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. Glycocin F (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 preserves conformational plasticity to maintain functionally relevant conformations. In stark contrast, α-methylation introduces local rigidity, locking the peptide into fewer metastable basins with significantly slower transition rates. This leads to the disruptions of α-helix structure which perturbs the loop-helix coupling and trapping the peptide into non-functional conformations. Together, these findings provide 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.

Keywords: *multi-drug resistant, antibiotics, Glycocin F, Variational Autoencoder, glycosylation, α-methylation, rational design.*

1 Introduction

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

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

As a general post-translational modification, glycosylation is well-known to confer enhanced biophysical stability to peptides and proteins. This includes a significant increase in both thermal stability and resistance to proteolytic degradation, which are critical properties for the viability of any therapeutic peptide [9-13]. A recent study revealed an instructive counterexample to its general utility by application of α-methylation to GccF at Ser18. While the modification is typically used to rigidify peptides and increase their stability, the study revealed substitution of the Cα proton with a methyl group at Ser18 resulted in a catastrophic loss of antimicrobial activity, reducing its potency by an astonishing 1000-fold [6] and hence showing a marked difference in the orientation and mobility of the critical interhelical loop.

Since, the study by Harjes *et al.*, 2023 [6], argued the structural basis of this functional loss by 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 [14]. 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 [15-17]. 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 [18]. By dynamically adjusting the β factor, we reconstructed a more robust latent representation of our projected trajectory data [16]. 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 [19-20]. 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 [21] with the CHARMM36 force field [22-23]. Three variants of GccF were modeled: (i) a *control* system without glycosylation, where both GlcNAc moieties were removed; (ii) a *glycosylated* system

111 based on the native NMR structure (PDB ID: 2KUY) [24]; and (iii) a *methylated* system prepared
112 using the NMR structure (PDB ID: 8DFZ) [6]. Bonded and non-bonded parameters for the α-
113 methylated Ser18 linked GlcNAc residue (SI Figure S1) were calculated using the DFT Seminario
114 method [25] within the Gaussian 16 software package [26] to derive force constants, which were
115 used for all-atom MD simulations. Disulfide bonds between Cys4–Cys39 and Cys9–Cys43 were
116 enforced in all systems.

117 Each system was solvated in a cubic box of TIP3P water [27-28], extending at least 1.0 nm beyond
118 the peptide surface, and neutralized with Na^+/Cl^- ions to achieve a physiological ionic strength of
119 150 mM. Systems were energy-minimized using the steepest-descent algorithm until the
120 maximum force was below 100 kJ mol^{-1} nm^{-1}. Equilibration was performed in two stages: (i) 2 ns
121 of NVT equilibration using the V-rescale thermostat [29] with position restraints on heavy atoms,
122 and (ii) 2 ns of NPT equilibration using the Parrinello–Rahman barostat [30]. Production MD
123 simulations were conducted in the NPT ensemble with coordinates saved every 100 ps. All bonds
124 were constrained using the LINCS algorithm [31], and long-range electrostatics were treated with
125 the Particle Mesh Ewald (PME) method [32] using a 1.2 nm cutoff for both Coulomb and van der
126 Waals interactions. Each system was simulated in triplicate to ensure reproducibility.

127 **2.2 Variational Autoencoder and Free-energy Surface Construction**

128 Molecular dynamics trajectories for the *control, glycosylated,* and *methylated* systems were
129 processed to remove solvent molecules and ions, and periodic boundary conditions were
130 corrected, to extract features for dimensionality reduction. For each trajectory, the protein
131 backbone dihedral angles, specifically the Φ and Ψ torsion angles, were computed. To properly
132 represent the circular nature of this angular data for the neural network, the cosine and sine of
133 each angle were calculated, resulting in four feature components for each residue. These
134 components were combined into a single feature vector for each frame, and the complete
135 dataset for each system was subsequently partitioned into training and test sets using a 90:10
136 split. A β-Variational Autoencoder (VAE) [16] was employed to construct a non-linear, low-
137 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.

