Leveraging Deep Learning and MD Simulations to Decipher the Molecular Basis of Attenuated Activity in Glycocin F

Dheeraj Kumar Sarkar,[*‡1] Subinoy Adhikari,[‡1] Avadhesha Surolia[*2,3]

[1]*Tata Institute of Fundamental Research, Hyderabad, Telangana, India-500046*

[2]*Molecular Biophysics Unit, Indian Institute of Science, Bangalore, India*

[3]*Dr. Reddy's Institute of Life Sciences, University of Hyderabad Campus, Hyderabad 500046, India*

‡ These authors contributed equally to the paper

*Corresponding author: dksarkar@tifrh.res.in, surolia@iisc.ac.in

Abstract

The escalating crisis of multi-drug resistant bacteria demand a new generation of antibiotics. GccF, a potent bacteriocin, is a promising candidate, but its function hinges on unique post-translational glycosylation. Intriguingly, a seemingly minor chemical tweak of α-methylation at Ser18 of GccF destroys its activity, reducing potency by 1000-fold. To quantify how this subtle chemical change leads to profound functional compromise, we used an advanced molecular dynamics framework guided by Variational Autoencoder to unravel GccF's complex dynamics. Our findings reveal that native glycosylation provides crucial rigidity, locking the peptide into a stable, functional shape. In stark contrast, α-methylation introduces unintended flexibility, compromising the peptide's ability to adopt its active conformation, indicating that the precise positioning and presentation of the glycan moiety are critical for GccF's function. This work provides a critical blueprint for the rational design of next generation antibiotics, demonstrating how precise chemical modifications can dictate a peptide's function by profoundly altering its structural dynamics.

53

54 **1 Introduction**

55 The emergence of multi-drug resistance bacteria constitute a major challenge to global health.
56 One such example is vancomycin-resistant *Enterococcus* species (VRE), which have rendered
57 many traditional antibiotics ineffective [1-2]. Thereby motivating exploration of new
58 antimicrobials agents with novel mechanisms of action, among those bacteriocins, ribosomally
59 synthesized and post-translationally modified peptides (RiPPs) hold a considerable promise [3-4].

60 Secreted by *Lactobacillus plantarum* KW30, GccF stands out as a particularly compelling
61 candidate within this class. GccF is a 43-residue bacteriocin that demonstrates potent and
62 reversible bacteriostatic activity against a range of Gram-positive bacteria at remarkably low
63 picomolar concentrations [5-6]. Beyond its disulfide connectivity (Cys-5 to Cys-28 and Cys-12 to
64 Cys-21), GccF is distinguished by its post-translational modifications (PTMs), specifically its
65 diglycosylated state which is obligatoty for the function of the protein. The first is a β-O-linked
66 GlcNAc at Ser18, which is situated on the interhelical loop of the peptide [5-7]. The second, and
67 more unusual, modification is a β-S-linked GlcNAc at the C-terminal Cys43 [5-6]. The removal of
68 this single sugar moiety completely abolishes the peptide's function, suggesting a primary and
69 indispensable role [6,8].

70 As a general post-translational modification, glycosylation is well-known to confer enhanced
71 biophysical stability to peptides and proteins. This includes a significant increase in both thermal
72 stability and resistance to proteolytic degradation, which are critical properties for the viability
73 of any therapeutic peptide [9-11b]. A recent study revealed an instructive counterexample to its
74 general utility by application of α-methylation to GccF at Ser18. While the modification is typically
75 used to rigidify peptides and increase their stability, the study revealed substitution of the Cα
76 proton with a methyl group at Ser18 resulted in a catastrophic loss of antimicrobial activity,
77 reducing its potency by an astonishing 1000-fold [6] and hence showing a marked difference in
78 the orientation and mobility of the critical interhelical loop.

79 Since, the study by Harjes *et al.*, 2023 [6], argued the structural basis of this functional loss by
80 GccF by a subtle chemical change, we here used a framework to investigate beyond the static

structural comparison in order to quantify how this modifications leads to a profound functional loss. We analyzed three molecular systems: control, glycosylated and methylated to describe a non-glycosylated control, the native glycosylated peptide (Ser18-GlcNAc, Cys43-GlcNAc), and an α-methyl-Ser18 glycosylated construct respectively. All-atom molecular dynamics (MD) simulations are the ideal starting point, as they provide atomic-level trajectories of biomolecular systems over time [12]. In order to interpret the high-dimensional trajectory data from unbiased MD of aggregated 3 μs data for each molecular systems, we employed a powerful integrative pipeline that leverages machine learning with Variational Autoencoder (VAE) for dimensionality reduction [13-15]. A key innovation in this framework is the use of a tailored β-annealing schedule for VAE training as standard VAE often struggles with "posterior collapse," where the model fails to learn a meaningful latent representation [16]. By dynamically adjusting the β factor, we reconstructed a more robust latent representation of our projected trajectory data [14]. The refined latent space from the VAE is then used as the basis for building Markov State Models (MSMs). MSMs are a statistical mechanics tool for extracting kinetic information from MD trajectories [17-18]. The combination of all-atom MD combined with β-annealing VAEs guided by MSMs provided an ensemble-level view of GccF's conformational dynamics, allowing for a precise quantification of how glycosylation and α-methylation modulating the dynamics and stability of the critical interhelical loop in GccF. We observed α-methylation may have locally rigidified the Ser18 residue, this local constraint appears to have propagated a global conformational effects. This in turn, compromised the peptide's ability to adopt the native, functional ensemble required for its activity, suggesting that the precise positioning and presentation of the glycan moiety is critical for its function. The results will not only elucidate the precise mechanism of action for this unique antimicrobial peptide but will also provide broader insights into the sophisticated role of PTMs in dictating peptide function and stability, the knowledge of which will pave the rational design of future generation of antibiotics.

2 Methods

2.1 Molecular Dynamics Simulations

Molecular dynamics (MD) simulations were conducted using GROMACS version 2021.6 [19] with the CHARMM36 force field [20-21]. Three variants of GccF were modeled: (i) a *control* system

without glycosylation, where both GlcNAc moieties were removed; (ii) a *glycosylated* system based on the native NMR structure (PDB ID: 2KUY) [22]; and (iii) a *methylated* system prepared using the NMR structure (PDB ID: 8DFZ) [6]. Bonded and non-bonded parameters for the α-Methylated Ser18 linked GlcNAc residue. were calculated using the DFT Seminario method [23] within the Gaussian 16 software package [24] to derive force constants, which were used for all-atom MD simulations. Disulfide bonds between Cys4–Cys39 and Cys9–Cys43 were enforced in all systems.

