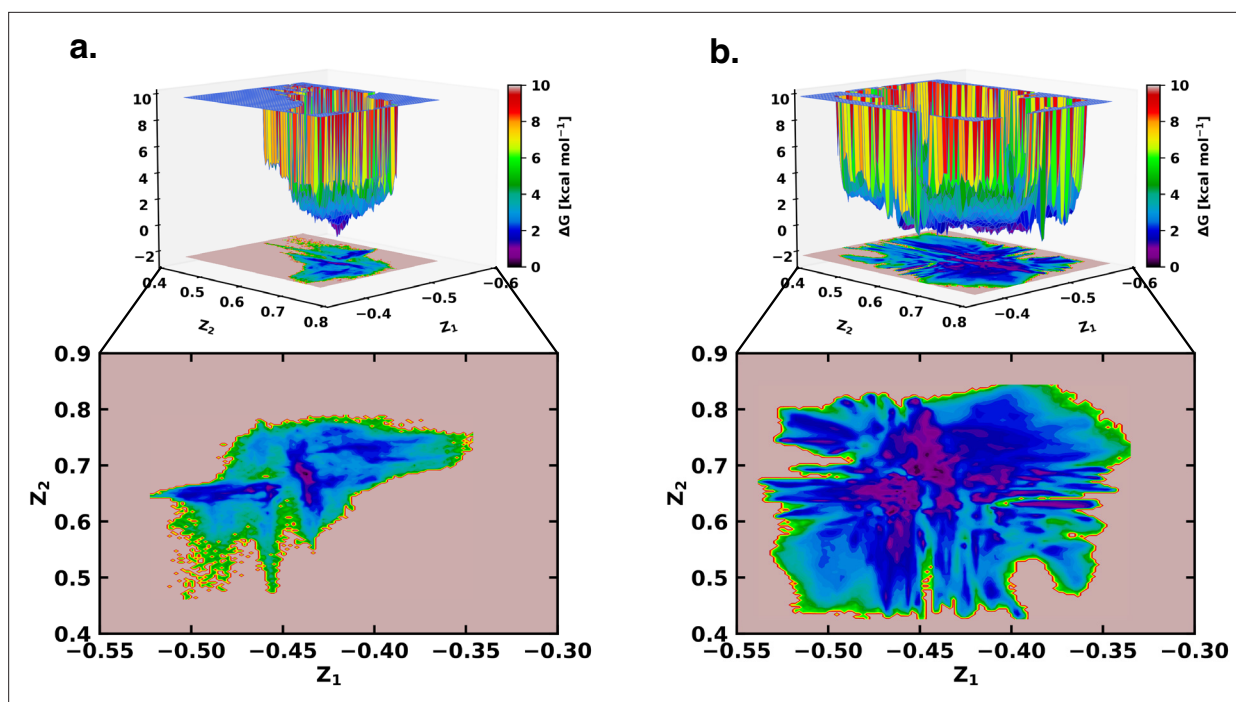


Figure 2. Free energy landscape using the latent space of β -VAE indicates greater conformational space sampled by α S in presence of fasudil. The free energy landscape of (a) α S and (b) α S-Fasudil system as determined from the latent dimensions of β -VAE.

Figure 3—figure supplements 1–2) as compared to tICA and hence better at representing slower dynamical processes in the α S and α S-Fasudil simulations.

In particular, as described in detail in the Methods section, we trained a VAE model using a large array of inter-residue pairwise distances using the β -VAE architecture. Our attempt to reconstruct the large-dimensional input feature via a β -VAE resulted in compression of the 1.5 ms MD simulation data into a two-dimensional latent space. The loss function steadily decreased and eventually plateaued providing a model optimized for this system (see **Figure 1c**). The rugged nature of the free energy landscape of α S, is evident from its projection along the latent space dimension of the simulated trajectory either in aqueous media (see **Figure 2a**) or in presence of fasudil (see **Figure 2b**). The free energy landscape represents a large number of spatially close local minima representative of energetically competitive conformations inherent in α S, which is also the source of the conformational heterogeneity and a hallmark of an IDP (see **Figure 2a–b**). Next we wanted to compare the changes in the conformational landscape of the apo α S conformations. Subsequent projection of the apo state on the low-dimensional subspace also brought out a spatially heterogeneous landscape.

However, a close comparison of free energy landscapes indicates that a set of local minima representing conformations of α S are less clustered and appear in small patches in the apo state. Moreover, a relative comparison of the two free energy landscapes suggests that α S spans a significantly larger space in presence of fasudil even in this low dimensional subspace, hinting at an expansion of conformational repertoire of this IDP in presence of the small molecule (see **Figure 2a–b**). Individual, one-dimensional projection of the conformational landscape along each of the two latent dimensions reveals that the presence of the small molecule would shift part of the conformational landscape to a distinct location, suggesting that fasudil would modulate the existing conformational ensemble as well as would create new conformational space. For a more discrete and clearer state-space decomposition of the conformational ensemble of α S (in apo or in ligand-bound form), we attempt to enumerate the optimum number of distinct metastable macrostates with non-negligible equilibrium populations by constructing MSM as explained in the following section.

MSM elucidates distinct binding competing states of α S in presence of the small-molecule drug

The two-dimensional feature obtained from β -VAE was used as the input feature for the MSM estimator provided by PyEMMA (Scherer et al., 2015). A flowchart of the implemented protocol is drawn in **Figure 3a**. A geometric clustering approach, k -means, was used to discretise the ensemble. Analysis of the implied time scales as a function of lag-time predicted a Markovian model of α S in water with three kinetically separated states (see **Figure 3d** for relative population). On the other hand, a similar analysis of α S in presence of fasudil (**Figure 3c**) indicated a clear increase in the number of spatially and temporally resolved conformational metastable states. Thus, a three-state MSM was built for the α S ensemble in water whereas a six-state model was built for the α S-fasudil ensemble. Hereafter, we will refer to the macrostates created in presence of fasudil prefixed with FS and those generated in water prefixed with MS. A bootstrapping analysis was performed to estimate the mean and standard deviations of the equilibrium populations of these metastable states (**Figure 3d**).

The MSM was also further validated for Markovianity i.e. ability to predict estimates at longer time scales using the Chapman-Kolmogorov (CK) test, as depicted in **Figure 3—figure supplement 3** for α S-Fasudil simulation. This test further validates the quality of the model that it estimates long-time scale behavior with reasonable accuracy. The MSM analyses essentially indicate that the small-molecule binding interactions with α S led to sampling or generation of additional conformational states than the states populated in the apo α S ensemble.

A pertinent question arises: Are the appearances of newer states in α S in presence of fasudil a result of the presence of longer trajectory in fasudil solution (1500 μ s) than that of the ensemble populated in neat water (62 μ s)? We verified this by building an MSM for 60 μ s of data of fasudil system using 10 μ s to 70 μ s segment of the trajectory (see **Figure 3—figure supplement 4**). We find that the MSM built using lesser data (and same amount of data in the α S simulation) also indicated the presence of six states of α S in presence of fasudil, as was observed in the MSM of the full trajectory. We compared the R_g distribution of this small chunk with the α S ensemble and found to be in good agreement (see **Figure 3—figure supplement 5**). Additionally, we selected a chunk spanning from 966 μ s to 1026 μ s and observed a similar distribution of R_g (see **Figure 3—figure supplement 6**). Moreover, we built a MSM on this new chunk and identified the presence of six states (see **Figure 3—figure supplement 7**). Together, this exercise invalidates the sampling argument and suggests that the increase in the number of metastable macrostates of α S in fasudil solution relative to that in water is a direct outcome of the interaction of α S with the small molecule.

Another aspect that we evaluated was in the choice of the number of latent dimensions of β -VAE. Even though an increase in the number of latent dimensions may make the model more accurate, this can also result in overfitting. The model can simply memorize the pattern in the data instead of generalizing them. A higher dimensional latent space is also more difficult to interpret; therefore, we chose two dimensions. Nevertheless, we trained another β -VAE with four neurons on the latent space and built a MSM by choosing the appropriate number of microstates (see **Figure 3—figure supplement 8**). The implied timescales (see **Figure 3—figure supplement 9**) indicate the presence of six states which is consistent with the model with two latent dimensions.

Structural characterisation of metastable states of α S monomer in presence of fasudil

We characterised the residue-wise intramonomer contact maps of the metastable states to identify the differences in the interaction patterns in these states that can be attributed to interactions of fasudil with the monomer. The average inter-residue contact probability maps for each of the metastable states populated in the presence of fasudil and in water are depicted in **Figures 4 and 5**, respectively. The residue-wise percentage secondary structure in each of the macrostates are presented in **Figure 4—figure supplement 1**, **Figure 5—figure supplement 1**, respectively. In state FS1, antiparallel β -sheet interactions are formed within the N-terminal region. This β -sheet network also includes long-range interactions of the H2 region of the N-terminus with the negatively charged C-terminal region. Residues 70–80 in the hydrophobic NAC region are also involved in antiparallel β -sheet interactions with residues 120–130 in the C-terminal region. Short parallel β -sheet interactions also exist between the NAC and C-terminal residues with the N-terminal residues. The residues in the C-terminus exhibit 40–90% of β -sheet propensity (see **Figure 4—figure supplement 1**). Long

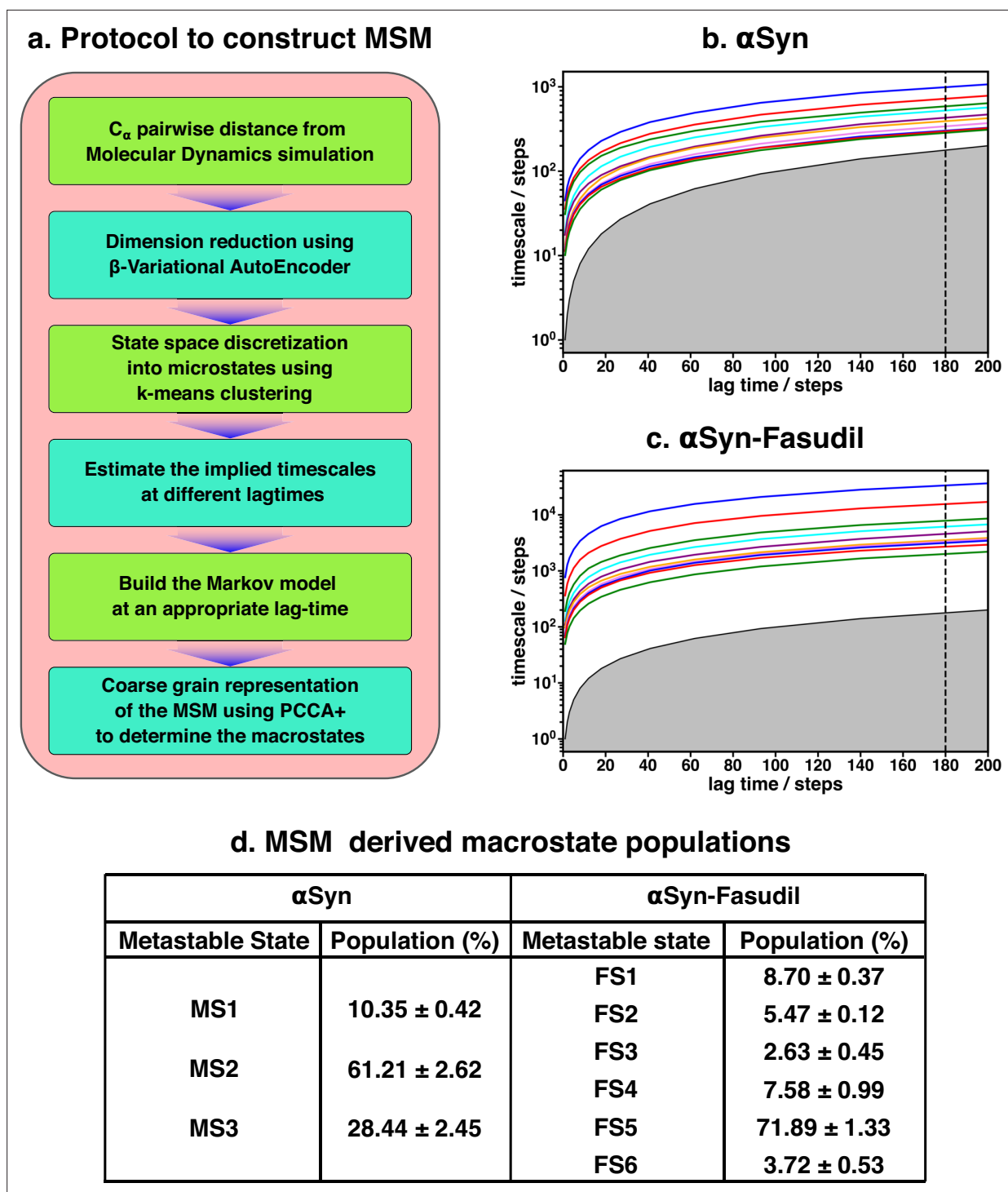


Figure 3. Markov state model captures three and six states of α S and α S-Fasudil system, respectively. (a) A flowchart of the process of building a Markov State Model (MSM). (b, c) Implied timescales plot of the α S and α S-Fasudil systems. (c) Macrostate populations of the 3-state and 6-state MSM of the α S and α S-Fasudil systems. Bootstrapping was used to estimate the mean and standard deviations of the macrostate populations.

The online version of this article includes the following figure supplement(s) for figure 3:

Figure supplement 1. Comparison of VAMP-2 score from VAE and tICA of pairwise distances, reduced to two dimensions for the α S-Fasudil simulation.

Figure supplement 2. Comparison of VAMP-2 score from VAE and tICA of pairwise distances, reduced to two dimensions for the α S simulation.

Figure supplement 3. The Chapman-Kolmogorov test performed for the six state Markov State Model of the α S-Fasudil ensemble.

Figure supplement 4. The implied timescale (ITS) plot α S-Fasudil simulation for the block of 10 μ s to 70 μ s.

Figure 3 continued on next page

Figure 3 continued

Figure supplement 5. Distribution of R_g of the αS ensemble and the 60 μs slice (10 μs -70 μs) of the αS -Fasudil simulation trajectory.