The decoder network was symmetrically structured, with layers of 16, 64, and 168 neurons. Training was performed for 1000 epochs using Stochastic Gradient Descent with a learning rate of 0.001, with batch sizes of 2048 for training and 2000 for validation. The total loss function was defined as

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

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

The β parameter was initially annealed from 0 to 1 during training using linear, cosine, and logistic schedules. For each system, the optimal β value was determined as the one that maximized the fraction of variance explained (FVE), following the procedure established in prior work. The VAEs were subsequently retrained by annealing β to these system-specific optimal values. Model selection among the annealing schemes was carried out using 5-fold cross-validated VAMP-2 scores (SI Figure S6). Among the tested approaches, cosine annealing VAE consistently yielded the highest scores, surpassing both the alternative annealing schedules and the leading two dimensions obtained via Principal Component Analysis (PCA). We investigated the free energy surface (FES) profile by plotting the first two principal components from PCA and projecting the trajectory data using the dihedral matrix into a lower dimensional space using our VAE approach. This indicated the efficiency of VAE over traditional PCA in capturing non-linear, collective variables that are crucial for describing complex protein dynamics (SI Figure S5 & S12).

2.3 Markov State Model (MSM) Analysis and Validation

We tested the convergence of all simulated MD trajectories of 3 µs each for *control*, *glycosylated* and *methylated* systems using time-dependent behavior and auto-correlation functions analysis [33]. We analyzed the change over time and the autocorrelation of two key metrics: the *Cα end-to-end distance* (the distance between the Cα atoms of the first and last residues) and the *Cα RMSD* (the root-mean-square deviation of each simulation frame from the initial frame). A convergence within 0.5 µs can be observed for all the three replicates (SI Figure S2) generated for the three molecular systems. The choice of the cosine-annealed VAE is supported by model

166 selection: across annealing schemes and PCA baselines. Consequently, the cosine-annealed VAE,
167 with optimal β values of 0.085 *(control)*, 0.135 *(glycosylated)*, and 0.024 *(methylated)* were
168 chosen to generate the two-dimensional latent spaces used for the construction of the MSMs
169 and cosine annealing yielded the highest VAMP-2 score (SI Figure S5-6). All MSMs were generated
170 using PyEMMA package [34]. The slow processes evolved from the MD trajectories were analyzed
171 using implied timescale (ITS) plot (SI Figure S8) and the number of macrostates were chosen
172 based on spectral analysis (SI Figure S9).

173 **2.3.1 Analysis and Validation**

174 The constructed MSMs were validated using Chapman-Kolmogorov [35] test (SI Figure 9-11)
175 against estimated models until lag time of 100ps. All trajectories were analyzed using MDAnalysis
176 [36] and custom Python scripts. Backbone root-mean-square deviation (RMSD) and per-residue
177 root-mean-square fluctuation (RMSF) were calculated relative to the initial conformations.
178 Secondary structure content was determined using the DSSP algorithm [36]. The S^2 order
179 parameters were computed using the cpptraj [38] module (version 6.18.1) from the AMBER
180 software suite [39], and the backbone NH bond vectors were used to determine the S^2 values
181 after removing the global rotational and translation motion using GROMACS v2021.6 [21].

182 **3 Results**

183 Using our MD-VAE-MSM framework, we studied the role of glycosylation in wild-type GccF
184 *(glycosylated)*, modified Ser18 GccF *(methylated)* along with *control* dataset with no glycosylation.
185 We simulated each molecular systems (Figure 1A) for an aggregated ~3 µs consisting three
186 replicate runs to study the distinct impacts of glycosylation and α-methylation on GccF's
187 conformational landscape and functional relationship. β-annealed VAE (Figure 1B) was able to
188 explore a wider conformational minima for the three molecular systems (Figure S5 & 12).