Each system was solvated in a cubic box of TIP3P water [25-26], extending at least 1.0 nm beyond the peptide surface, and neutralized with Na^+/Cl^- ions to achieve a physiological ionic strength of 150 mM. Systems were energy-minimized using the steepest-descent algorithm until the maximum force was below 100 kJ mol^{-1} nm^{-1}. Equilibration was performed in two stages: (i) 2 ns of NVT equilibration using the V-rescale thermostat [27] with position restraints on heavy atoms, and (ii) 2 ns of NPT equilibration using the Parrinello–Rahman barostat [28]. Production MD simulations were conducted in the NPT ensemble with coordinates saved every 100 ps. All bonds were constrained using the LINCS algorithm [29], and long-range electrostatics were treated with the Particle Mesh Ewald (PME) method [30] using a 1.2 nm cutoff for both Coulomb and van der Waals interactions. Each system was simulated in triplicate to ensure reproducibility.

2.2 Variational Autoencoder and Free-energy Surface Construction

Molecular dynamics trajectories for the *control, glycosylated,* and *methylated* systems were processed to remove solvent molecules and ions, and periodic boundary conditions were corrected, to extract features for dimensionality reduction. For each trajectory, the protein backbone dihedral angles, specifically the Φ and Ψ torsion angles, were computed. To properly represent the circular nature of this angular data for the neural network, the cosine and sine of each angle were calculated, resulting in four feature components for each residue. These components were combined into a single feature vector for each frame, and the complete dataset for each system was subsequently partitioned into training and test sets using a 90:10 split. A β-Variational Autoencoder (VAE) [14] was employed to construct a non-linear, low-dimensional representation of the dihedral angle space. The encoder network consisted of three

138 fully connected layers with 168, 64, and 16 neurons, mapping into a two-dimensional latent space.
139 The decoder network was symmetrically structured, with layers of 16, 64, and 168 neurons.
140 Training was performed for 1000 epochs using Stochastic Gradient Descent with a learning rate
141 of 0.001, with batch sizes of 2048 for training and 2000 for validation. The total loss function was
142 defined as

143 $$\mathcal{L}_{VAE} = \alpha \cdot \mathcal{L}_{reconstruction} + \beta \cdot \mathcal{L}_{KL} \quad (1)$$

144 where the reconstruction loss was given by the mean squared error (MSE), $\mathcal{L}_{KL}$ denotes the
145 Kullback–Leibler divergence, α was fixed at 10, and β was annealed during training.

146 The β parameter was initially annealed from 0 to 1 during training using linear, cosine, and logistic
147 schedules. For each system, the optimal β value was determined as the one that maximized the
148 fraction of variance explained (FVE), following the procedure established in prior work. The VAEs
149 were subsequently retrained by annealing β to these system-specific optimal values. Model
150 selection among the annealing schemes was carried out using 5-fold cross-validated VAMP-2
151 scores (SI Figure S5). Among the tested approaches, cosine annealing VAE consistently yielded
152 the highest scores, surpassing both the alternative annealing schedules and the leading two
153 dimensions obtained via Principal Component Analysis (PCA). We investigated the free energy
154 surface (FES) profile by plotting the first two principal components from PCA and projecting the
155 trajectory data using the dihedral matrix into a lower dimensional space using our VAE approach.
156 This indicated the efficiency of VAE over traditional PCA in capturing non-linear, collective
157 variables that are crucial for describing complex protein dynamics (SI Figure S4 & S11).

158 **2.3 Markov State Model (MSM) Analysis and Validation**

159 We tested the convergence of all simulated MD trajectories of 3 µs each for *control*, *glycosylated*
160 and *methylated* systems using time-dependent behavior and auto-correlation functions analysis
161 [31]. We analyzed the change over time and the autocorrelation of two key metrics: the *Cα end-*
162 *to-end distance* (the distance between the Cα atoms of the first and last residues) and the *Cα*
163 *RMSD* (the root-mean-square deviation of each simulation frame from the initial frame). A
164 convergence within 0.5 µs can be observed for all the three replicates (SI Figure S1) generated

165 for the three molecular systems. Consequently, the cosine-annealed VAE, with optimal β values
166 of 0.085 *(control)*, 0.135 *(glycosylated)*, and 0.024 *(methylated)* were chosen to generate the two-
167 dimensional latent spaces used for the construction of the MSMs. All MSMs were generated
168 using PyEMMA package [32]. The slow processes evolved from the MD trajectories were analyzed
169 using implied timescale (ITS) plot (SI Figure S7) and the number of macrostates were chosen
170 based on spectral analysis (SI Figure S8).

171 **2.3.1 Analysis and Validation**

172 The constructed MSMs were validated using Chapman-Kolmogorov [33] test (SI Figure 8-10)
173 against estimated models until lag time of 100ps. All trajectories were analyzed using MDAnalysis
174 [34] and custom Python scripts. Backbone root-mean-square deviation (RMSD) and per-residue
175 root-mean-square fluctuation (RMSF) were calculated relative to the initial conformations.
176 Secondary structure content was determined using the DSSP algorithm [34]. The S^2 order
177 parameters were computed using the cpptraj [36] module (version 6.18.1) from the AMBER
178 software suite [37], and the backbone NH bond vectors were used to determine the S^2 values
179 after removing the global rotational and translation motion using GROMACS v2021.6 [19].

180 **3 Results**

181 Using our MD-VAE-MSM framework, we studied the role of glycosylation in wild-type GccF
182 *(glycosylated)*, modified Ser18 GccF *(methylated)* along with *control* dataset with no glycosylation.
183 We simulated each molecular systems (Figure 1A) for an aggregated ~3 µs consisting three
184 replicate runs to study the distinct impacts of glycosylation and α-methylation on GccF's
185 conformational landscape and functional relationship. β-annealed VAE (Figure 1B) was able to
186 explore a wider conformational minima for the three molecular systems (Figure S4 & 11).