Figure supplement 6. Distribution of R_g of the αS ensemble and the 60 μs slice (966 μs to 1026 μs) of the αS -Fasudil simulation trajectory.

Figure supplement 7. The implied timescale (ITS) plot αS -Fasudil simulation for the block of 966 μs to 1026 μs .

Figure supplement 8. VAMP-2 score as a function of the number of microstates in for the αS -Fasudil ensembles for a latent space of four dimensions.

Figure supplement 9. ITS plot of the αS -Fasudil simulation using the four dimension of the VAE.

Figure supplement 10. VAMP-2 score as a function of the number of microstates in for the αS -Fasudil ensemble.

Figure supplement 11. VAMP-2 score as a function of the number of microstates in for the αS ensemble.

range interactions between N-terminus and C-terminus are relatively diminished in state FS2. This state has multiple parallel β -sheet interactions between short stretches of residues in the NAC region and C-terminal region with the N-terminal region. In addition, antiparallel β -sheet interactions are also present in NAC:N-terminus, NAC:C-terminus and within the N-terminal region. Residues 8–13 in the N-terminus and 86–92 in the NAC region exhibit significant helicity. State FS3 is characterised by ~20% of β -sheet interactions across the entire protein. Hydrophobic and polar, uncharged residues in the range 61–80 in the NAC region are involved in β -sheet interactions with the N-terminus(residues 1–10) and C-terminus(100–120). Moreover, β -sheet interaction networks are also present within the NAC and N-terminal regions. The C-terminal region 100–120 forms long-range interactions with residues 1–10 in the N-terminal region. Extensive short and long-range β -sheet networks are prevalent in state FS4. The entire NAC region forms antiparallel β -sheet network with the N-terminus. The C-terminus forms parallel β -sheets with the NAC and N-terminus. In the most populated state FS5, the key interactions are β -sheets formed between the oppositely charged termini. The NAC region also exhibits α -helical propensity. Finally, in state FS6, the NAC region (residues 60–80) forms antiparallel β -sheet network with the residues 30–60 of the N-terminal segment as well as the C-terminal region 100–120. These residues have ~40% β -sheet propensity.

On the other hand, the metastable states that are populated in neat water are distinct from those formed in the presence of fasudil. In state MS1, the NAC region residues 61–80 interacts with the N-terminal region by both parallel and antiparallel β -sheet interactions. The β -sheet network of NAC also incorporates interactions with residues 100–130 in the C-terminal region. These residues have ~60% β -sheet propensity (see **Figure 5—figure supplement 1**). State MS2, the most populated macrostate in neat water, is characterised by β -sheet interactions of the NAC region with H2 region (residues 30–60) of the N-terminus. Furthermore, the N and C-termini also form β -sheet interactions. State MS3 mainly has long-range interactions as β -sheets between the N and C-terminal regions. Moreover, a β -sheet network also exists within the NAC region and NAC:N-terminal (H2) interface. The residues in the NAC region have ~50% β -sheet propensity in this macrostate.

We note that the differences in the structural features that is intra-protein contacts of the macrostates arise from the interactions of fasudil with the protein residues. For instance, in state FS2 in which fasudil interacts with multiple sites across the entire length of the protein, the secondary structure properties of this state are dictated by β -sheets formed by short stretches of residues rather than an extensive network. Furthermore, owing to the relatively stronger small-molecule binding interactions with hydrophobic residues in the N-terminus, the long-range interactions of the oppositely charged terminal regions are suppressed. Similarly, in state FS4, since fasudil primarily interacts with residues 120–140 of the C-terminus and other interactions are transient, the conformational properties of this metastable state is marked by a network of β -sheets across the rest of the protein. Essentially, small-molecule binding interactions via stabilisation of various intra-protein interactions modulate the conformational properties of αS .

Mapping the macrostates in αS and αS -Fasudil ensembles using denoising convolutional variational autoencoder

To ascertain the distinctions among these contact maps for the αS and αS -Fasudil system, we chose to filter the essential features and patterns by representing it in a lower dimensional space. Initially, this was attempted using a Convolutional Variational AutoEncoder (CVAE) with the contact maps of αS and αS -Fasudil macrostates as inputs. The model was initially trained using the six contact maps of αS -Fasudil system as a test case. However, the model displayed a tendency to overfit by learning an

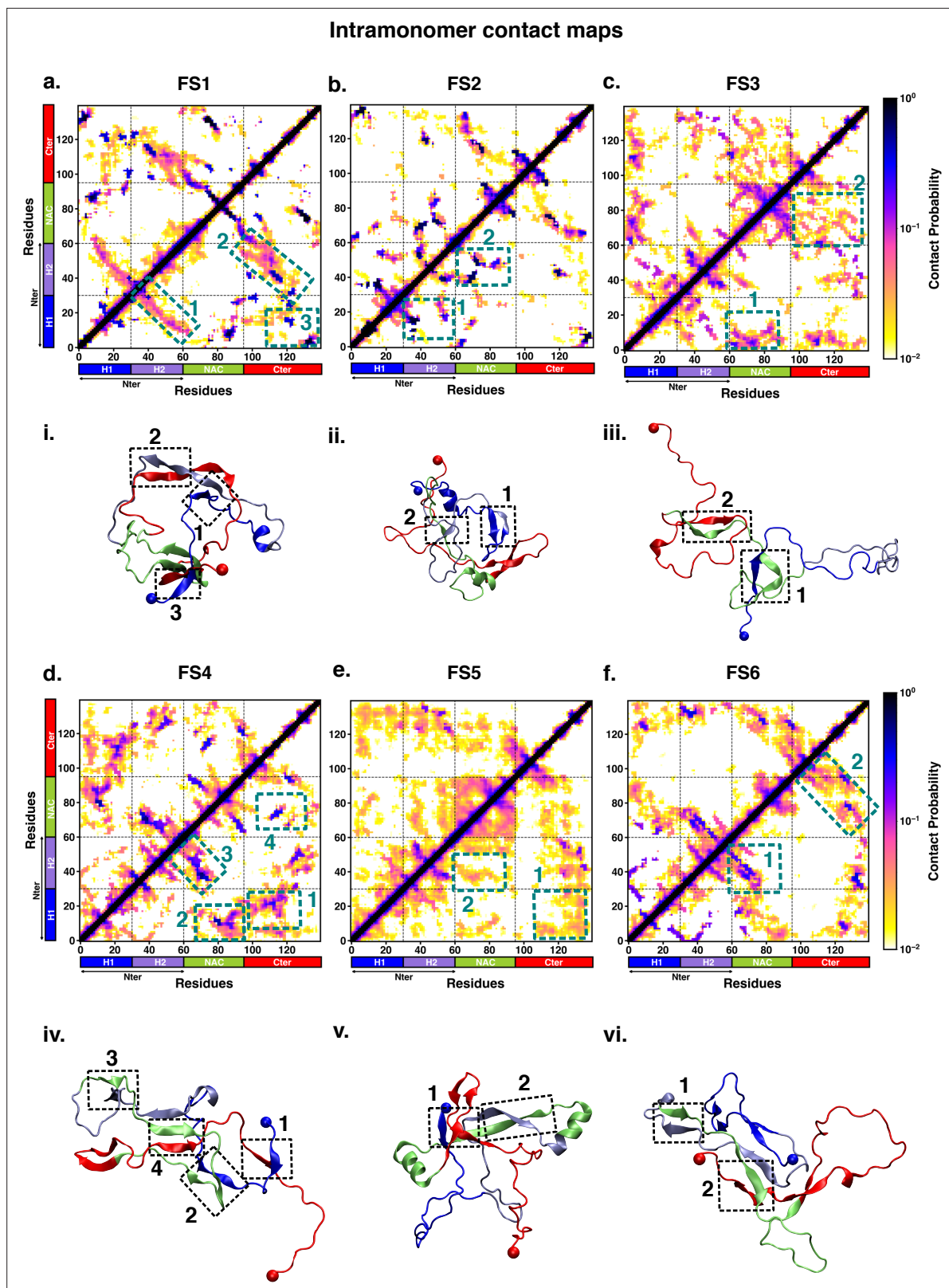


Figure 4. Intrapeptide residue-wise contact map of the macrostates of α S-Fasudil system indicates that the six states represent distinct ensembles. Intrapeptide residue-wise contact probability maps of (a) six macrostates (FS1 to FS6) in the presence of fasudil. A contact is considered if the C α atoms of two residues are within a distance of 8Å of each other. Axes denote the residue numbers. The color scale for the contact probability is shown at the extreme right of each panel of maps. The color bar along the axes of the plots represents the segments in the α S monomer. Specific contact regions are

Figure 4 continued on next page

Figure 4 continued

marked by boxes and numbered. These contacts are illustrated by representative snapshots and the corresponding contacts are similarly marked and numbered.

The online version of this article includes the following figure supplement(s) for figure 4:

Figure supplement 1. Residue-wise percentage secondary structure estimated for each of the six macrostates of α S monomer simulated in the presence of fasudil.

identity function between the input and output, as evident from the large reconstruction error of the test dataset (see **Figure 6—figure supplement 1a**). This overfitting issue was addressed by training a Denoising CVAE (DCVAE), where the contact map was intentionally corrupted with noise prior to training. This approach successfully mitigated the problem of overfitting as indicated by the loss profile of the test dataset (see **Figure 6—figure supplement 1b**). An implementation is provided in [GitHub](#) (copy archived at [JMLab-tifrh, 2024](#)).

The denoising performance was measured by calculating the structural similarity index measure (SSIM) between the original and denoised contact map. SSIM measures the similarity between two images as perceived by the human eye by considering the luminance, contrast and structure of the image. SSIM values can range from -1 to 1 , which represents perfect anti-correlation and perfect similarity, respectively ([Hore and Ziou, 2010](#)). The SSIM values is tabulated in **Figure 6**. We find that the SSIM values for all the metastable states of α S and α S-Fasudil system are close to 1 (see **Figure 6**), thereby indicating that the denoised contact map as obtained from the trained DCVAE model is very close to the quality of the original contact map (see **Figure 6—figure supplements 2–4**). This is also supported by the high Peak signal-to-noise ratio (PSNR). PSNR is the ratio of the maximum possible power of a signal to the power of corrupting noise that measures the fidelity of a reconstructed image compared to its original in decibels (dB) ([Hore and Ziou, 2010](#)). A higher value of PSNR indicates smaller difference between the original and denoised contact map (see **Figure 6**). Upon visualizing the latent space, it became evident that the contact map of the metastable states of α S and α S-Fasudil occupy distinct locations within the latent space, which indicates that the DCVAE model effectively captured the distinct patterns among the metastable states.

Fasudil exhibits conformation-dependent interactions with individual metastable states

We next sought to understand the specificities of the interactions of fasudil with the different metastable states of α S. Accordingly, we calculated the contact probabilities between fasudil and each α S residue. A contact between fasudil and a protein residue is considered to be present when the minimum distance between any heavy atom of fasudil and a protein residue is less than 0.6 nm. This is calculated for the entire population of each macrostate to determine the contact probabilities. The α S-Fasudil contact maps for the six macrostates are presented in **Figure 7a** and snapshots of the interactions are presented in **Figure 7b**.

The residue-interaction profiles of each of the six individual macrostates of α S with fasudil revealed distinct motifs of interaction across the length of the protein for the different macrostates (**Figure 7**). Previous investigations [Robustelli et al., 2022](#); [Tatenhorst et al., 2016](#) have reported dominant fasudil- α S interaction via the C-terminal region of the protein while it weakly interacts with the entire α S sequence. In state FS1, fasudil dominantly interacts with residues 133–139 of the C-terminal region, which consists of negatively charged residues D135 and E137 as well as the polar, neutral tyrosine residues (Y133 and Y136). In addition, there are also non-specific interactions of fasudil with hydrophobic residues in the NAC region (residues 71–74). In state FS2, fasudil interactions are spread across the entirety of the protein. These comprise hydrophobic residues in the N-terminal region, neutral polar residues (62–66) in the NAC region and residues 130–136 in the C-terminal region. Fasudil interacts strongly with the NAC region consisting of hydrophobic and polar, neutral residues in state FS3. Interactions also prevail with the negatively charged C-terminal region and residues 1–6 of the N-terminal region. While interacting non-specifically with residues in the amphipathic N-terminal region of state FS4, fasudil preferentially interacts with residues 121–140 of the C-terminal region via charge-charge and π -stacking interactions with the negatively charged residues (D135, E137, E139) and tyrosine side chains (Y125, Y133, Y136). In the most populated state FS5, fasudil interacts weakly with