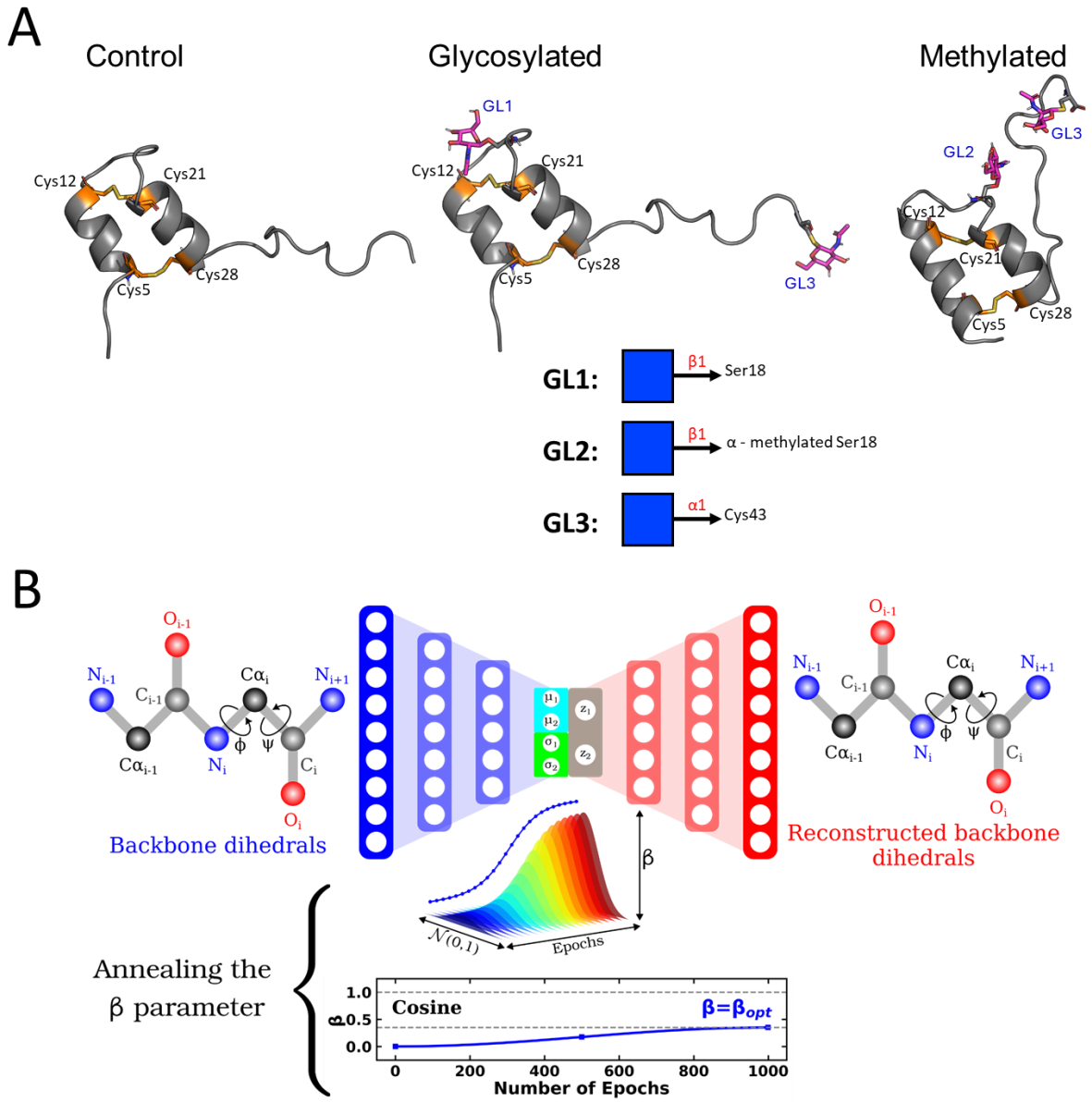


Figure 1. A. Structural representation of Glycocin F (GccF) showing the two glycosylation sites at Ser18 (β-O-GlcNAc) and Cys43 (β-S-GlcNAc), stabilized by nested disulfide bonds (Cys4–Cys39 and Cys9–Cys43). B. Schematic of β-annealed Variational AutoEncoder. A schematic of the reconstructed dihedrals 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