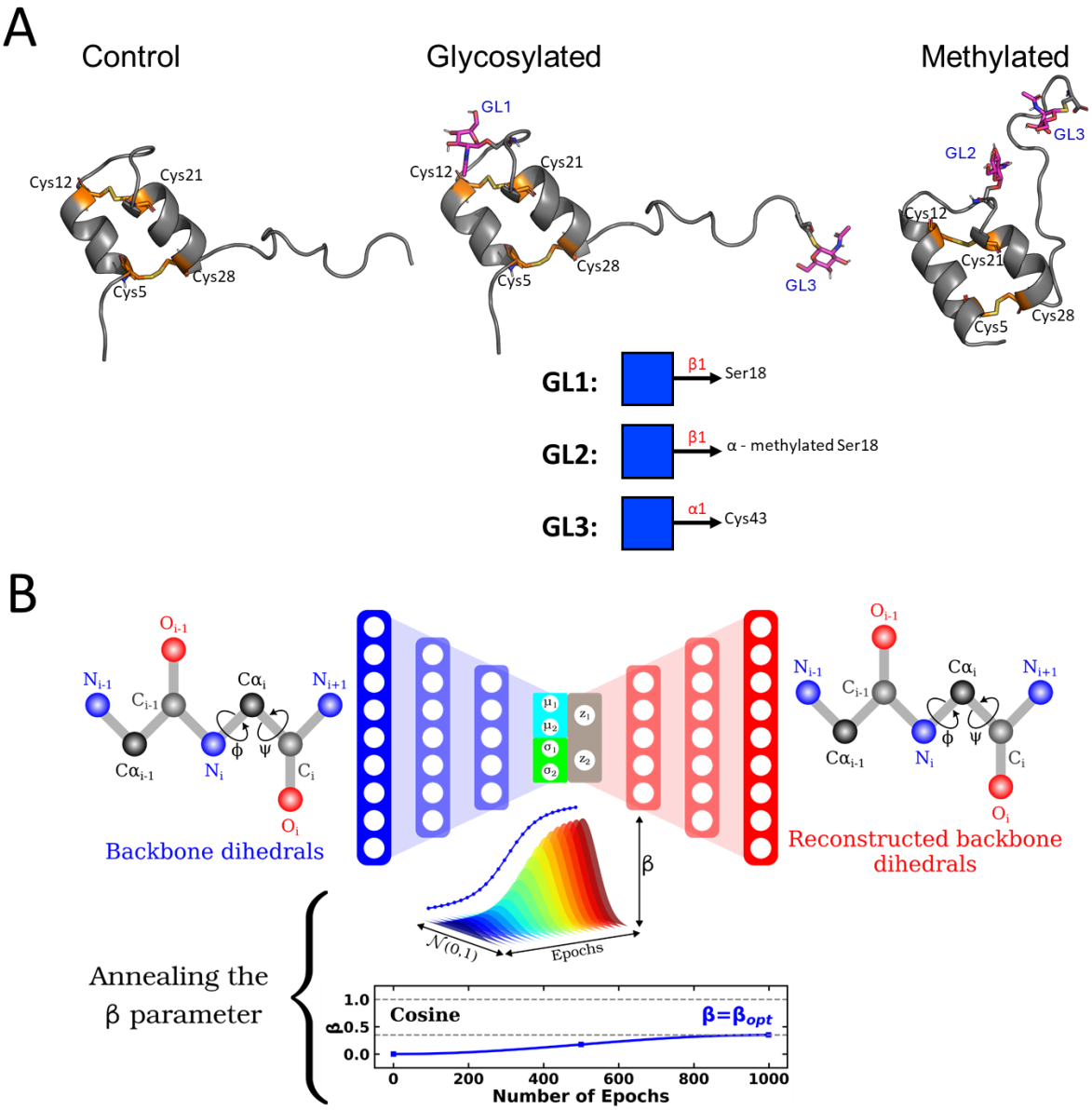


Figure 1. A. Structural representation of Glycocin F (GccF) showing the two glycosylation sites at Ser18 (β-O-GlcNAc) and Cys43 (β-S-GlcNAc), stabilized by nested disulfide bonds (Cys4–Cys39 and Cys9–Cys43). B. Schematic of β-annealed Variational AutoEncoder. A schematic of the reconstructed dihedrals are shown along with the β optimal (β_{opt}) annealed using cosine method.

3.1 Stability and flexibility profiles: glycosylation and α-methylation stabilizes key loop dynamics

Backbone RMSD profiles revealed that α-methylation promoted overall conformational stability to GccF structure followed by glycosylation with >2.0 Å on average compared to ~2.5 Å in the control (Figure 1A). RMSF analysis pinpointed reduced fluctuations and prominent rigidity in the

interhelical loop (residues 10–19) upon O-glycosylation in wild-type GccF and modified GccF, whereas α-methylation at Ser18 introduced enhanced rigidity in this region (Figure 2B & 2D). These findings align with previous reports linking O-GlcNAc modification to enhanced peptide rigidity and resistance to unfolding [6]. MSM results revealed the transition time from closed to opened state is much higher for the glycosylated and methylated models (Figure 2C) compared to control dynamics depicting much improved and effective role of both O-/S-linked GlcNAc and α-methylation for the stabilization of the scaffold, with the largest effects concentrated in loop residues that are poised to modulate glycan presentation. S^2 order analysis (Figure 1D) revealed Ser18 is the most responsive region: glycosylation dampens fluctuations across this segment and the C-terminus. The α-methyl–Ser18 variant further rigidifies the Ser18-proximal loop.

Secondary structure analyses (SI Figure S2) confirmed that glycosylation preserved helical segments while dampening loop disorder, in contrast to the methylated system which exhibited partial loss of α-helical content. Hydrogen-bond analyses (SI Figure S3) revealed persistent glycan–peptide contacts in the glycosylated system, reinforcing loop rigidity and stabilizing the disulfide-anchored scaffold. In contrast, α-methylation disrupted key H-bonding patterns and reoriented the GlcNAc moiety, reducing the stabilizing effect of glycosylation.