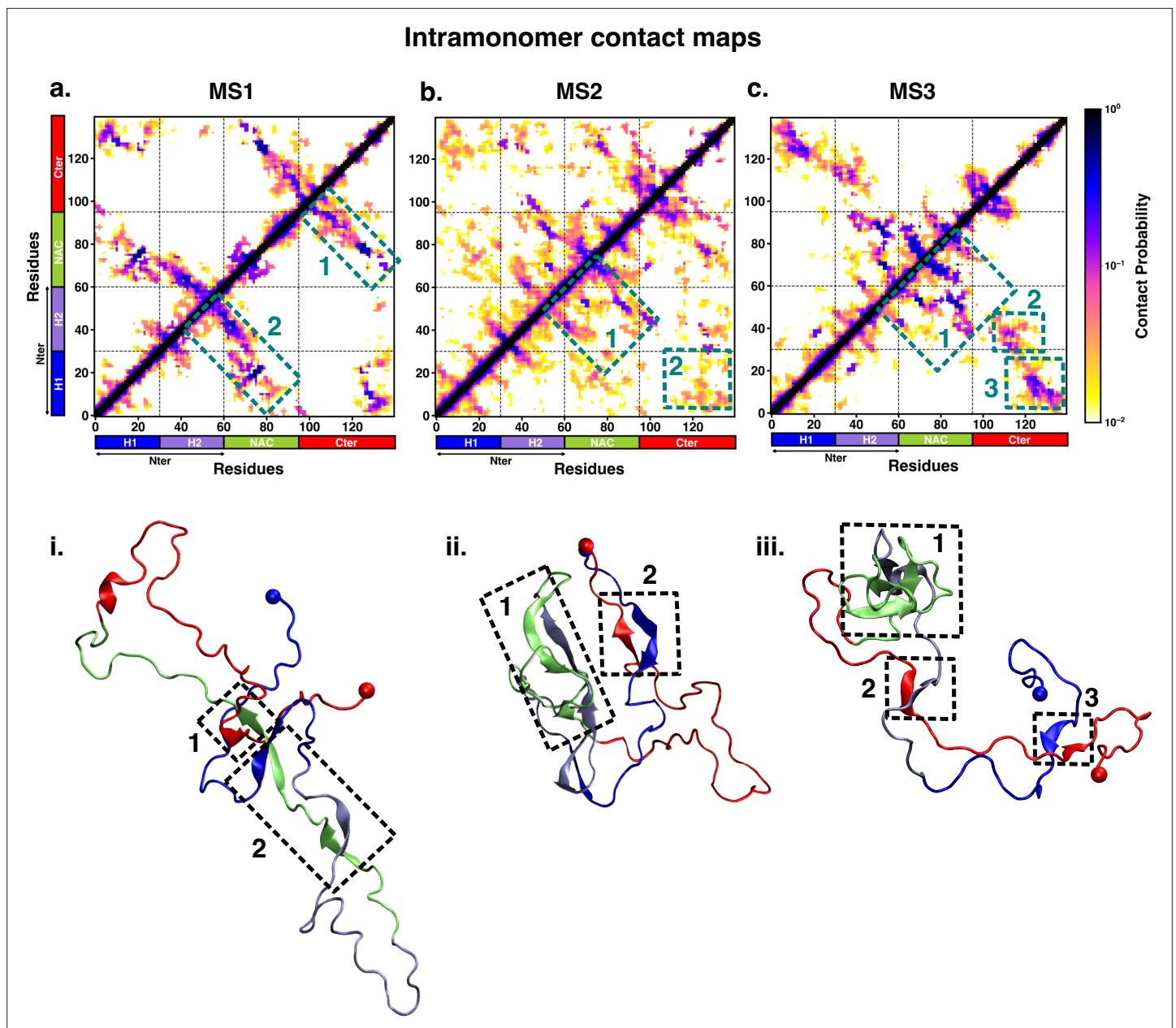


Figure 5. Intra-peptide residue-wise contact map of the macrostates of α S system indicates that the three states represent distinct ensembles. Intra-peptide residue-wise contact probability maps of the three macrostates (MS1 to MS3) in neat water. A contact is considered if the C_{α} atoms of two residues are within a distance of 8\AA of each other. Axes denote the residue numbers. The color scale for the contact probability is shown at the extreme right. The color bar along the axes of the plots represents the segments in the α S monomer. Specific contact regions are marked by boxes and numbered. These contacts are illustrated by representative snapshots and the corresponding contacts are similarly marked and numbered.

The online version of this article includes the following figure supplement(s) for figure 5:

Figure supplement 1. Residue-wise percentage secondary structure estimated for each of the three macrostates of α S monomer simulated in neat water.

the hydrophobic residues and glutamic acid residues while interactions with the C-terminal region 107–140 comprising of tyrosines and acidic amino acids are dominant. Lastly, in state FS6, the strongest interactions of fasudil comprise hydrophobic and polar residues in the range 118–121. There are also weak interactions with hydrophobic residues in the N-terminus(11–16, 38, 39) and NAC regions (71–74, 80–95).

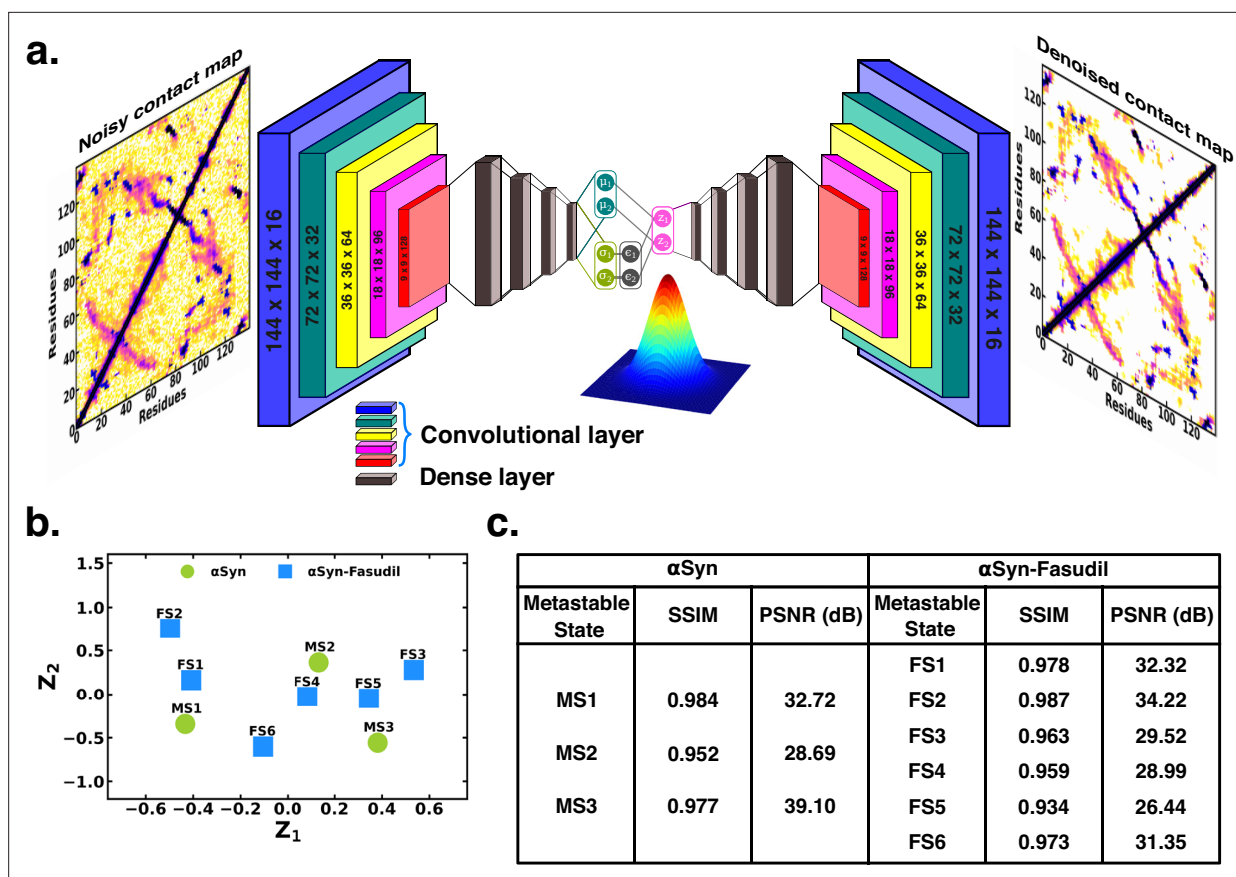


Figure 6. Projection of contact maps of α S and α S-Fasudil macrostates using denoising convolutional variational autoencoder (b) Latent space of the contact map of α S and α S-Fasudil metastable states. (c) SSIM and PSNR values for α S and α S-Fasudil metastable states.

The online version of this article includes the following figure supplement(s) for figure 6:

Figure supplement 1. Loss profile of (a) CVAE indicating overfitting (b) DCVAE indicating no overfitting.

Figure supplement 2. Denoised output of the contact map of the α S macrostates.

Figure supplement 3. Denoised output of the contact map of the α S-Fasudil macrostates FS1, FS2, and FS3.

Figure supplement 4. Denoised output of the contact map of the α S-Fasudil macrostates FS4, FS5, and FS6.

Robustelli et al., 2022 have reported that in the bound state, fasudil interactions with α S are favorable charge-charge and π -stacking interactions that form and break in a mechanism, they term as *dynamic shuttling*. The stacking interactions arise from the favourable orientation of fasudil's isoquinoline ring and the aromatic ring in the side chain of tyrosine residue. Analysis of the time-series of the formation of different intermolecular interactions indicated the formation of charge-charge and π -stacking interaction with residues Y125, Y133, and Y136 and the shuttling among these interactions causes fasudil to remain localized to the C-terminal region. Notably, we observe that each of the metastable, while preferentially interacting with the C-terminus also has distinct yet significant interactions with hydrophobic residues in the N-terminus and NAC regions.

Entropic signatures of small molecule binding

The overall analysis of the conformational ensemble of α S in presence of fasudil demonstrates that small-molecule binding substantially modulates the ensemble of α S. Our results indicate that the interaction of fasudil with α S residues governs the structural features of the protein. Estimation of the metastable states of α S shows that compared to aqueous conditions in which three metastable states are identified, the ensemble populated in the presence of fasudil in solution contains six metastable states. Since IDPs are described as an ensemble unlike folded proteins with a fixed structure, the

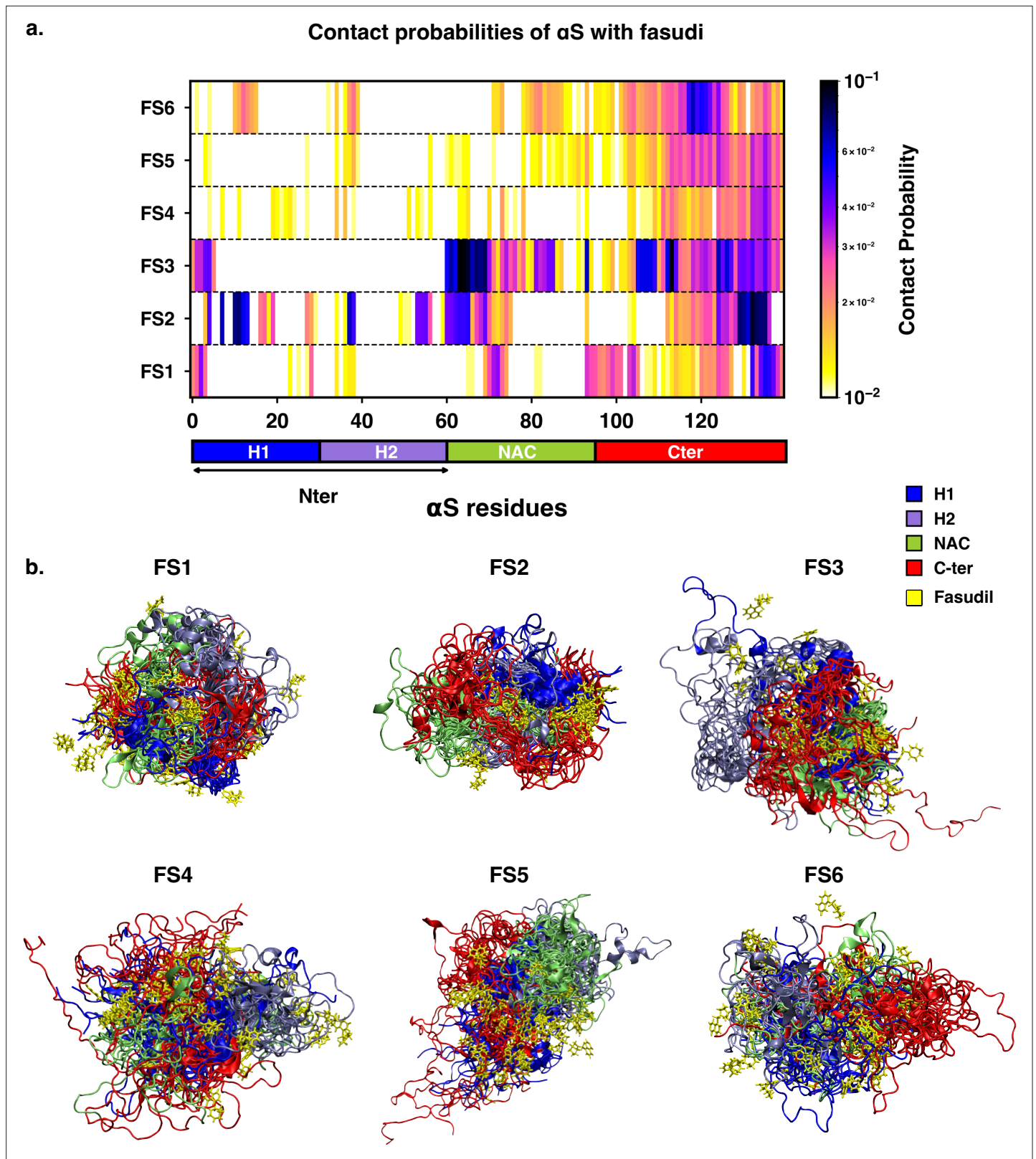


Figure 7. Contact map between fasudil and α S, indicates that fasudil interacts with distinct regions of α S in the six macrostates. (a) Contact probability map of residue-wise interaction of fasudil with α S. (b) Overlay of representative conformations from the six macrostates (FS1 to FS6) along with the bound fasudil molecules. The conformations are colored segment-wise as shown in the legend. The fasudil molecules, in licorice representation, are colored yellow.

effect of small molecules binding on disordered proteins or regions, is manifested as modulation of the whole ensemble by shifting the population of the states leading to a change in its conformational entropy (Heller et al., 2018; Heller et al., 2015; Flock et al., 2014). The common mechanism that is generally postulated is the *entropic collapse* or *folding upon binding* model. (Dyson and Wright, 2002) It describes the disorder-to-order transition that IDPs typically undergo upon interacting with its biological target. This transition leads to the shift in population to a more ordered, predominant state causing a loss of conformational entropy that is compensated by a corresponding enthalpic gain. Conversely, small-molecule binding may also lead to *entropic expansion* in which newer conformational states are populated that are otherwise undetectable (Heller et al., 2015). Additionally, in cases where there is no gross change in the conformational entropy and the binding is considered *fuzzy* (Bonomi et al., 2016), referred to as *isentropic shift*. The observations reported in the present investigation for α S monomer plausibly conform with the entropy-expansion model (Heller et al., 2018). According to this model, the interactions of small molecules with IDPs can promote the exploration of conformations with discernible probability, thereby expanding the disorderedness (conformational entropy) of the ensemble (Heller et al., 2018; Heller et al., 2015; Heller et al., 2020; Löhner et al., 2022). The small molecule interaction with α S operates transiently, and not strongly as seen in structured proteins. These transient local interactions stabilise the conformations that are inaccessible in the absence of the small molecule.