199 interhelical loop (residues 10–19) upon O-glycosylation in wild-type GccF and modified GccF,
200 whereas α-methylation at Ser18 introduced enhanced rigidity in this region (Figure 2B & 2D).
201 These findings align with previous reports linking O-GlcNAc modification to enhanced peptide
202 rigidity and resistance to unfolding [6]. MSM results revealed the transition time from "closed"
203 to "open" state is much higher for the *glycosylated* and *methylated* models (Figure 2C) compared
204 to *control* dynamics depicting much improved and effective role of both O-/S-linked GlcNAc and
205 α-methylation for the stabilization of the scaffold, with the largest effects concentrated in the
206 interhelical loop residues 10-19, that are poised to modulate glycan presentation. S^2 order
207 analysis (Figure 2B & 2D) revealed Ser18 is the most responsive region which is in good
208 agreement with the root mean square fluctuation (RMSF) analysis. Overall, the α-methylated
209 variant further rigidifies the Ser18-proximal interhelical loop.

210 Secondary structure analyses (SI Figure S3) confirmed that glycosylation preserved helical
211 segments, in contrast to the *methylated* system which exhibited partial loss of α-helical content.
212 Hydrogen-bond analyses (SI Figure S4) revealed prominent stabilization effect by glycan-peptide
213 contacts in the *glycosylated* system, reinforcing glycan modulated loop dynamics and stabilizing
214 the disulfide-anchored scaffold. In contrast, Ser18 α-methylation in *methylated* model showed
215 less prominent H-bonds compared to *glycosylated* model that leads to disrupted H-bonding
216 patterns in helical and interhelical loop regions due to steric constraints facilitated by the O-/S-
217 linked GlcNAc moiety. This altered backbone dihedral preference could result in reduced
218 stabilizing effect of glycosylation for the *methylated* model.