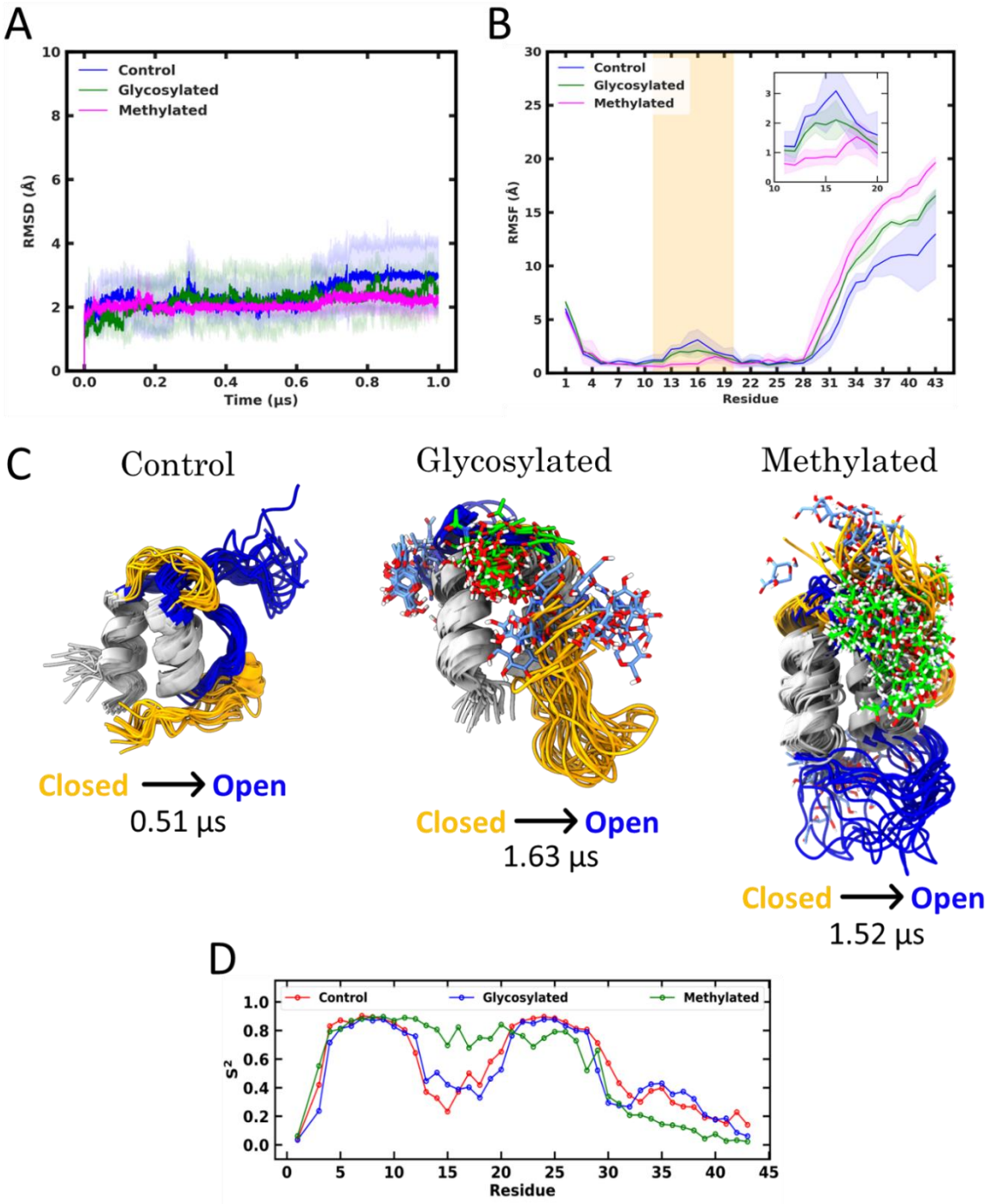


Figure 2. A. Backbone RMSD and B. per-residue RMSF comparisons among *control, glycosylated,* and *methylated* systems. C. Transition time of closed to opened metastable states obtained from MSM analysis for the three molecular systems. D. S^2 order analysis for *control, glycosylated,* and *methylated* systems.

3.2 Kinetic decomposition and metastable states in the glycosylated and α-methylated GccF compared to control

Mapping backbone dihedrals into ϕ–ψ space of Ramachandran plot reveals a clear, modification-dependent topology (Figure 3). The deglycosylated *control* samples a most residues in favored α/β regions and generated more diffuse set of basins (Figure 3A). Glycosylation concentrates population into fewer, deeper basins, indicating a narrowed conformational ensemble, as glycan–protein contacts stabilize turns/β-like conformations (Figure 3B). In contrast, the α-methyl–Ser18 variant shows a broader yet stabilized landscape, with density centered on loop-rigidified conformers. α-methylation at Ser18 adds local steric bulk and subtly shift ϕ–ψ preferences, with a mild redistribution toward the β/PPII sector (ψ ≈ 120–150°) in exposed segment (Figure 3C).

FES derived from β-annealed VAE latent coordinates revealed that glycosylation compressed the conformational ensemble into well-defined states with higher residence times (Figure 4A & SI Figure 11B), while methylation fragmented the conformational ensemble into multiple rapidly interconverting states (Figure 5A & SI Figure 11C). Consistent with these site-localized dynamics, state-resolved intra-residue contact maps from MSM analyses reveal tighter packing in the stabilized variants compared with the control. In the glycosylated ensemble, contacts around the Ser18 loop are enriched relative to the control, in line with the reduced RMSF observed in Figure 2,4B & SI Figure S13.

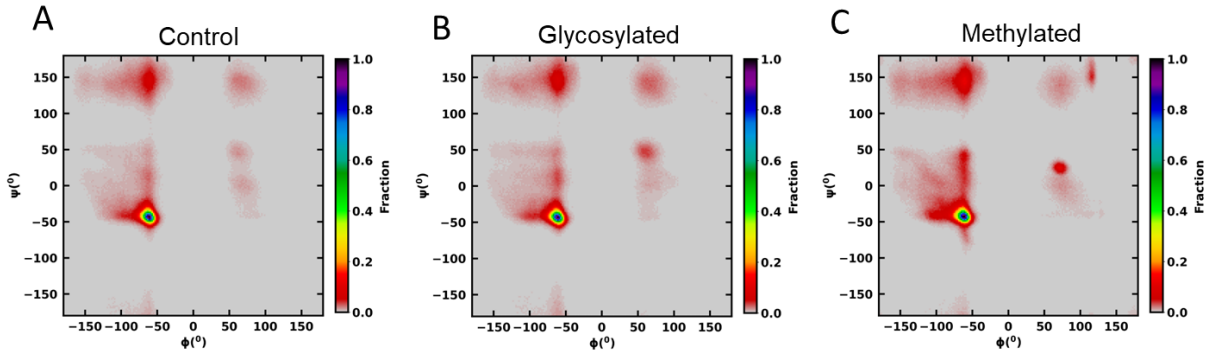


Figure 3. Ramachandran plot comparing backbone conformations of GccF models for A. *control*, B. *glycosylated* and C. *methylated* molecular systems.