To gain deeper insights into the entropic effect of small-molecule binding to α S, we further estimated protein conformational entropy as well as water (solvent) entropy. We evaluated the entropy of water using the 2PT (Lin et al., 2003; Lin et al., 2010) method. First, we determined the molar entropy of pure water that yielded an entropy value of $55.5 \text{ J mol}^{-1} \text{ K}^{-1}$, which is close to the reported value of $54.4 \text{ J mol}^{-1} \text{ K}^{-1}$ from free energy perturbation calculation (Mukherjee and Schäfer, 2023). However, the molar entropy is marginally reduced in all the macrostates of both α S and α S-fasudil systems, with values ranging between $55.01 \text{ J mol}^{-1} \text{ K}^{-1}$ to $55.25 \text{ J mol}^{-1} \text{ K}^{-1}$, with difference from that of bulk water being negligible. Upon closer investigation of the translational and rotational components of the total water entropy, we find that the decrease of the translational component is more pronounced as compared to the rotational component. This implies that α S in its apo state or in the presence of fasudil has a comparatively greater impact on the translational component than on the rotational component of the entropy of water (see **Figure 8—figure supplement 1**). Nonetheless, the entropy analysis suggests that the presence of fasudil has an insignificant impact on the entropy of water.

Next, we estimated the protein conformational entropy using the PDB2ENTROPY program (Fogolari et al., 2018) that implements the Maximum Information Spanning Tree (MIST) approach (King and Tidor, 2009; King et al., 2012) on sets of torsional angles of the protein. The entropy is calculated from probability distributions of torsional degrees of freedom of the protein relative to random distribution i.e. corresponding to a fully flexible protein chain. The total protein entropy values for the metastable states in water (MS1 to MS3) and fasudil-bound states (FS1 to FS6) are shown in **Figure 8**. While the conformational entropies of the ligand-free states, MS1 to MS3, range from -72 to $-75 \text{ J mol}^{-1} \text{ K}^{-1}$, the entropy of fasudil-bound states (FS1 to FS6) vary from -70 to $-86 \text{ J mol}^{-1} \text{ K}^{-1}$, indicating that binding of fasudil to α S leads to a broadening of the span of entropy of these states. The highest entropy among the metastable states in the two environments correspond to FS5 and MS2 with values of $-70.8 \text{ J mol}^{-1} \text{ K}^{-1}$ and $-72.3 \text{ J mol}^{-1} \text{ K}^{-1}$, respectively. These states are also the most populated metastable states. MS1 and MS3 states have entropy values of $-73.9 \text{ J mol}^{-1} \text{ K}^{-1}$ and $-74.5 \text{ J mol}^{-1} \text{ K}^{-1}$, respectively. Notably, the presence of fasudil and its interactions with the protein chains can elicit the exploration of states that are more ordered in nature or increasingly deviated from a random chain. The largest deviation from a fully disordered chain is attained by the state FS1 with entropy value of $-85.3 \text{ J mol}^{-1} \text{ K}^{-1}$, followed by FS2 having entropy of $-83.1 \text{ J mol}^{-1} \text{ K}^{-1}$. FS4 is also relatively more ordered than FS5 with entropy of $-76.2 \text{ J mol}^{-1} \text{ K}^{-1}$. FS3 and FS6 are closely disordered as FS5 with entropy values of $-73.2 \text{ J mol}^{-1} \text{ K}^{-1}$ and $-72.5 \text{ J mol}^{-1} \text{ K}^{-1}$, respectively. These observations of the total extent of disorderedness of the metastable states clearly indicate that the interactions of fasudil with α S residues can restrict their conformational degrees of freedom thus entrapping the protein in specific conformational states, thus manifesting as overall decrease in the protein entropy. Considering the diffuse binding of fasudil across several residues along the entire length of the protein, we further estimated and analysed how the entropic contributions are affected at the residue level (shown

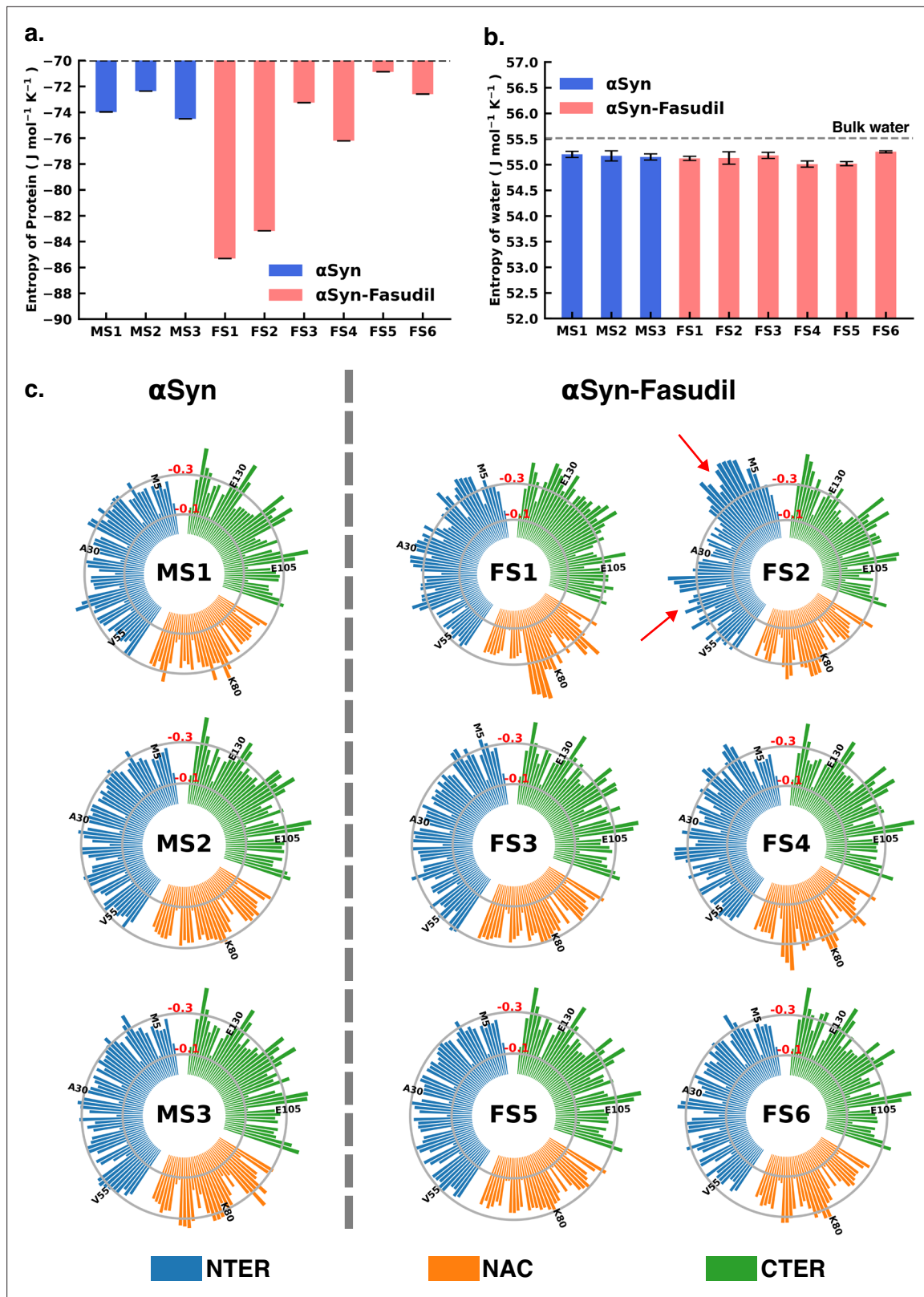


Figure 8. Entropy of αS system is modulated to a greater extent in presence of fasudil, with insignificant contribution from the solvent entropy. (a) Total Protein entropy of the metastable states of αS and $\alpha\text{S-Fasudil}$ calculated from the torsion angles, relative to a fully flexible chain (b) Total water entropy in αS and $\alpha\text{S-Fasudil}$ system and (c) Residue-wise backbone entropy within the αS and $\alpha\text{S-Fasudil}$ states.

Figure 8 continued on next page

Figure 8 continued

The online version of this article includes the following figure supplement(s) for figure 8:

Figure supplement 1. Translational and Rotational entropy of water.

Figure supplement 2. Residue-wise entropy of α S states populated in water and in the presence of fasudil.

Figure supplement 3. Residue-wise sidechain entropy of α S states populated in water and in the presence of fasudil.

in **Figure 8**, **Figure 8—figure supplements 2 and 3**). The total residue-wise entropy presented in **Figure 8—figure supplement 2** shows that the residues in the N-terminal region of fasudil-bound states FS1 and FS2, exhibit greater entropic fluctuations i.e. more negative values from the apo metastable states (MS1 to MS3), suggesting more ordered or restricted fluctuations of these residues. We note that the propensity of fasudil binding to the N-terminal residues is higher in FS1 and FS2, relative to other states (see **Figure 7**). We further decoupled the residue entropy into backbone and sidechain components; this analysis is depicted in **Figure 8**, **Figure 8—figure supplement 3**, respectively. Comparison of the sidechain entropies, computed from the χ torsion angles, for the unbound and bound states, does not show significant alteration upon small-molecule binding. Conversely, increased fluctuations in residue backbone entropies are observed in response to small-molecule binding (see **Figure 8**). The metastable states, MS1 to MS3, populated in water have minimally fluctuating residue entropies amongst the states. In contrast, significant fluctuations in residue entropies are noted predominantly in the N-terminal region (residues 1–60) and to some extent in the NAC region for states FS1 and FS2. These residues in the N-terminal region in FS1 and FS2 experience a drop in backbone entropy owing to the restricted degrees of freedom from fasudil binding. The decrease in residue backbone entropy correlates with fasudil binding hotspots and therefore can be ascribed to small-molecule interactions. Hydrophobic residues in the NAC region show decrease in entropy in ligand-binding locations mainly in state FS1, FS2, and FS4. However, intriguingly, we note that the correlation between fasudil interactions and residue entropy is not observed in state FS3. Finally, in the negatively charged C-terminal region, where fasudil preferentially targets the entire region in the α S protein, residue-wise entropic fluctuations are minimal compared to the apo state. Robustelli et al. refer to the interaction of fasudil with the C-terminal region as dynamic shuttling, in which breaking and formation of interactions with nearby residues occur simultaneously. This dynamic shuttling mechanism could possibly minimise the backbone entropic fluctuations compared to the relatively specific binding in the N-terminus and NAC regions.

Exploring interplay of conformational entropy and mean first passage time in protein conformational transitions

The Mean First Passage Time (MFPT) for transition among the metastable states provides valuable information for understanding the dynamics and conformational changes in the system. MFPT rate defines the average time it takes for a system to transition between one metastable state to another. A lower MFPT value suggests a favourable transition among the states. The transition timescales for α S indicate that in the presence and absence of fasudil, the transition to the most populated state is preferred (see **Figure 9**). For α S-Fasudil system, the transition to the most populated states takes place over timescales ranging from 74 μ s and 766.5 μ s, whereas for the apo state it occurs between 29.8 μ s and 37.4 μ s. Upon comparing these transition times with those to other macrostates, it becomes evident that these timescales represent the shortest durations. This observation suggests that the transition to the MS2 and FS5 state in α S and α S-Fasudil system is the most favoured over all other states in the system.

While MFPT offers insights into the kinetics of the metastable states, the thermodynamic property, entropy, characterizes the diversity and disorder within the system. Although a state with higher population might not always have the higher entropy, we observed that in our case, the state with the highest population also has the highest protein entropy, for both α S and α S-Fasudil system. States with higher entropy are thermodynamically favored which in turn can affect the rates of state transitions, potentially influencing MFPT values. To decipher this, we have compared the transition times of each metastable state to the most populated state with the entropy of the state, and found a strong degree of correlation (see **Figure 9e**). This suggests that a state with higher entropy, which also has more number of accessible microstates, potentially provide multiple pathways for other macrostates

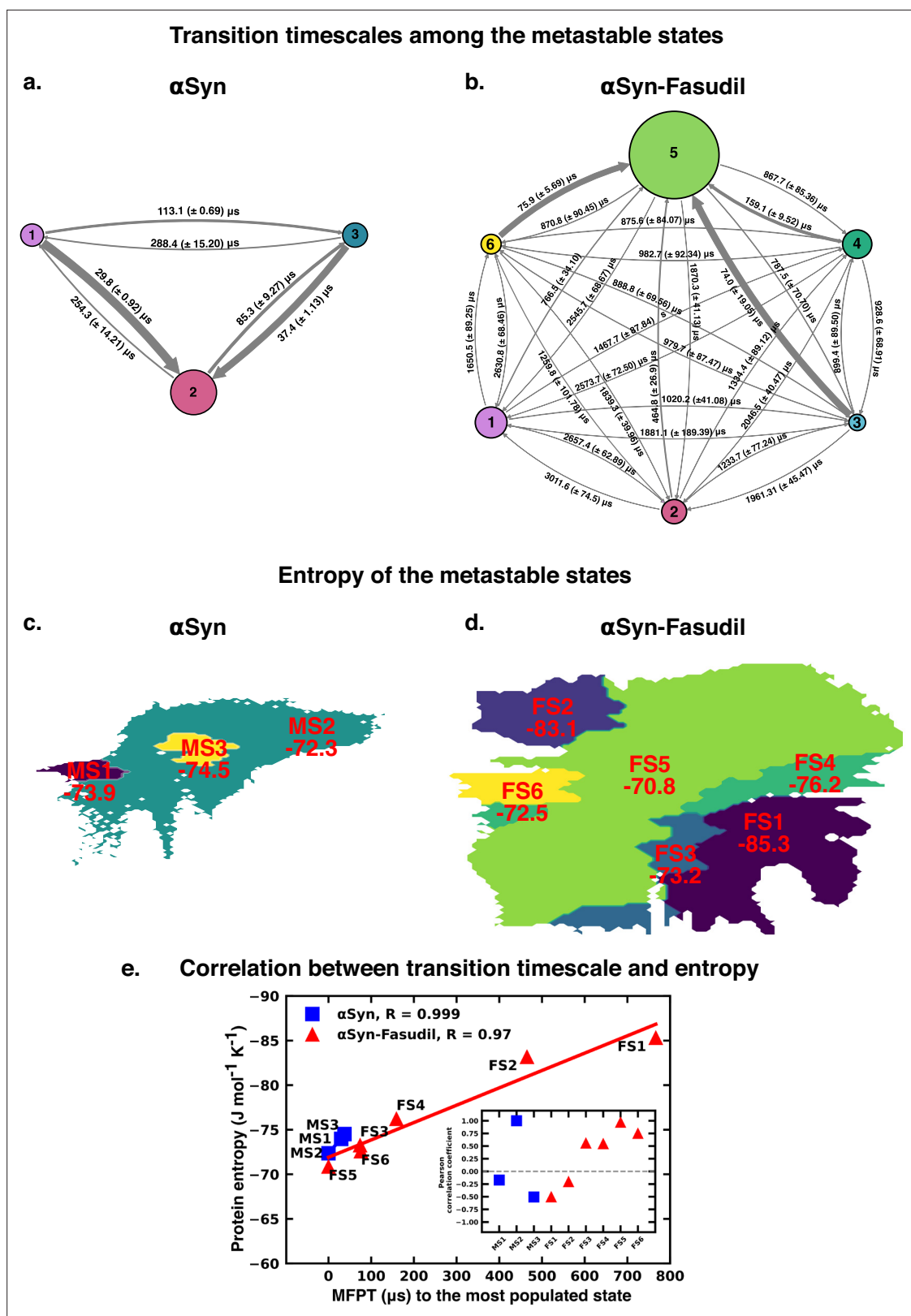


Figure 9. Mean first passage time for α S and α S-Fasudil ensembles indicates that a decrease in the transition time to the most entropic state in presence of fasudil, thereby potentially trapping S in the intermediate states. The rate network obtained from MFPT analysis for the (a) 3-state MSM of the α S ensemble in neat water and (b) 6-state MSM of the α S-Fasudil ensemble. The sizes of the discs is proportional to the stationary population of the respective state. The thickness of the arrows connecting the states is proportional to the transition rate ($1/\text{MFPT}$) between the two states and the MFPT