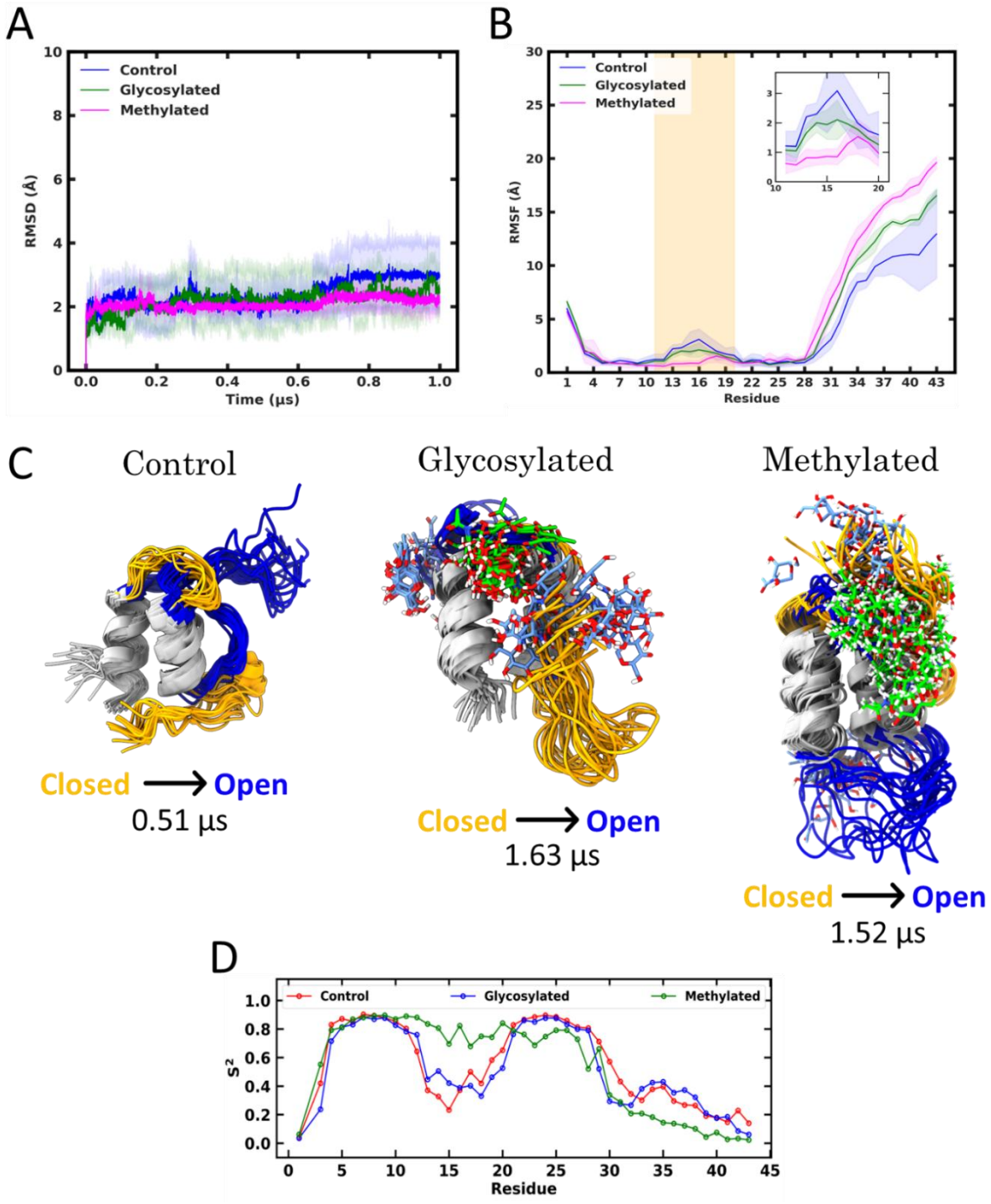


Figure 2. A. Backbone RMSD and B. per-residue RMSF comparisons among *control, glycosylated,* and *methylated* systems. C. Transition time of closed to opened metastable states obtained from MSM analysis for the three molecular systems. D. S^2 order analysis for *control, glycosylated,* and *methylated* 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 more clear, modification-dependent topology (Figure 3). The deglycosylated *control* samples most residues in favored α/β regions and generated a diffuse Ramachandran landscape characteristic of a flexible and dynamically unrestrained peptide backbone (Figure 3A). For the *glycosylated* model, the ϕ–ψ distributions collapse into discrete states located within canonical α-helical (ϕ ≈ –65°, ψ ≈ –35°) and β-turn regions (ϕ ≈ –75°, ψ ≈ 140°), indicating more dynamic conformational ensemble, as glycan-protein contacts could stabilize turns-like conformations (Figure 3B). Moreover, the reduced dispersion of dihedral angles demonstrates that glycan-peptide h-bonding (Figure 3B & SI Figure S4) could stabilize specific backbone orientations, particularly around Ser18 and the inter-helical loop. In contrast, the α-methyl-Ser18 variant shows a broader yet stabilized landscape, with density centered on loop-rigidified conformers. α-methylation at Ser18 could add 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).