The choice of the cosine-annealed VAE is supported by model selection: across annealing schemes and PCA baselines, cosine annealing yielded the highest five fold cross-validated VAMP-2 scores, and system-specific β_opt values (control 0.0085, glycosylated 0.0135, methylated 0.0024) were used to generate the latent spaces for FES and downstream kinetics (SI Figure S5).

The same qualitative ordering is recapitulated in SI Figure S11, which shows the latent-space FES for all three systems computed with the β-annealed VAE: the *control* covers a larger area with multiple shallow minima; the glycosylated peptide collapses into compact, well-defined minima; and the *methylated* system partitions its population across fewer but distinct basins compatible with loop rigidification inferred from Figure 2.

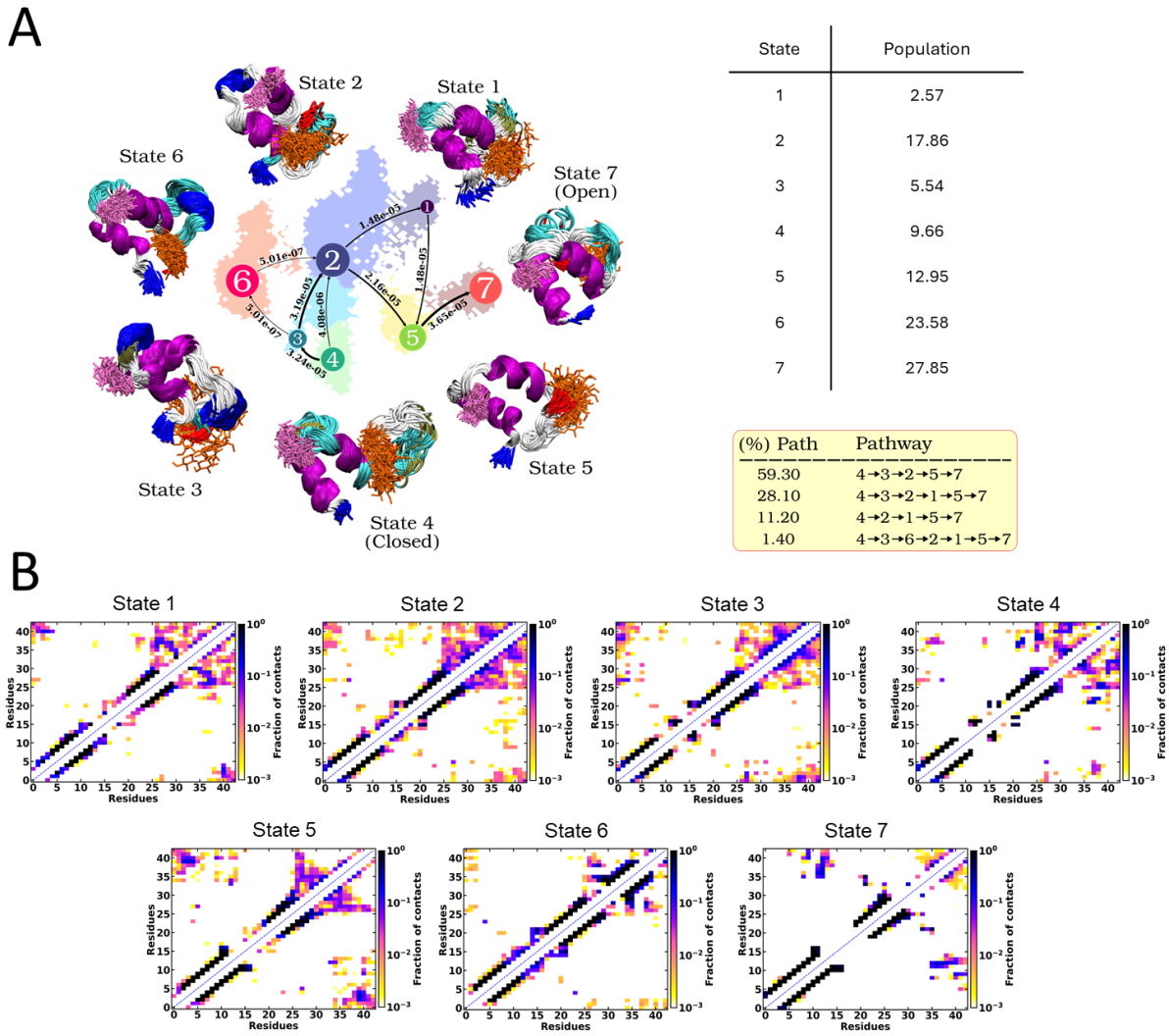
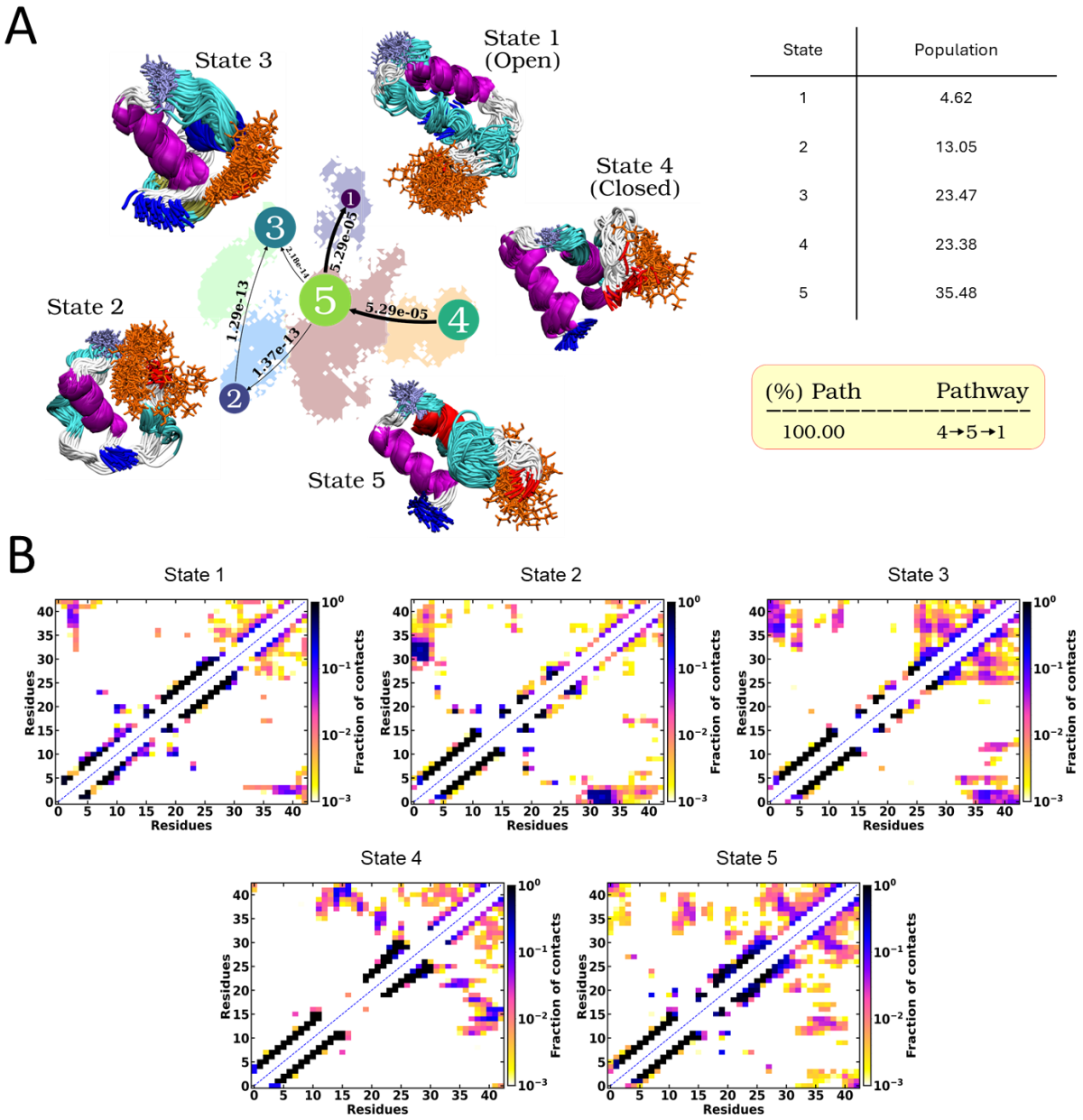