Figure 9 continued on next page

Figure 9 continued

values are shown on the arrows. (c, d) Projection of conformation sub-ensemble PCCA + for α S and α S-Fasudil system in latent space, with respective protein entropy values (in unit of $\text{J mol}^{-1} \text{K}^{-1}$) annotated on top of it. (e) Correlation between protein entropy and transition time to the major state for α S and α S-Fasudil system. The inset plot corresponds to the correlation between entropy and transition to the macrostate for α S and α S-Fasudil system.

to access this state for both α S and α S-Fasudil system. However, as compared to α S in its apo state, the apparent increase (~ 2 – 20 times) in the time to transition to the major state in α S-Fasudil system, suggests that fasudil may potentially trap the protein conformations in the intermediate states, thereby slowing down α S in exploring the large conformational space in presence of fasudil.

Conclusions

The conformationally dynamic disordered proteins challenge the application of conventional structure-based drug-design methods that relies on the specificity of a binding pocket, based on which the drug molecule is designed and optimized. A general framework describes the effect of small-molecule binding to IDPs by modulation of the disordered ensemble by increasing or decreasing its conformational entropy. Here, we characterised the effect of the small molecule, fasudil, on the conformational characteristics of the quintessential IDP, α -synuclein. Fasudil has been experimentally reported to curb the aggregation of α S both in vivo and in vitro. The ensemble view of long time scale atomic simulations of fasudil binding to α S monomer displays a fuzzy ensemble of α S with preferential binding of fasudil to the C-terminal region via charge-charge interactions and aromatic stacking. Using a deep neural-network based machine learning approach, variational autoencoder, we reconstructed a high dimensional feature that is C α -pairwise distance into a two-dimensional latent space. We built an MSM using this data to delineate the metastable conformational states of α S and identify the differences that arise from small-molecule interactions. Comparison of the metastable states of α S in the presence and absence of the small molecule fasudil revealed that the small-molecule binding navigates the structural landscape of the protein. Importantly, more number of metastable states are populated indicating that small-molecule mediated conformational modulation led to entropy expansion (see **Figure 9d–e**). Atomic-level MD simulations have slowly emerged as a valuable tool for complementing experimental measurements of disordered proteins and providing detailed descriptions of their conformational ensembles (**Dedmon et al., 2005; Allison et al., 2009; Ullman et al., 2011; Esteban-Martín et al., 2013**). With improvements in force fields (**Best et al., 2014; Huang et al., 2017; Robustelli et al., 2018**) that model disordered proteins, the accuracy of MD simulations has dramatically increased, as assessed by their agreement with a wide variety of experimental measurements. Integrative approaches use a combination of experimental restraints such as chemical shifts, paramagnetic relaxation enhancement (PRE; **Dedmon et al., 2005; Bertocini et al., 2005; Allison et al., 2009; Esteban-Martín et al., 2013**), residual dipolar coupling constants (RDCs; **Bertocini et al., 2005; Bernadó et al., 2005; Marsh et al., 2008**), small-angle X-ray scattering (SAXS; **Bernadó et al., 2007; Ahmed et al., 2021**) and computational models from MD simulations to generate a more accurate description of the ensemble of disordered proteins. More recent advances include statistical approaches such as Bayesian inference models like MetaInference (**Bonomi et al., 2016**), Experimental Inferential Structure Determination (EISD; **Brookes and Head-Gordon, 2016; Lincoff et al., 2020**) and Maximum Entropy formulations (**Roux and Weare, 2013; Boomsma et al., 2014; Cavalli et al., 2013**) that also take into account the experimental and back-calculation model errors and uncertainties. Such improved approaches to leverage MD simulations with improved force fields and experimental knowledge-based ensemble generation have shown tremendous potential for describing molecular recognition mechanisms of IDPs with other protein partners; scenarios such as folding-upon-binding (**Dyson and Wright, 2002; Robustelli et al., 2020**) or self-dimerisation. More importantly, a set of recent initiatives have focussed on molecular characterisation of the interaction of small-molecules with IDPs via effective combination of atomistic simulation and biophysical experiments (**Heller et al., 2020; Robustelli et al., 2022**).

In this regard, the present investigation takes a step ahead via a quantitative and statistical dissection of the α S ensemble in presence of a small-molecule. The implementation of Markov state model in the present work clearly brings out the feasibility of metastable states that would have a finite 'stationary' or equilibrium population in presence of small-molecule or in neat water. The uniqueness of these states is supported from the projection of their inter-residue C α contact map of the macrostates

using DCVAE model. Our systematic investigation of the impact of fasudil on α S monomer ensemble reveals the altering effects of small-molecule interaction on α S ensemble by tuning the entropy of these states. Insights from studies targeting structural ensemble modulation by small-molecule can be effectively harnessed in designing drug-design strategies for aggregation diseases.

Methods

Monomer simulation protocol

A 1500 μ s long all-atom MD simulation trajectory of α S monomer in aqueous fasudil solution was simulated by D. E. Shaw Research with the Anton supercomputer that is specially purposed for running long-time-scale simulations (Robustelli et al., 2022). The protein was solvated in a cubic water-box with 108 sides and containing 50 mM concentration of Na or Cl ions. Protein, water molecules and ions present in the system were parameterised with the a99SB-*disp* force field (Robustelli et al., 2018). The small-molecule, fasudil, was parameterised using the generalised Amber force-field (GAFF). Production run of the system was performed in the NPT ensemble. Further details of the MD simulation can be obtained by referring to the original work by Robustelli et al., 2022. The extensive simulation trajectories were generously made available by D. E. Shaw Research upon request. To generate α S ensemble in water, we performed multiple simulations of α S monomer. An IDP is not represented by a fixed structure. Therefore, it would be most appropriate to run multiple simulations starting from different initial structures and simulate the local environment around those structures; thus generating

Table 1. The size of the simulation box for the 23 α S simulations.

Conformation number	Cubic Box size (nm)
1	12.14
2	11.33
3	13.55
4	10.41
5	10.97
6	10.44
7	10.10
8	11.44
9	10.57
10	10.75
11	11.34
12	10.35
13	13.38
14	10.08
15	10.38
16	11.78
17	11.92
18	9.37
19	10.15
20	11.16
21	11.24
22	11.46
23	12.00

an ensemble effectively sampling the phase space. Accordingly, for initiating the apo simulations, instead of biasing the initial structure (using the starting structure used for simulations with fasudil), we randomly chose 23 different conformations from previously reported (Robustelli et al., 2018) multi- μ s α S trajectory in water. We have also determined the R_g values corresponding to the starting structures used to start the apo simulations. It is evident that the 23 starting conformations chosen represent the entire span of the R_g space that is sampled in the fasudil ensemble (see **Figure 1—figure supplement 1**). In addition, we have also calculated the RMSD of the starting structures with respect to the starting structure of the fasudil simulation (see **Figure 1—figure supplement 7**), secondary structure content (see **Figure 1—figure supplements 8–10**) and solvent accessible surface area (SASA; see **Figure 1—figure supplements 11–12**) to validate our choice of starting structures. The simulation details are described below. Each protein conformation is solvated in a cubic box (see **Table 1**) of explicit water molecules and Na or Cl ions were added to a concentration of 50 mM, to maintain the same conditions as used in the simulations with fasudil. The all-atom a99SB-*disp* force field and water model (Robustelli et al., 2018) were used to simulate the protein and solvent. The unbiased MD simulations were performed using the Gromacs simulation package (Lindahl et al., 2020). A timestep of 2 fs and the leap-frog integrator was used. The simulation was performed in the isothermal-isobaric (NPT) ensemble at an average temperature of 300 K, maintained using the v-rescale thermostat (Bussi et al., 2007) with a relaxation time of 0.1 ps and a pressure of 1 bar using the Parrinello-Rahman barostat (Parrinello and Rahman, 1981) with a time constant of 2 ps for coupling. The Verlet cutoff scheme (Páll and Hess, 2013) with 1.0 nm cutoff was applied for Lennard-Jones interactions and short-range electrostatic interactions. Long-range electrostatic interactions were computed using the Particle-Mesh Ewald (PME) summation method (York et al., 1993) and covalent bonds involving hydrogen atoms were constrained using the LINCS algorithm (Hess et al., 1997). All the systems were minimized using the steepest descent algorithm, followed by equilibration in the isothermal-isochoric (NVT) and subsequently in the NPT ensemble with position restraints on the heavy atoms of the protein. Twenty-three simulations, each starting from different conformations, were performed. These simulation timescales are variable, ranging from 1 to 4 μ s, amounting to a cumulative length of \sim 62 μ s.

We have compared the ensemble properties of the α S in the presence and absence of fasudil. As seen in **Figure 1—figure supplement 1**, the distributions of R_g for both the simulations have a significant overlap, thereby indicating that in terms of extent of compaction, the conformational sampling of the protein is comparable in the presence and absence of fasudil. We have also compared other structural features such as end-to-end distance (R_{ee} ; see **Figure 1—figure supplement 2**), SASA (see **Figure 1—figure supplement 3**) and secondary structure content (see **Figure 1—figure supplements 4–6**). We have found that these features are highly comparable for both the ensembles, which clearly indicates that the α S simulations we performed from different starting structures have effectively sampled the phase space as the single long simulation of the α S-Fasudil system.

Dimension reduction using β -variational autoencoder

The probability distribution $p(x)$ of a multi-dimensional variable x can be determined using various approaches (Bengio and Bengio, 2000; Larochelle and Murray, 2011), one of them being the latent variable model. In this model, the probability distribution is modeled as a joint distribution $p(x, z)$ of the observable variable x and the hidden latent variable z which is given as $p(x, z) = p(x|z)p(z)$, where $p(z)$ is the prior distribution over the latent variables z and $p(x|z)$ is the likelihood of generating a true data, given z . The $p(x)$ is determined by marginalization over all the latent variables giving us,

$$p(x) = \int p(x|z)p(z)dz \quad (2)$$

However, this integral is often intractable thereby constraining the calculation of the true posterior distribution $p(z|x)$. This can be solved using, variational inference, which aims to estimate the true posterior using an approximate distribution $q(z|x)$, using the encoder part of VAE which parameterizes the distribution using a mean vector and a variance vector. The goal is to determine $q(z|x)$ which is closest to the true posterior by minimizing the Kullback-Leibler (KL) divergence between them which is given as

$$D_{KL}(q(z|x)||p(z|x)) = \int q(z|x) \log \left(\frac{q(z|x)}{p(z|x)} \right) dz \quad (3)$$

Upon expanding and rearranging, we have

$$\log p(x) - D_{KL}(q(z|x)||p(z|x)) = \mathbb{E}_{z \sim q(z|x)} \log p(x|z) - D_{KL}(q(z|x)||p(z)) \quad (4)$$

where $p(x)$ is the evidence. The VAE is trained so that we maximize the log-likelihood of the evidence and minimize the KL divergence between the estimated and true posterior. The loss function of VAE is given by the negation of **Equation 2** and is given as,

$$\text{Loss}_{VAE} = -\mathbb{E}_{z \sim q(z|x)} \log p(x|z) + D_{KL}(q(z|x)||p(z)) \quad (5)$$

The training objective of VAE being minimizing this loss function to generate a realistic data. The VAE model can be further improved by keeping the distance between the approximated and the true posterior under a small distance, δ . This is achieved by maximizing the equation

$$\mathcal{F}(\beta; x, z) = \mathbb{E}_{z \sim q(z|x)} \log p(x|z) - \beta(D_{KL}(q(z|x)||p(z)) - \delta) \quad (6)$$

which modifies the loss function to

$$\text{Loss}_{\beta-VAE} = -\mathbb{E}_{z \sim q(z|x)} \log p(x|z) + \beta D_{KL}(q(z|x)||p(z)) \quad (7)$$

where β is the Lagrangian multiplier (under the Karush-Kuhn-Tucker conditions). This enhances the probability of generating an actual data and is known as β -VAE (**Higgins et al., 2016**). The expectation term in the loss function of β -VAE requires a point z to be sampled from the posterior distribution. This is performed using the reparametrization trick

$$z = \mu + \sigma \odot \epsilon \quad (8)$$

where ϵ is sampled from a Standard Normal distribution. This makes the VAE model trainable using the backpropagation algorithm.