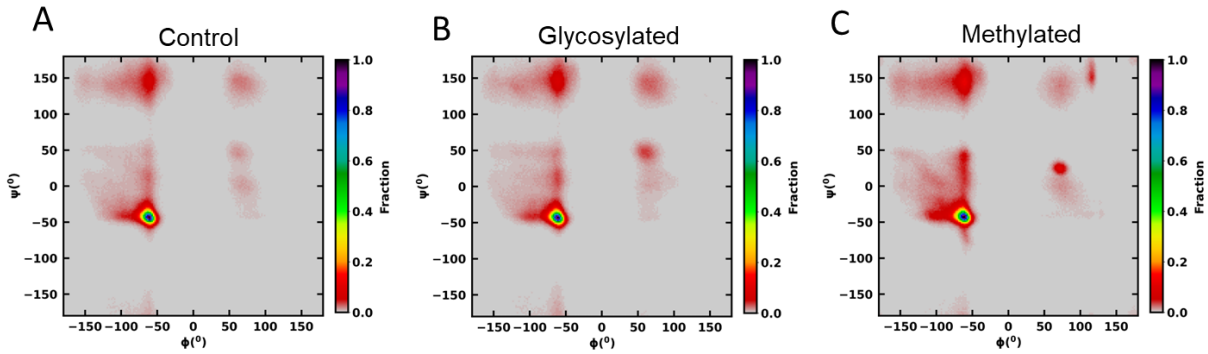


Figure 3. Ramachandran plot comparing backbone conformations of GccF models for A. *control*, B. *glycosylated* and C. *methylated* molecular systems.

Free energy surface (FES) derived from β-annealed VAE latent coordinates revealed that *glycosylated* model resolved into multiple rapidly interconverting states (Figure 4A, SI Tables S1 & SI Figure 12B), while methylation compressed the conformational ensemble into fewer metastable states with much slower transition rated from "closed" to "open" metastates (Figure 5A, SI Table S1 & SI Figure 12C). 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 2B,4B & SI Figure S14.