Figure 4. Markov state models and the proportion of each state with unidirectional transition probability and pathways between states derived using latent space of Cosine Annealed VAE from glycosylated dataset. The N-terminal and C-terminal residues are colored blue and red, and the helix and loop color are as per VMD default scheme. (B) Inter-residue wise contact maps for MSM states for glycosylated system.

The α-methylated system's MSM (Figure 5A) partitions the landscape into a handful of states whose connectivity reflects preferred routes between loop-rigidified conformers, in agreement with residue-wise contact analysis (Figure 5B) and suppressed loop fluctuations (Figure 2). These states remain more interconnected than those of the glycosylated peptide, consistent with a stabilized yet somewhat more permissive landscape. The methylation metastable states revealed destabilization in one of the alpha-helix facilitated by interactions of both the two S/O-linked GlcNAc (Figure 4A & 5A).

In contrast, the *control* exhibits a more ramified kinetic graph (SI Figure S12) and state-resolved contact maps with fewer persistent contacts around the Ser18 loop (SI Figure S13), matching its broader FES and higher RMSF. This control-glycosylated contrast anchors the interpretation of Figure 5 that the loop-rigidifying modifications compress the kinetic network by stabilizing intra-peptide contacts that restrict access to peripheral basins sampled by the *control*.

Taken together, the stability metrics (Figure 2), FES topology (Figure 3; SI Figure S11), and MSMs (Figure 4A,5A & SI Figure S12) converge on a consistent mechanism: post-translational modifications that reinforce local packing around Ser18, either through glycan-mediated interactions or α-methylation-induced backbone restrictions that narrow the conformational and kinetic ensembles by stabilizing loop-packed conformers and elevating barriers to excursions into flexible states. This modification-dependent reshaping of the energy landscape provides a structural–kinetic basis for the reduced conformational plasticity observed in the modified peptides relative to the control.



State	Population
1	4.62
2	13.05
3	23.47
4	23.38
5	35.48

Figure 5. (A) Markov state models and the proportion of each state with unidirectional transition probability and pathways between states derived using latent space of Cosine Annealed VAE from methylated dataset. The N-terminal and C-terminal residues are colored blue and red, and the helix and loop color are as per VMD default scheme. (B) Inter-residue wise contact maps for MSM states for methylated system.

4 Discussion

Across three GccF variants, control (non-glycosylated), native glycosylated, and α-methyl-Ser18, we find that post-translational context around Ser18 is the dominant lever shaping the peptide's

286 conformational and kinetic organization. Glycosylation narrows the accessible ensemble,
287 stabilizes the inter-helical loop (residues ~10–20), and lowers backbone flexibility, whereas α-
288 methylation perturbs the native hydrogen-bonding network and expands conformational
289 heterogeneity. These trends are concordant across stability metrics (RMSD/RMSF), latent-space
290 free-energy landscapes, and MSM kinetics.

291 Latent-space FES from the cosine-annealed VAE consolidate this picture: glycosylation
292 compresses population into fewer, deeper basins, while α-methylation broadens and destabilizes
293 metastable regions. The MSMs depicts glycosylated GccF occupies fewer, longer-lived states with
294 limited interconversion, whereas the α-methylated ensemble fragments into multiple rapidly
295 interconverting states. Together, these results indicate that glycan-mediated packing around
296 Ser18 kinetically "locks in" loop-packed conformers, while α-methylation unlocks alternative
297 routes and increases plasticity.