In our work we have used C_{α} - C_{α} pairwise distance (excluding i , $i + 1$ and $i + 2$ distances) sampled at every 9 ns of the entire 1.5 millisecond of the α S protein in presence of fasudil trajectory as our input. We have built our model using Tensorflow backend (**Abadi et al., 2016**) with Keras library. Training was performed with 80% of the data while 20% was used for testing. We have employed a symmetric VAE where the input layer consists of 9453 nodes followed by four hidden layers having 4096, 512, 128 and 16 nodes respectively, and 2 nodes in the latent layer. Hyperbolic tangent (*tanh*) is used as our activation function with weights being initialized using the *glorot_uniform* (**Glorot and Bengio, 2010**) method in all the layers except for the output layer where we have used *sigmoid* activation. Adaptive Moment Estimation (Adam **Kingma and Ba, 2014**) was used as the optimizer with a learning rate of 1×10^{-4} . The model was trained for 300 epochs with a batch size of 64. The relevant Python implementations are provided in [Github](#) (copy archived at **JMLab-tifrh, 2024**).

Building the denoising convolutional variational autoEncoder

The Denoising Convolutional Variational Autoencoder (DCVAE) model employs five convolutional layers which incorporate filters of sizes 16, 32, 64, 96, and 128, with corresponding kernels of size 3x3. Strides are set at 1, 2, 2, 2, 2, and 1 for efficient feature extraction. This is followed by six densely connected layers with neuron counts of 4096, 2048, 1024, 256, 64 and 16 in the successive dense layers. The latent layer is comprised of 2 neurons. The decoder mirrors the encoder architecture, to reconstruct back the input contact maps. Hyperbolic tangent activation is used in all the layers, excluding the output layer where *sigmoid* activation is used. Adam optimizer with a learning rate of 5×10^{-4} is used for training the DCVAE. The model was trained for 300 epochs with a batch size of 32 with $\beta = 1 \times 10^{-12}$. Training was performed with 90% of the data while 10% was used for testing.

At first, we added a noise to 200 replicas of contact map for each of the states of α S and α S-fasudil. The noise was generated stochastically from a standard normal distribution which was multiplied by a noise factor of 0.035 and added to each of the replicas, followed by scaling the dataset so that the values range between 0 and 1 using the formula

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (9)$$

where x_{min} and x_{max} are the minimum and maximum values of the input contact map. In order to assess the denoising performance, we have calculated the structural similarity index measure (SSIM) values (**Hore and Ziou, 2010**). The SSIM value can range between -1 to $+1$, thereby indicating the structural dissimilarity or similarity between the original (x) and denoised (y) contact maps. Mathematically, this is expressed as

$$SSIM(x, y) = [l(x, y)]^\alpha \cdot [c(x, y)]^\beta \cdot [s(x, y)]^\gamma \quad (10)$$

where the functions $l(x, y)$, $c(x, y)$, and $s(x, y)$ compare the luminance, contrast and structure between the contact maps and is given as

$$\begin{aligned} l(x, y) &= \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \\ c(x, y) &= \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \\ s(x, y) &= \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \end{aligned} \quad (11)$$

The variables μ_i and σ_i represent the pixel sample mean and variance of i , where $i = \{x, y\}$. σ_{xy} is the covariance of the x and y contact maps. The exponents α , β and γ controls the weighted contribution to the SSIM value and are set to 1. The constants $C_1 = (k_1L)^2$, $C_2 = (k_2L)^2$ are small positive values introduced to prevent instability and division by zero. They are used for numerical stability and to avoid problems when the means and variances of the contact maps are close to zero. We have set $k_1 = 0.01$, $k_2 = 0.03$, and $C_3 = C_2/2$ as per convention (**Wang et al., 2004**).

In addition, we have also calculated the peak Signal-to-Noise Ratio to evaluate our denoising performance and is calculated as

$$PSNR = 10 \cdot \log_{10} \left(\frac{R^2}{MSE} \right) \quad (12)$$

where, R is the maximum pixel value (255 in our case) and MSE is the mean-square error between the original and denoised image and is given as

$$MSE = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N [O_{ij} - D_{ij}] \quad (13)$$

where M is the height (number of rows) of the image, N is the width (number of columns) of the image and O_{ij} and D_{ij} is the pixel intensity at position (i, j) for the original and denoised image, respectively. The relevant Python implementations are provided in [Github](#) (copy archived at **JMLab-tifrh, 2024**).

Building a Markov State Model of monomer ensemble

The 2D latent space reconstructed using β -VAE described in the previous subsection was then used to build an MSM to elicit the metastable states underlying the ensembles. We employed PyEMMA (**Scherer et al., 2015**), an MSM building and analysis package. The 2D data is subjected to k -means clustering. We used the VAMP-2 score to determine the number of microstates. By varying the number of microstates from 10 to 220, we found that the VAMP-2 score saturated at higher numbers for both α S and α S-Fasudil simulations. We find that the VAMP-2 score has saturated at 180 microstates (see **Figure 3—figure supplements 10–11**). A transition matrix was then built by counting the number of transitions among the microstates at a specific lag time. To choose the lag time, the implied timescale (ITS) or relaxation timescales of the systems were calculated over a range of lag times and plotted as a function of lag time. The timescale at which the ITS plot levels off is chosen as the lag time to build the final MSM. The ITS plots corresponding to the neat water and aqueous fasudil systems are presented in **Figure 3**. Accordingly, MSM lag times of 36 and 32 ns were selected for neat water fasudil systems, respectively. At these lag times, by identifying the gaps between the ITS, a three-state model for neat water and a six-state model for aqueous fasudil system was built. Lastly, the transition path theory (**Weinan and Vanden-Eijnden, 2006; Metzner et al., 2009; Noé et al., 2009**) was used to ascertain the transition paths, fluxes and timescales in these models. Bootstrapping was performed to estimate the errors associated with the state populations and transition timescales. In ten iterations, the model

was rebuilt after eliminating a randomly selected trajectory followed by calculation of the state populations and transition timescales. The mean value and standard deviations of the values collected by bootstrapping are reported. A Chapman-Kolmogorov test was performed for the MSMs of the neat water and small molecule ensembles, that tests the validity of the model prediction at longer time scales (see **Figure 3—figure supplement 3**).

Entropy of water

In order to calculate the entropy of water, simulations were performed using GROMACS (version 2022.3). A single snapshot was chosen from each macrostate of α -synuclein in its apo state and in presence of fasudil, which were then solvated in a cubic box, placed within 1 nm from the box edge. The system was neutralized using NaCl and the salt concentration was kept at 50 mM. The system was energy minimized using the steepest descent algorithm, followed by equilibration in the canonical ensemble for 2 ps. The average temperature was maintained at 300 K using velocity rescale thermostat with a time constant of 0.1 ps. Following this, NPT equilibration was performed for 2 ps to maintain the pressure at 1.0 bar using the Berendsen barostat. Finally, three independent production runs were performed from the equilibrated structures in the NPT ensemble using the velocity-verlet integrator with a timestep of 1 fs for a timescale of 20 ps with frames saved at every 4 fs. In addition, we have performed three simulations of neat water using the parameters as described for the macrostates.

The entropy of water was estimated using the DoSPT program (Caro et al., 2016; Caro et al., 2017) that computes the thermodynamic properties from MD simulation in the framework of the 2-phase-thermodynamics (2PT) method (Lin et al., 2003; Lin et al., 2010). This is achieved from calculating the density of state (DoS) function, $I(v)$ which is the Fourier transform of the velocity auto correlation function,

$$I(v) = \frac{1}{k_B T} \sum_{l=1}^N \sum_{k=1}^3 \lim_{\tau \rightarrow \infty} \frac{m_l}{\tau} \left| \int_{-\tau}^{\tau} v_l^k(t) \exp(-2i\pi vt) dt \right|^2 \quad (14)$$

where m_l is the mass of atom l , v_l^k is the velocity of atom l along the k direction and N is the total number of atoms in the system. DoS can be represented as a sum of translational (trn), rotational (rot), and vibrational (vib) motions (Lin et al., 2010).

$$I(v) = I_{trn}(v) + I_{rot}(v) + I_{vib} \quad (15)$$

For water, the 2PT model overcomes the anharmonic nature of the low frequency modes (diffusive motions and libration) by decomposing the DoS into solid-like (I^s) and gas-like (I^g) contribution for the translational and rotational components. The solid-like DoS is treated using the harmonic oscillator model, whereas the gas-like DoS is described using the Enskog hard sphere (HS) theory for translation and the rigid rotor (RR) model for rotation. The entropy is calculated by integrating the DoS, weighted by the corresponding weighting functions W ,

$$S_k = k_B \left[\int_0^\infty I_k^s(v) W_k^s(v) dv + \int_0^\infty I_k^g(v) W_k^g(v) dv \right] \quad (16)$$

where $k = \text{trn, rot, or vib}$ and the weighting functions are given as

$$\begin{aligned} W^s(v) &= \frac{\beta h v}{\exp(\beta h v) - 1} - \ln[1 - \exp(-\beta h v)] \\ W_{trn}^g &= \frac{S^{HS}}{3k_B} \\ W_{rot}^g &= \frac{S^{RR}}{3k_B} \end{aligned} \quad (17)$$

where $\beta = (k_B T)^{-1}$ and h is the planck constant (Huang et al., 2011).

Protein conformational entropy

We calculated the protein conformational entropy using the program PDB2ENTROPY (Fogolari et al., 2018). This program is based on the nearest-neighbor approach from probability distributions

of torsion angles at a given temperature relative to uniform distributions. The entropy is computed using the Maximum Information Spanning Tree (MIST) approach (*King and Tidor, 2009; King et al., 2012*).

Acknowledgements

We sincerely thank D E Shaw research for providing us with access to long simulation trajectories of monomeric alpha-synuclein in presence of fasudil which seeded the project. We acknowledge support of the Department of Atomic Energy, Government of India, under Project Identification No. RTI 4007. JM acknowledges Core Research grants provided by the Department of Science and Technology (DST) of India (CRG/2023/001426).

Additional information

Funding

Funder	Grant reference number	Author
Department of Atomic Energy, Government of India	RTI 4007	Jagannath Mondal
Department of Science and Technology, Ministry of Science and Technology, India	CRG/2023/001426	Jagannath Mondal

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

Sneha Menon, Data curation, Formal analysis, Validation, Investigation, Visualization, Methodology, Writing – original draft, Writing – review and editing; Subinoy Adhikari, Data curation, Software, Formal analysis, Validation, Investigation, Visualization, Methodology, Writing – original draft, Writing – review and editing; Jagannath Mondal, Supervision, Funding acquisition, Validation, Investigation, Writing – original draft, Project administration, Writing – review and editing

Author ORCIDs

Sneha Menon  <http://orcid.org/0000-0003-1079-8928>

Jagannath Mondal  <https://orcid.org/0000-0003-1090-5199>

Peer review material

Reviewer #2 (Public Review): <https://doi.org/10.7554/eLife.97709.3.sa1>

Reviewer #3 (Public Review): <https://doi.org/10.7554/eLife.97709.3.sa2>

Author response <https://doi.org/10.7554/eLife.97709.3.sa3>

Additional files

Supplementary files

- MDAR checklist

Data availability

The raw files of the data presented in this article and the key simulation trajectories have been uploaded to zenodo (<https://doi.org/10.5281/zenodo.14177307>) for public access. The implementations of Machine learning protocols are publicly available in our group GitHub webpage (https://github.com/JMLab-tifrh/Protein_Ligand_Variational_Autoencoder; copy archived at *JMLab-tifrh, 2024*).

The following dataset was generated:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Subinoy A, Sneha M, Jagannath M	2024	An Integrated Machine Learning Approach Delineates an Entropic Expansion Mechanism for the Binding of a Small Molecule to α -Synuclein	https://doi.org/10.5281/zenodo.14177307	Zenodo, 10.5281/zenodo.14177307