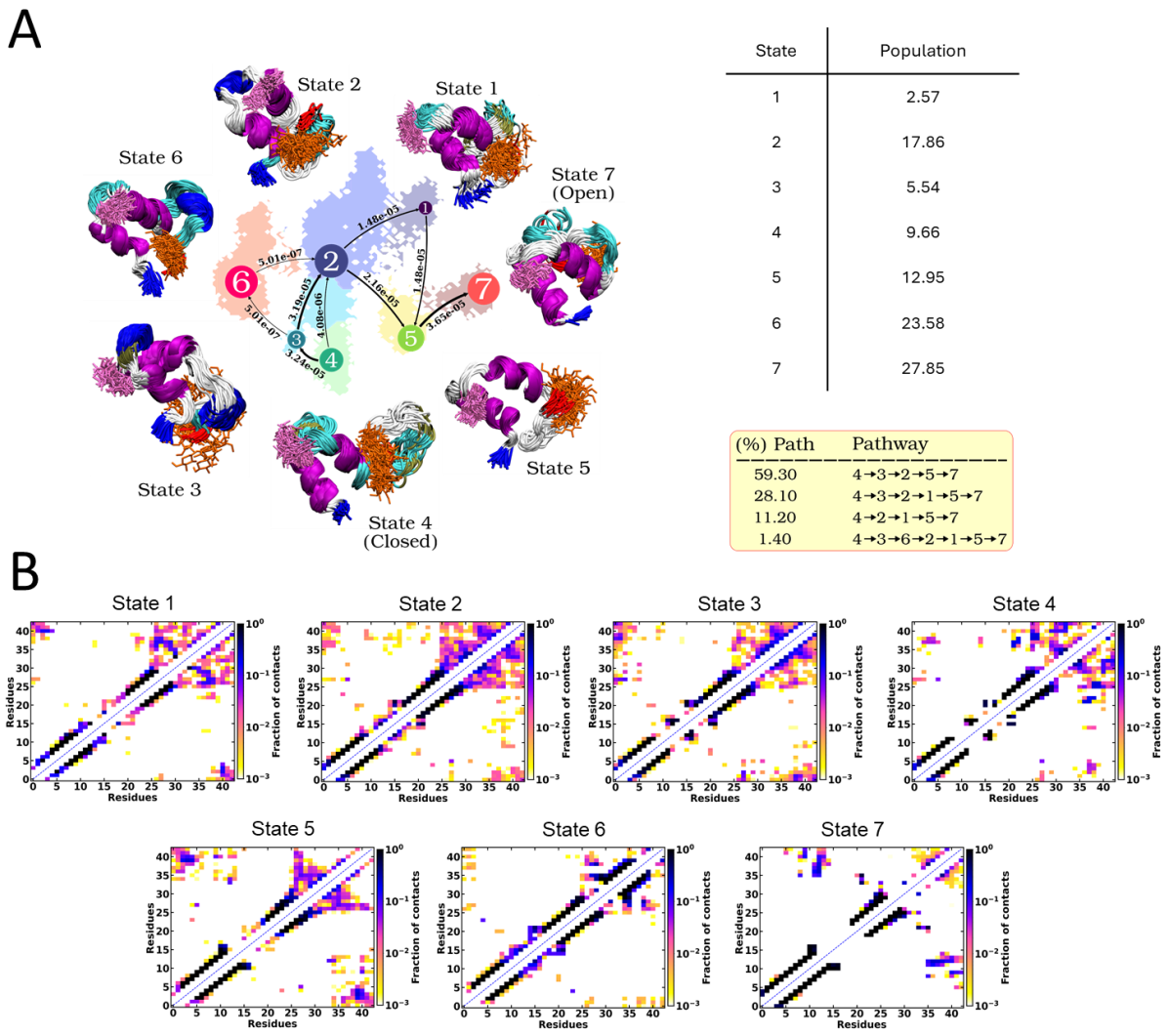


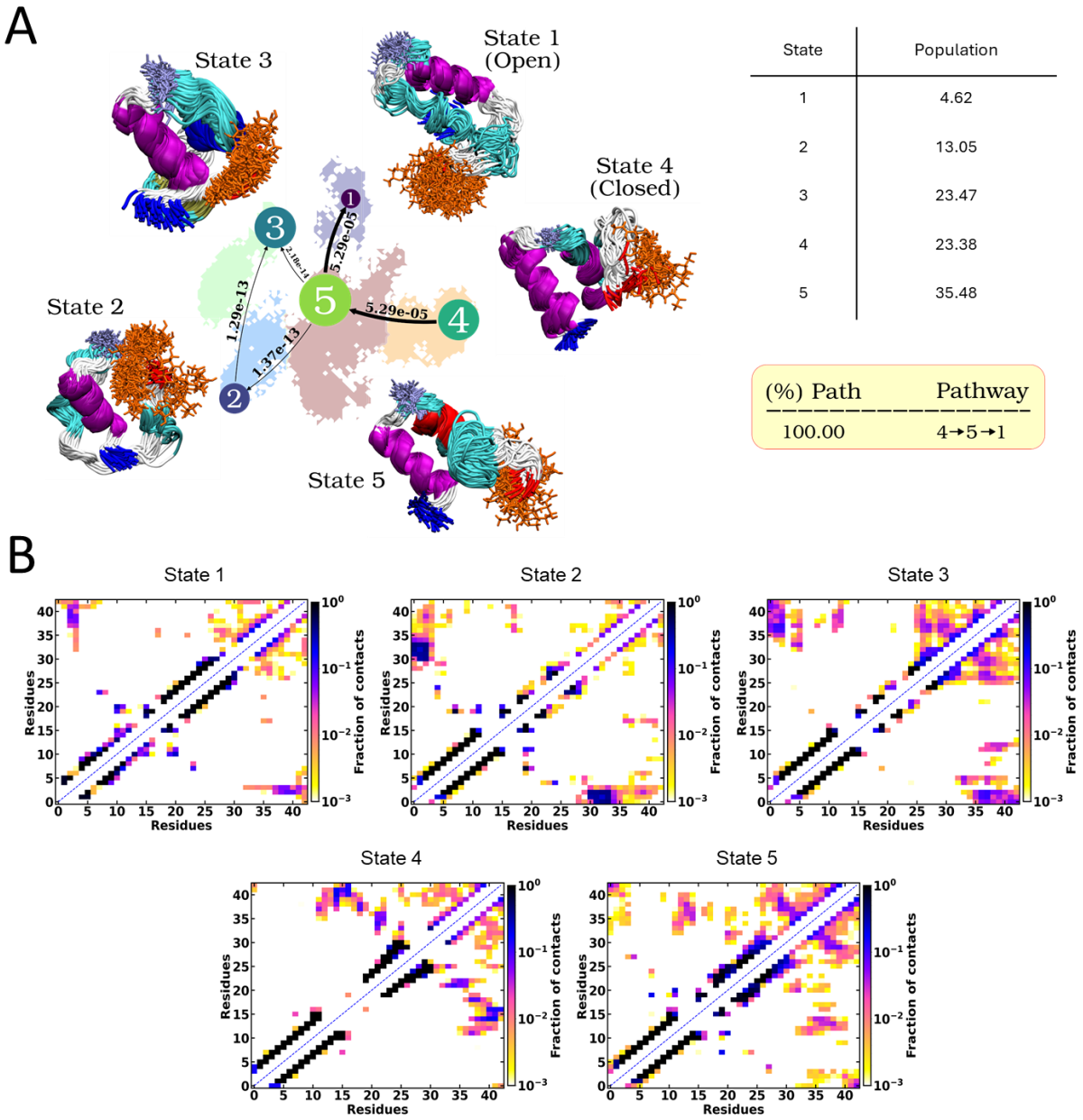
Figure 4. Markov state models and the proportion of each state with unidirectional transition probability and pathways between states derived using latent space of Cosine Annealed VAE from glycosylated dataset. The N-terminal and C-terminal residues are colored blue and red, and the helix and loop color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