298 Methodologically, the cosine-annealed VAE provided the most faithful dynamical embedding
299 (highest cross-validated VAMP2), enabling robust state decomposition and model validation; the
300 MSM and FES analyses (SI Figure S11-12), and state-resolved contact maps (SI Figure S13 & Figure
301 4B & 5B) support these assignments and mechanistic links.

302 Overall, our MD–VAE–MSM framework reveals a modification-dependent reshaping of GccF's
303 energy landscape: native glycosylation enforces structural order and kinetic persistence, whereas
304 α-methylation at Ser18 disrupts native interactions and increases conformational flux. These
305 insights provide a structural-kinetic rationale for differences in functional behavior and suggest
306 concrete design principles centred around Ser18-proximal packing that endeavour to preserve
307 the native glycan-peptide contacts to tune activity in glycocin-like antimicrobial peptides.

308

309 **Authors contribution**

310 *Conceptualization: A.S.; Data curation: D.K.S. and S.A.; Formal analysis: D.K.S., S.A. Investigation: D.K.S.*
311 *(Simulations and Structural Analysis), S.A. (Simulations, VAE and MSMs); Methodology: D.K.S. (Simulations*
312 *and Parametrization), Validation: A.S.; Writing-original draft: D.K.S.; Writing-review and editing: A.S.*

313

Acknowledgement

This work was supported by the Department of Atomic Energy, Government of India, under Project Identification No. RTI4007. D.K.S acknowledges Dr. Jagannath Mondal, Associate Professor, Tata Institute of Fundamental Research, Hyderabad, Telangana, India-500046 for providing the computing facilities.

References

1. Almeida-Santos, A. C.; Novais, C.; Peixe, L.; Freitas, A. R. Enterococcus spp. as a Producer and Target of Bacteriocins: A Double-Edged Sword in the Antimicrobial Resistance Crisis Context. *Antibiotics* **2021**, *10* (10), 1215. https://doi.org/10.3390/antibiotics10101215.
2. Jain, M.; Stitt, G.; Son, L.; Enioutina, E. Y. Probiotics and Their Bioproducts: A Promising Approach for Targeting Methicillin-Resistant Staphylococcus aureus and Vancomycin-Resistant Enterococcus. *Microorganisms* **2023**, *11* (10), 2393. https://doi.org/10.3390/microorganisms11102393.
3. Cotter, P. D.; Ross, R. P.; Hill, C. Bacteriocins — a viable alternative to antibiotics? *Nature Reviews Microbiology* **2012**, *11* (2), 95–105. https://doi.org/10.1038/nrmicro2937.
4. Simons, A.; Alhanout, K.; Duval, R. E. Bacteriocins, Antimicrobial Peptides from Bacterial Origin: Overview of Their Biology and Their Impact against Multidrug-Resistant Bacteria. *Microorganisms* **2020**, *8* (5), 639. https://doi.org/10.3390/microorganisms8050639.
5. Venugopal, H.; Edwards, P. J. B.; Schwalbe, M.; Claridge, J. K.; Libich, D. S.; Stepper, J.; Loo, T.; Patchett, M. L.; Norris, G. E.; Pascal, S. M. Structural, dynamic, and chemical characterization of a novel S-Glycosylated bacteriocin. *Biochemistry* **2011**, *50* (14), 2748–2755. https://doi.org/10.1021/bi200217u.
6. Harjes, E.; Edwards, P. J. B.; Bisset, S. W.; Patchett, M. L.; Jameson, G. B.; Yang, S.-H.; Navo, C. D.; Harris, P. W. R.; Brimble, M. A.; Norris, G. E. NMR shows why a small chemical change almost abolishes the antimicrobial activity of glycocin F. *Biochemistry* **2023**, *62* (17), 2669–2676. https://doi.org/10.1021/acs.biochem.3c00197.
7. Stepper, J.; Shastri, S.; Loo, T. S.; Preston, J. C.; Novak, P.; Man, P.; Moore, C. H.; Havlíček, V.; Patchett, M. L.; Norris, G. E. Cysteine S -glycosylation, a new post-translational modification found in glycopeptide bacteriocins. *FEBS Letters* **2011**, *585* (4), 645–650. https://doi.org/10.1016/j.febslet.2011.01.023.
8. Norris, G. E.; Patchett, M. L. The glycocins: in a class of their own. *Current Opinion in Structural Biology* **2016**, *40*, 112–119. https://doi.org/10.1016/j.sbi.2016.09.003.

9. Lee, H. S.; Qi, Y.; Im, W. Effects of N-glycosylation on protein conformation and dynamics: Protein Data Bank analysis and molecular dynamics simulation study. *Scientific Reports* **2015**, *5* (1). https://doi.org/10.1038/srep08926.
10. Bednarska, N. G.; Wren, B. W.; Willcocks, S. J. The importance of the glycosylation of antimicrobial peptides: natural and synthetic approaches. *Drug Discovery Today* **2017**, *22* (6), 919–926. https://doi.org/10.1016/j.drudis.2017.02.001.
11. Jayaprakash, N. G.; Surolia, A. Role of glycosylation in nucleating protein folding and stability. *Biochemical Journal* **2017**, *474* (14), 2333–2347. https://doi.org/10.1042/bcj20170111.
    a. Mitra, N.; Sharon, N.; Surolia, A. Role of N-Linked Glycan in the Unfolding Pathway of Erythrina corallodendron Lectin. *Biochemistry* **2003**, *42* (42), 12208–12216. https://doi.org/10.1021/bi035169e.
    b. Erratum: N-linked oligosaccharides as outfitters for glycoprotein folding, form and function. *Trends in Biochemical Sciences* **2006**, *31* (5), 251. https://doi.org/10.1016/j.tibs.2006.03.010.
12. Robustelli, P.; Piana, S.; Shaw, D. E. Developing a molecular dynamics force field for both folded and disordered protein states. *Proceedings of the National Academy of Sciences* **2018**, *115* (21). https://doi.org/10.1073/pnas.1800690115.
13. Ghorbani, M.; Prasad, S.; Klauda, J. B.; Brooks, B. R. Variational embedding of protein folding simulations using Gaussian mixture variational autoencoders. *The Journal of Chemical Physics* **2021**, *155* (19). https://doi.org/10.1063/5.0069708.
14. Adhikari, S.; Mondal, J. Elucidating Protein Dynamics through the Optimal Annealing of Variational Autoencoders. *Journal of Chemical Theory and Computation* **2025**, *21* (13), 6367–6379. https://doi.org/10.1021/acs.jctc.5c00365.
15. Kingma, D. P.; Welling, M. *Auto-Encoding variational Bayes*. arXiv.org. https://arxiv.org/abs/1312.6114.
16. Ichikawa, Y.; Hukushima, K. Learning Dynamics in Linear VAE: Posterior Collapse Threshold, Superfluous Latent Space Pitfalls, and Speedup with KL Annealing. *arXiv (Cornell University)* **2023**. https://doi.org/10.48550/arxiv.2310.15440.
17. Chodera, J. D.; Noé, F. Markov state models of biomolecular conformational dynamics. *Current Opinion in Structural Biology* **2014**, *25*, 135–144. https://doi.org/10.1016/j.sbi.2014.04.002.
18. Husic, B. E.; Pande, V. S. Markov State Models: From an Art to a Science. *Journal of the American Chemical Society* **2018**, *140* (7), 2386–2396. https://doi.org/10.1021/jacs.7b12191.

19. Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry* **2005**, *26* (16), 1701–1718. https://doi.org/10.1002/jcc.20291.
20. Huang, J.; MacKerell, A. D. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *Journal of Computational Chemistry* **2013**, *34* (25), 2135–2145. https://doi.org/10.1002/jcc.23354.
21. Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of Computational Chemistry* **2009**, *31* (4), 671–690. https://doi.org/10.1002/jcc.21367.
22. Venugopal, H.; Edwards, P. J. B.; Schwalbe, M.; Claridge, J. K.; Libich, D. S.; Stepper, J.; Loo, T.; Patchett, M. L.; Norris, G. E.; Pascal, S. M. Structural, dynamic, and chemical characterization of a novel S-Glycosylated bacteriocin. *Biochemistry* **2011**, *50* (14), 2748–2755. https://doi.org/10.1021/bi200217u.
23. Seminario, J.M. Calculation of intramolecular force fields from second-derivative tensors. Int. J. *Quantum Chem.,* **1996** 60: 1271-1277. https://doi.org/10.1002/(SICI)1097-461X(1996)60:7<1271::AID-QUA8>3.0.CO;2-W
24. Frisch, M.J., Trucks, G.W., Schlegel, H.B., et al. Gaussian 16, **2016** Revision B.01. Gaussian, Inc., Wallingford.
25. Price, D. J.; Brooks, C. L. A modified TIP3P water potential for simulation with Ewald summation. *The Journal of Chemical Physics* **2004**, *121* (20), 10096–10103. https://doi.org/10.1063/1.1808117.
26. Boonstra, S.; Onck, P. R.; Van Der Giessen, E. CHARMM TIP3P water model suppresses peptide folding by solvating the unfolded state. *The Journal of Physical Chemistry B* **2016**, *120* (15), 3692–3698. https://doi.org/10.1021/acs.jpcb.6b01316.
27. Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics* **2007**, *126* (1). https://doi.org/10.1063/1.2408420.
28. Ke, Q.; Gong, X.; Liao, S.; Duan, C.; Li, L. Effects of thermostats/barostats on physical properties of liquids by molecular dynamics simulations. *Journal of Molecular Liquids* **2022**, *365*, 120116. https://doi.org/10.1016/j.molliq.2022.120116.
29. Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry* **1997**, *18* (12), 1463–1472. https://doi.org/10.1002/(sici)1096-987x(199709)18:12.

30. Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* **1993**, *98* (12), 10089–10092. https://doi.org/10.1063/1.464397.
31. Robustelli, P.; Ibanez-De-Opakua, A.; Campbell-Bezat, C.; Giordanetto, F.; Becker, S.; Zweckstetter, M.; Pan, A. C.; Shaw, D. E. Molecular basis of Small-Molecule binding to A-Synuclein. *Journal of the American Chemical Society* **2022**, *144* (6), 2501–2510. https://doi.org/10.1021/jacs.1c07591.
32. Scherer, M. K.; Trendelkamp-Schroer, B.; Paul, F.; Pérez-Hernández, G.; Hoffmann, M.; Plattner, N.; Wehmeyer, C.; Prinz, J.-H.; Noé, F. PyEMMA 2: A software package for estimation, validation, and analysis of Markov models. *Journal of Chemical Theory and Computation* **2015**, *11* (11), 5525–5542. https://doi.org/10.1021/acs.jctc.5b00743.
33. Prinz, J.-H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J. D.; Schütte, C.; Noé, F. Markov models of molecular kinetics: Generation and validation. *The Journal of Chemical Physics* **2011**, *134* (17). https://doi.org/10.1063/1.3565032.
34. Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *Journal of Computational Chemistry* **2011**, *32* (10), 2319–2327. https://doi.org/10.1002/jcc.21787.

35. Kabsch, W.; Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, *22* (12), 2577–2637. https://doi.org/10.1002/bip.360221211.

36. Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for processing and analysis of Molecular Dynamics Trajectory data. *Journal of Chemical Theory and Computation* **2013**, *9* (7), 3084–3095. https://doi.org/10.1021/ct400341p.

37. Case, D. A.; Aktulga, H. M.; Belfon, K.; Cerutti, D. S.; Cisneros, G. A.; Cruzeiro, V. W. D.; Forouzesh, N.; Giese, T. J.; Götz, A. W.; Gohlke, H.; Izadi, S.; Kasavajhala, K.; Kaymak, M. C.; King, E.; Kurtzman, T.; Lee, T.-S.; Li, P.; Liu, J.; Luchko, T.; Luo, R.; Manathunga, M.; Machado, M. R.; Nguyen, H. M.; O'Hearn, K. A.; Onufriev, A. V.; Pan, F.; Pantano, S.; Qi, R.; Rahnamoun, A.; Risheh, A.; Schott-Verdugo, S.; Shajan, A.; Swails, J.; Wang, J.; Wei, H.; Wu, X.; Wu, Y.; Zhang, S.; Zhao, S.; Zhu, Q.; Cheatham, T. E.; Roe, D. R.; Roitberg, A.; Simmerling, C.; York, D. M.; Nagan, M. C.; Merz, K. M. AmberTools. *Journal of Chemical Information and Modeling* **2023**, *63* (20), 6183–6191. https://doi.org/10.1021/acs.jcim.3c01153.