References

- Abadi M**, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, Corrado GS. 2016. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv*. <https://arxiv.org/abs/1603.04467> DOI: <https://doi.org/10.48550/arXiv.1603.04467>
- Adhikari S**, Mondal J. 2023. Machine learning subtle conformational change due to phosphorylation in intrinsically disordered proteins. *The Journal of Physical Chemistry B* **127**:9433–9449. DOI: <https://doi.org/10.1021/acs.jpcc.3c05136>, PMID: 37905972
- Ahmed MC**, Skaanning LK, Jussupow A, Newcombe EA, Kragelund BB, Camilloni C, Langkilde AE, Lindorff-Larsen K. 2021. Refinement of α -synuclein ensembles against SAXS data: comparison of force fields and methods. *Frontiers in Molecular Biosciences* **8**:654333. DOI: <https://doi.org/10.3389/fmolb.2021.654333>, PMID: 33968988
- Akoury E**, Gajda M, Pickhardt M, Biernat J, Soraya P, Griesinger C, Mandelkow E, Zweckstetter M. 2013. Inhibition of tau filament formation by conformational modulation. *Journal of the American Chemical Society* **135**:2853–2862. DOI: <https://doi.org/10.1021/ja312471h>, PMID: 23360400
- Alberti S**, Gladfelter A, Mittag T. 2019. Considerations and challenges in studying liquid-liquid phase separation and biomolecular condensates. *Cell* **176**:419–434. DOI: <https://doi.org/10.1016/j.cell.2018.12.035>, PMID: 30682370
- Allison JR**, Varnai P, Dobson CM, Vendruscolo M. 2009. Determination of the free energy landscape of alpha-synuclein using spin label nuclear magnetic resonance measurements. *Journal of the American Chemical Society* **131**:18314–18326. DOI: <https://doi.org/10.1021/ja904716h>, PMID: 20028147
- Attanasio F**, Convertino M, Magno A, Cafilisch A, Corazza A, Haridas H, Esposito G, Cataldo S, Pignataro B, Milardi D, Rizzarelli E. 2013. Carnosine inhibits A β (42) aggregation by perturbing the H-bond network in and around the central hydrophobic cluster. *Chembiochem* **14**:583–592. DOI: <https://doi.org/10.1002/cbic.201200704>, PMID: 23440928
- Babu MM**, van der Lee R, de Groot NS, Gsponer J. 2011. Intrinsically disordered proteins: regulation and disease. *Current Opinion in Structural Biology* **21**:432–440. DOI: <https://doi.org/10.1016/j.sbi.2011.03.011>, PMID: 21514144
- Bengio S**, Bengio Y. 2000. Taking on the curse of dimensionality in joint distributions using neural networks. *IEEE Transactions on Neural Networks* **11**:550–557. DOI: <https://doi.org/10.1109/72.846725>, PMID: 18249784
- Bernadó P**, Blanchard L, Timmins P, Marion D, Ruigrok RWH, Blackledge M. 2005. A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. *PNAS* **102**:17002–17007. DOI: <https://doi.org/10.1073/pnas.0506202102>, PMID: 16284250
- Bernadó P**, Mylonas E, Petoukhov MV, Blackledge M, Svergun DI. 2007. Structural characterization of flexible proteins using small-angle X-ray scattering. *Journal of the American Chemical Society* **129**:5656–5664. DOI: <https://doi.org/10.1021/ja069124n>, PMID: 17411046
- Bertoncini CW**, Jung YS, Fernandez CO, Hoyer W, Griesinger C, Jovin TM, Zweckstetter M. 2005. Release of long-range tertiary interactions potentiates aggregation of natively unstructured alpha-synuclein. *PNAS* **102**:1430–1435. DOI: <https://doi.org/10.1073/pnas.0407146102>, PMID: 15671169
- Best RB**, Zheng W, Mittal J. 2014. Balanced protein-water interactions improve properties of disordered proteins and non-specific protein association. *Journal of Chemical Theory and Computation* **10**:5113–5124. DOI: <https://doi.org/10.1021/ct500569b>, PMID: 25400522
- Biesaga M**, Frigolé-Vivas M, Salvatella X. 2021. Intrinsically disordered proteins and biomolecular condensates as drug targets. *Current Opinion in Chemical Biology* **62**:90–100. DOI: <https://doi.org/10.1016/j.cbpa.2021.02.009>, PMID: 33812316
- Bonomi M**, Camilloni C, Cavalli A, Vendruscolo M. 2016. Metainference: a bayesian inference method for heterogeneous systems. *Science Advances* **2**:e1501177. DOI: <https://doi.org/10.1126/sciadv.1501177>, PMID: 26844300
- Boomsma W**, Ferkinghoff-Borg J, Lindorff-Larsen K. 2014. Combining experiments and simulations using the maximum entropy principle. *PLOS Computational Biology* **10**:e1003406. DOI: <https://doi.org/10.1371/journal.pcbi.1003406>, PMID: 24586124
- Bottaro S**, Lindorff-Larsen K. 2018. Biophysical experiments and biomolecular simulations: A perfect match? *Science* **361**:355–360. DOI: <https://doi.org/10.1126/science.aat4010>, PMID: 30049874

- Bowman GR**, Beauchamp KA, Boxer G, Pande VS. 2009. Progress and challenges in the automated construction of Markov state models for full protein systems. *The Journal of Chemical Physics* **131**:124101. DOI: <https://doi.org/10.1063/1.3216567>, PMID: 19791846
- Bronowska A**. 2011. *Thermodynamics of Ligand-Protein Interactions: Implications for Molecular Design, Thermodynamics*. Interaction Studies - Solids, Liquids and Gases. DOI: <https://doi.org/10.5772/19447>
- Brookes DH**, Head-Gordon T. 2016. Experimental inferential structure determination of ensembles for intrinsically disordered proteins. *Journal of the American Chemical Society* **138**:4530–4538. DOI: <https://doi.org/10.1021/jacs.6b00351>, PMID: 26967199
- Bussi G**, Donadio D, Parrinello M. 2007. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics* **126**:014101. DOI: <https://doi.org/10.1063/1.2408420>, PMID: 17212484
- Cao K**, Zhu Y, Hou Z, Liu M, Yang Y, Hu H, Dai Y, Wang Y, Yuan S, Huang G. 2022. α -Synuclein as a Target for Metallo-Anti-Neurodegenerative Agents. *Angewandte Chemie International Edition*. DOI: <https://doi.org/10.1002/ange.202215360>
- Caro MA**, Laurila T, Lopez-Acevedo O. 2016. Accurate schemes for calculation of thermodynamic properties of liquid mixtures from molecular dynamics simulations. *The Journal of Chemical Physics* **145**:244504. DOI: <https://doi.org/10.1063/1.4973001>, PMID: 28049340
- Caro MA**, Lopez-Acevedo O, Laurila T. 2017. Redox potentials from ab initio molecular dynamics and explicit entropy calculations: application to transition metals in aqueous solution. *Journal of Chemical Theory and Computation* **13**:3432–3441. DOI: <https://doi.org/10.1021/acs.jctc.7b00314>, PMID: 28715635
- Cavalli A**, Camilloni C, Vendruscolo M. 2013. Molecular dynamics simulations with replica-averaged structural restraints generate structural ensembles according to the maximum entropy principle. *The Journal of Chemical Physics* **138**:094112. DOI: <https://doi.org/10.1063/1.4793625>, PMID: 23485282
- Chong PA**, Forman-Kay JD. 2016. Liquid-liquid phase separation in cellular signaling systems. *Current Opinion in Structural Biology* **41**:180–186. DOI: <https://doi.org/10.1016/j.sbi.2016.08.001>, PMID: 27552079
- Csizmok V**, Follis AV, Kriwacki RW, Forman-Kay JD. 2016. Dynamic protein interaction networks and new structural paradigms in signaling. *Chemical Reviews* **116**:6424–6462. DOI: <https://doi.org/10.1021/acs.chemrev.5b00548>, PMID: 26922996
- Das P**, Moll M, Stamati H, Kaviraki LE, Clementi C. 2006. Low-dimensional, free-energy landscapes of protein-folding reactions by nonlinear dimensionality reduction. *PNAS* **103**:9885–9890. DOI: <https://doi.org/10.1073/pnas.0603553103>
- Dedmon MM**, Lindorff-Larsen K, Christodoulou J, Vendruscolo M, Dobson CM. 2005. Mapping long-range interactions in alpha-synuclein using spin-label NMR and ensemble molecular dynamics simulations. *Journal of the American Chemical Society* **127**:476–477. DOI: <https://doi.org/10.1021/ja044834j>, PMID: 15643843
- Dyson HJ**, Wright PE. 2002. Coupling of folding and binding for unstructured proteins. *Current Opinion in Structural Biology* **12**:54–60. DOI: [https://doi.org/10.1016/s0959-440x\(02\)00289-0](https://doi.org/10.1016/s0959-440x(02)00289-0), PMID: 11839490
- Dyson HJ**. 2016. Making sense of intrinsically disordered proteins. *Biophysical Journal* **110**:1013–1016. DOI: <https://doi.org/10.1016/j.bpj.2016.01.030>, PMID: 26958875
- Ehrnhoefer DE**, Bieschke J, Boeddrich A, Herbst M, Masino L, Lurz R, Engemann S, Pastore A, Wanker EE. 2008. ECGC redirects amyloidogenic polypeptides into unstructured, off-pathway oligomers. *Nature Structural & Molecular Biology* **15**:558–566. DOI: <https://doi.org/10.1038/nsmb.1437>, PMID: 18511942
- Emin D**, Zhang YP, Lobanova E, Miller A, Li X, Xia Z, Dakin H, Sideris DI, Lam JYL, Ranasinghe RT, Kouli A, Zhao Y, De S, Knowles TPJ, Vendruscolo M, Ruggeri FS, Aigbirhio FI, Williams-Gray CH, Klenerman D. 2022. Small soluble α -synuclein aggregates are the toxic species in Parkinson's disease. *Nature Communications* **13**:5512. DOI: <https://doi.org/10.1038/s41467-022-33252-6>, PMID: 36127374
- Esteban-Martín S**, Silvestre-Ryan J, Bertoncini CW, Salvatella X. 2013. Identification of fibril-like tertiary contacts in soluble monomeric α -synuclein. *Biophysical Journal* **105**:1192–1198. DOI: <https://doi.org/10.1016/j.bpj.2013.07.044>, PMID: 24010662
- Faloutsos C**, Lin KI. 1995. FastMap: a fast algorithm for indexing, data-mining and visualization of traditional and multimedia datasets. Proceedings of the 1995 ACM SIGMOD International Conference on Management of Data. 163–174.
- Fields CR**, Bengoa-Vergniory N, Wade-Martins R. 2019. Targeting alpha-synuclein as a therapy for parkinson's disease. *Frontiers in Molecular Neuroscience* **12**:299. DOI: <https://doi.org/10.3389/fnmol.2019.00299>, PMID: 31866823
- Fisher RA**. 1936. The use of multiple measurements in taxonomic problems. *Annals of Eugenics* **7**:179–188. DOI: <https://doi.org/10.1111/j.1469-1809.1936.tb02137.x>
- Flock T**, Weatheritt RJ, Latysheva NS, Babu MM. 2014. Controlling entropy to tune the functions of intrinsically disordered regions. *Current Opinion in Structural Biology* **26**:62–72. DOI: <https://doi.org/10.1016/j.sbi.2014.05.007>, PMID: 24930020
- Fogolari F**, Maloku O, Dongmo Fomthum CJ, Corazza A, Esposito G. 2018. PDB2ENTROPY and PDB2TENT: conformational and translational-rotational entropy from molecular ensembles. *Journal of Chemical Information and Modeling* **58**:1319–1324. DOI: <https://doi.org/10.1021/acs.jcim.8b00143>, PMID: 29897235
- Follis AV**, Hammoudeh DI, Wang H, Prochownik EV, Metallo SJ. 2008. Structural rationale for the coupled binding and unfolding of the C-Myc oncoprotein by small molecules. *Chemistry & Biology* **15**:1149–1155. DOI: <https://doi.org/10.1016/j.chembiol.2008.09.011>
- Glorot X**, Bengio Y. 2010. Understanding the difficulty of training deep feedforward neural networks. Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics. 249–256.

- Gomes GNW**, Krzeminski M, Namini A, Martin EW, Mittag T, Head-Gordon T, Forman-Kay JD, Gradinaru CC. 2020. Conformational ensembles of an intrinsically disordered protein consistent with NMR, SAXS, and single-molecule FRET. *Journal of the American Chemical Society* **142**:15697–15710. DOI: <https://doi.org/10.1021/jacs.0c02088>, PMID: 32840111
- Heller GT**, Sormanni P, Vendruscolo M. 2015. Targeting disordered proteins with small molecules using entropy. *Trends in Biochemical Sciences* **40**:491–496. DOI: <https://doi.org/10.1016/j.tibs.2015.07.004>, PMID: 26275458
- Heller GT**, Bonomi M, Vendruscolo M. 2018. Structural ensemble modulation upon small-molecule binding to disordered proteins. *Journal of Molecular Biology* **430**:2288–2292. DOI: <https://doi.org/10.1016/j.jmb.2018.03.015>, PMID: 29596915
- Heller GT**, Aprile FA, Michaels TCT, Limbocker R, Perni M, Ruggeri FS, Mannini B, Löhner T, Bonomi M, Camilloni C, De Simone A, Felli IC, Pierattelli R, Knowles TPJ, Dobson CM, Vendruscolo M. 2020. Small-molecule sequestration of amyloid- β as a drug discovery strategy for Alzheimer's disease. *Science Advances* **6**:eabb5924. DOI: <https://doi.org/10.1126/sciadv.abb5924>, PMID: 33148639
- Hess B**, Bekker H, Berendsen HJC, Fraaije J. 1997. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry* **18**:1463–1472. DOI: [https://doi.org/10.1002/\(SICI\)1096-987X\(199709\)18:12<1463::AID-JCC4>3.0.CO;2-H](https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H)
- Higgins I**, Matthey L, Pal A, Burgess C, Glorot X, Botvinick M, Mohamed S, Lerchner A. 2016. beta-vae: learning basic visual concepts with a constrained variational framework. International Conference on Learning Representations. .
- Hinton GE**, Salakhutdinov RR. 2006. Reducing the dimensionality of data with neural networks. *Science* **313**:504–507. DOI: <https://doi.org/10.1126/science.1127647>, PMID: 16873662
- Hore A**, Ziou D. 2010. Image quality metrics: PSNR vs. SSIM. 2010 20th International Conference on Pattern Recognition. 2366–2369.
- Huang SN**, Pascal TA, Goddard WA III, Maiti PK, Lin ST. 2011. Absolute entropy and energy of carbon dioxide using the two-phase thermodynamic model. *Journal of Chemical Theory and Computation* **7**:1893–1901. DOI: <https://doi.org/10.1021/ct200211b>, PMID: 26596450
- Huang J**, Rauscher S, Nawrocki G, Ran T, Feig M, de Groot BL, Grubmüller H, MacKerell AD Jr. 2017. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature Methods* **14**:71–73. DOI: <https://doi.org/10.1038/nmeth.4067>, PMID: 27819658
- Husic BE**, Pande VS. 2018. Markov state models: from an art to a science. *Journal of the American Chemical Society* **140**:2386–2396. DOI: <https://doi.org/10.1021/jacs.7b12191>, PMID: 29323881
- Hyvärinen A**. 2013. Independent component analysis: recent advances. *Philosophical Transactions of the Royal Society A* **371**:20110534. DOI: <https://doi.org/10.1098/rsta.2011.0534>
- JMLab-tifrh**. 2024. Protein_Ligand_Variational_Autoencoder. swh:1:rev:e6f7c4452d8afc12ff309cdc44adcbff1a4bcc79. Software Heritage. https://archive.softwareheritage.org/swh:1:dir:43a88fcd4f57adaee81df77be3e4f973fdd353cc;origin=https://github.com/JMLab-tifrh/Protein_Ligand_Variational_Autoencoder;visit=swh:1:snp:fc6d8246cc6281fdada9c2e711293033f079b06d;anchor=swh:1:rev:e6f7c4452d8afc12ff309cdc44adcbff1a4bcc79
- King BM**, Tidor B. 2009. MIST: maximum information spanning trees for dimension reduction of biological data sets. *Bioinformatics* **25**:1165–1172. DOI: <https://doi.org/10.1093/bioinformatics/btp109>, PMID: 19261718
- King BM**, Silver NW, Tidor B. 2012. Efficient calculation of molecular configurational entropies using an information theoretic approximation. *The Journal of Physical Chemistry. B* **116**:2891–2904. DOI: <https://doi.org/10.1021/jp2068123>, PMID: 22229789
- Kingma DP**, Welling M. 2013. Auto-Encoding Variational Bayes. *arXiv*. <https://arxiv.org/abs/1312.6114>
- Kingma DP**, Ba J. 2014. Adam: A Method for Stochastic Optimization. *arXiv*. <https://arxiv.org/abs/1412.6980>
- Klebe G**. 2015. Applying thermodynamic profiling in lead finding and optimization. *Nature Reviews. Drug Discovery* **14**:95–110. DOI: <https://doi.org/10.1038/nrd4486>, PMID: 25614222
- Klema V**, Laub A. 1980. The singular value decomposition: Its computation and some applications. *IEEE Transactions on Automatic Control* **25**:164–176. DOI: <https://doi.org/10.1109/TAC.1980.1102314>
- Krishnan N**, Koveal D, Miller DH, Xue B, Akshinthala SD, Kragelj J, Jensen MR, Gauss C-M, Page R, Blackledge M, Muthuswamy SK, Peti W, Tonks NK. 2014. Targeting the disordered C terminus of PTP1B with an allosteric inhibitor. *Nature Chemical Biology* **10**:558–566. DOI: <https://doi.org/10.1038/nchembio.1528>, PMID: 24845231
- Kruskal JB**, Wish M. 1978. *Multidimensional Scaling*. Sage. DOI: <https://doi.org/10.4135/9781412985130>
- Kurzbach D**, Schwarz TC, Platzer G, Höfler S, Hinderberger D, Konrat R. 2014. Compensatory adaptations of structural dynamics in an intrinsically disordered protein complex. *Angewandte Chemie* **53**:3840–3843. DOI: <https://doi.org/10.1002/anie.201308389>, PMID: 24604825
- Larochelle H**, Murray I. 2011. The neural autoregressive distribution estimator. Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics. 29–37.
- Lin ST**, Blanco M, Goddard WA III. 2003. The two-phase model for calculating thermodynamic properties of liquids from molecular dynamics: Validation for the phase diagram of Lennard-Jones fluids. *The Journal of Chemical Physics* **119**:11792–11805. DOI: <https://doi.org/10.1063/1.1624057>
- Lin ST**, Maiti PK, Goddard WA III. 2010. Two-phase thermodynamic model for efficient and accurate absolute entropy of water from molecular dynamics simulations. *The Journal of Physical Chemistry. B* **114**:8191–8198. DOI: <https://doi.org/10.1021/jp103120q>, PMID: 20504009
- Lincoff J**, Haghighatlari M, Krzeminski M, Teixeira JMC, Gomes GNW, Gradinaru CC, Forman-Kay JD, Head-Gordon T. 2020. Extended experimental inferential structure determination method in determining the

- structural ensembles of disordered protein states. *Communications Chemistry* **3**:74. DOI: <https://doi.org/10.1038/s42004-020-0323-0>, PMID: 32775701
- Lindahl E, Abraham M, Hess B, Spoel D. 2020. GROMACS 2020.1 source code. 2020.1. Zenodo. <https://zenodo.org/records/3685919>
- Löhr T, Kohlhoff K, Heller GT, Camilloni C, Vendruscolo M. 2022. A small molecule stabilizes the disordered native state of the alzheimer's A β peptide. *ACS Chemical Neuroscience* **13**:1738–1745. DOI: <https://doi.org/10.1021/acscchemneuro.2c00116>, PMID: 35649268
- Marsh JA, Baker JMR, Tollinger M, Forman-Kay JD. 2008. Calculation of residual dipolar couplings from disordered state ensembles using local alignment. *Journal of the American Chemical Society* **130**:7804–7805. DOI: <https://doi.org/10.1021/ja802220c>, PMID: 18512919
- Metallo SJ. 2010. Intrinsically disordered proteins are potential drug targets. *Current Opinion in Chemical Biology* **14**:481–488. DOI: <https://doi.org/10.1016/j.cbpa.2010.06.169>, PMID: 20598937
- Metzner P, Schütte C, Vanden-Eijnden E. 2009. Transition path theory for markov jump processes. *Multiscale Modeling & Simulation* **7**:1192–1219. DOI: <https://doi.org/10.1137/070699500>
- Mukherjee S, Schäfer LV. 2023. Thermodynamic forces from protein and water govern condensate formation of an intrinsically disordered protein domain. *Nature Communications* **14**:5892. DOI: <https://doi.org/10.1038/s41467-023-41586-y>, PMID: 37735186
- Noé F, Schütte C, Vanden-Eijnden E, Reich L, Weikl TR. 2009. Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations. *PNAS* **106**:19011–19016. DOI: <https://doi.org/10.1073/pnas.0905466106>
- O'Flynn BG, Mittag T. 2021. The role of liquid-liquid phase separation in regulating enzyme activity. *Current Opinion in Cell Biology* **69**:70–79. DOI: <https://doi.org/10.1016/j.ceb.2020.12.012>, PMID: 33503539
- Ono K, Li L, Takamura Y, Yoshiike Y, Zhu L, Han F, Mao X, Ikeda T, Takasaki J, Nishijo H, Takashima A, Teplow DB, Zagorski MG, Yamada M. 2012. Phenolic compounds prevent amyloid β -protein oligomerization and synaptic dysfunction by site-specific binding. *Journal of Biological Chemistry* **287**:14631–14643. DOI: <https://doi.org/10.1074/jbc.M111.325456>
- Páll S, Hess B. 2013. A flexible algorithm for calculating pair interactions on SIMD architectures. *Computer Physics Communications* **184**:2641–2650. DOI: <https://doi.org/10.1016/j.cpc.2013.06.003>
- Parrinello M, Rahman A. 1981. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics* **52**:7182–7190. DOI: <https://doi.org/10.1063/1.328693>
- Prinz JH, Wu H, Sarich M, Keller B, Senne M, Held M, Chodera JD, Schütte C, Noé F. 2011. Markov models of molecular kinetics: generation and validation. *The Journal of Chemical Physics* **134**:174105. DOI: <https://doi.org/10.1063/1.3565032>, PMID: 21548671
- Rezende DJ, Mohamed S, Wierstra D. 2014. Stochastic backpropagation and approximate inference in deep generative models. International Conference on Machine Learning. 1278–1286.
- Robustelli P, Piana S, Shaw DE. 2018. Developing a molecular dynamics force field for both folded and disordered protein states. *PNAS* **115**:E4758–E4766. DOI: <https://doi.org/10.1073/pnas.1800690115>
- Robustelli P, Piana S, Shaw DE. 2020. Mechanism of coupled folding-upon-binding of an intrinsically disordered protein. *Journal of the American Chemical Society* **142**:11092–11101. DOI: <https://doi.org/10.1021/jacs.0c03217>, PMID: 32323533
- Robustelli P, Ibanez-de-Opakua A, Campbell-Bezat C, Giordanetto F, Becker S, Zweckstetter M, Pan AC, Shaw DE. 2022. Molecular basis of small-molecule binding to α -synuclein. *Journal of the American Chemical Society* **144**:2501–2510. DOI: <https://doi.org/10.1021/jacs.1c07591>, PMID: 35130691
- Roux B, Weare J. 2013. On the statistical equivalence of restrained-ensemble simulations with the maximum entropy method. *The Journal of Chemical Physics* **138**:084107. DOI: <https://doi.org/10.1063/1.4792208>, PMID: 23464140
- Roweis ST, Saul LK. 2000. Nonlinear dimensionality reduction by locally linear embedding. *Science* **290**:2323–2326. DOI: <https://doi.org/10.1126/science.290.5500.2323>, PMID: 11125150
- Scherer MK, Trendelkamp-Schroer B, Paul F, Pérez-Hernández G, Hoffmann M, Plattner N, Wehmeyer C, Prinz JH, Noé F. 2015. PyEMMA 2: a software package for estimation, validation, and analysis of markov models. *Journal of Chemical Theory and Computation* **11**:5525–5542. DOI: <https://doi.org/10.1021/acs.jctc.5b00743>, PMID: 26574340
- Schölkopf B, Smola A, Müller KR. 2005. Kernel principal component analysis. Artificial Neural Networks-ICANN 97: 7th International Conference Lausanne. 583–588.
- Schwantes CR, Pande VS. 2015. Modeling molecular kinetics with tICA and the kernel trick. *Journal of Chemical Theory and Computation* **11**:600–608. DOI: <https://doi.org/10.1021/ct5007357>, PMID: 26528090
- Shin Y, Brangwynne CP. 2017. Liquid phase condensation in cell physiology and disease. *Science* **357**:eaaf4382. DOI: <https://doi.org/10.1126/science.aaf4382>, PMID: 28935776
- Stelzl LS, Pietrek LM, Holla A, Oroz J, Sikora M, Köfinger J, Schuler B, Zweckstetter M, Hummer G. 2022. Global structure of the intrinsically disordered protein tau emerges from its local structure. *JACS Au* **2**:673–686. DOI: <https://doi.org/10.1021/jacsau.1c00536>, PMID: 35373198
- Stephens AD, Kölbl J, Moons R, Chung CW, Ruggiero MT, Mahmoudi N, Shmool TA, McCoy TM, Nietlispach D, Routh AF, Sobott F, Zeitler JA, Kaminski Schierle GS. 2023. Decreased water mobility contributes to increased α -synuclein aggregation. *Angewandte Chemie* **135**:e202212063. DOI: <https://doi.org/10.1002/ange.202212063>, PMID: 38516046
- Sumithra V, Surendran S. 2015. A review of various linear and non linear dimensionality reduction techniques. *International Journal of Computer Science and Information Technologies* **6**:2354–2360.

- Tatenhorst L**, Eckermann K, Dambeck V, Fonseca-Ornelas L, Walle H, Lopes da Fonseca T, Koch JC, Becker S, Tönges L, Bähr M, Outeiro TF, Zweckstetter M, Lingor P. 2016. Fasudil attenuates aggregation of α -synuclein in models of Parkinson's disease. *Acta Neuropathologica Communications* **4**:39. DOI: <https://doi.org/10.1186/s40478-016-0310-y>, PMID: 27101974
- Tenenbaum JB**, de Silva V, Langford JC. 2000. A global geometric framework for nonlinear dimensionality reduction. *Science* **290**:2319–2323. DOI: <https://doi.org/10.1126/science.290.5500.2319>, PMID: 11125149
- Tóth G**, Gardai SJ, Zago W, Bertoncini CW, Cremades N, Roy SL, Tambe MA, Rochet J-C, Galvagnion C, Skibinski G, Finkbeiner S, Bova M, Regnstrom K, Chiou S-S, Johnston J, Callaway K, Anderson JP, Jobling MF, Buell AK, Yednock TA, et al. 2014. Targeting the intrinsically disordered structural ensemble of α -synuclein by small molecules as a potential therapeutic strategy for parkinson's disease. *PLOS ONE* **9**:e87133. DOI: <https://doi.org/10.1371/journal.pone.0087133>
- Ubbiali D**, Fratini M, Piersimoni L, Ihling CH, Kipping M, Heilmann I, Iacobucci C, Sinz A. 2022. Direct observation of “elongated” conformational states in α -synuclein upon liquid-liquid phase separation. *Angewandte Chemie* **61**:e202205726. DOI: <https://doi.org/10.1002/anie.202205726>, PMID: 36115020
- Ullman O**, Fisher CK, Stultz CM. 2011. Explaining the structural plasticity of α -synuclein. *Journal of the American Chemical Society* **133**:19536–19546. DOI: <https://doi.org/10.1021/ja208657z>, PMID: 22029383
- Uversky VN**, Oldfield CJ, Dunker AK. 2008. Intrinsically disordered proteins in human diseases: introducing the D2 concept. *Annual Review of Biophysics* **37**:215–246. DOI: <https://doi.org/10.1146/annurev.biophys.37.032807.125924>, PMID: 18573080
- Wang Z**, Bovik AC, Sheikh HR, Simoncelli EP. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing* **13**:600–612. DOI: <https://doi.org/10.1109/TIP.2003.819861>
- Wang Y**, Fisher JC, Mathew R, Ou L, Otieno S, Sublet J, Xiao L, Chen J, Roussel MF, Kriwacki RW. 2011. Intrinsic disorder mediates the diverse regulatory functions of the Cdk inhibitor P21. *Nature Chemical Biology* **7**:214–221. DOI: <https://doi.org/10.1038/nchembio.536>, PMID: 21358637
- Wehmeyer C**, Noé F. 2018. Time-lagged autoencoders: Deep learning of slow collective variables for molecular kinetics. *The Journal of Chemical Physics* **148**:241703. DOI: <https://doi.org/10.1063/1.5011399>, PMID: 29960344
- Weinan E**, Vanden-Eijnden E. 2006. Towards a theory of transition paths. *Journal of Statistical Physics* **123**:503–523. DOI: <https://doi.org/10.1007/s10955-005-9003-9>
- Winner B**, Jappelli R, Maji SK, Desplats PA, Boyer L, Aigner S, Hetzer C, Loher T, Vilar M, Campioni S, Tzitzilonis C, Soragni A, Jessberger S, Mira H, Consiglio A, Pham E, Masliah E, Gage FH, Riek R. 2011. In vivo demonstration that alpha-synuclein oligomers are toxic. *PNAS* **108**:4194–4199. DOI: <https://doi.org/10.1073/pnas.1100976108>, PMID: 21325059
- Wright PE**, Dyson HJ. 2015. Intrinsically disordered proteins in cellular signalling and regulation. *Nature Reviews Molecular Cell Biology* **16**:18–29. DOI: <https://doi.org/10.1038/nrm3920>
- York DM**, Darden TA, Pedersen LG. 1993. The effect of long-range electrostatic interactions in simulations of macromolecular crystals: a comparison of the ewald and truncated list methods. *The Journal of Chemical Physics* **99**:8345–8348. DOI: <https://doi.org/10.1063/1.465608>
- Zhu M**, De Simone A, Schenk D, Toth G, Dobson CM, Vendruscolo M. 2013. Identification of small-molecule binding pockets in the soluble monomeric form of the A β 42 peptide. *The Journal of Chemical Physics* **139**:035101. DOI: <https://doi.org/10.1063/1.4811831>, PMID: 23883055