260 The α-methylated system's MSM (Figure 5A) partitions the landscape into fewer states whose
261 connectivity reflects preferred routes between loop-rigidified conformers, in agreement with
262 residue-wise contact analysis (Figure 5B) and suppressed loop fluctuations (Figure 2B-C).
263 Whereas, the metastates remain more interconnected with faster transition time in the
264 *glycosylated* model, consistent with a stabilized yet somewhat more permissive landscape. The
265 methylation metastable states revealed destabilization in one of the alpha-helix facilitated by
266 interactions of both the two S/O-linked GlcNAc (Figure 4A & 5A).

267 In contrast, the *control* exhibits a more ramified kinetic graph (SI Figure S13) and state-resolved
268 contact maps with fewer persistent contacts around the Ser18 loop (SI Figure S14), matching its
269 broader FES and higher RMSF. Thus, establishing a progressive tightening of conformational
270 space from the flexible *control* to the *glycosylated* and ultimately to the *methylated* form. (SI
271 Table S1 & Figure 3).

272 Taken together, the stability profiles (Figure 2), FES topology (Figure 3; SI Figure S12), and MSMs
273 (Figure 4A,5A & SI Figure S13) converge on a consistent mechanism: post-translational
274 modifications that reinforce local packing around Ser18, either through glycan-mediated
275 interactions or α-methylation-induced backbone restrictions that narrow the conformational and
276 kinetic ensembles by stabilizing loop-packed conformers and elevating barriers to excursions into
277 flexible states. This modification-dependent reshaping of the energy landscape provides a
278 structural-kinetic basis for the reduced conformational plasticity observed in the *methylated*
279 model relative to the *control* and *glycosylated* models (Figure 4A, 5A, SI Table S1). While
280 glycosylation stabilizes the functional loop-helix coupling through organized yet interconvertible
281 basins, α-methylation locks the backbone into sterically constrained geometries, distorting one
282 α-helix as can be observed from the metastates 1-3 (Figure 5A) and kinetically trapping the
283 peptide in non-functional conformations.

284



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

294 conformational and kinetic organization. Glycosylation maintains the conformational plasticity
295 by stabilizing the helical and inter-helical loop (residues ~10–20), whereas α-methylation
296 perturbs the native hydrogen-bonding network and restricts the conformational heterogeneity.
297 These trends are concordant across stability metrics (RMSD/RMSF), latent-space free-energy
298 landscapes, and MSM metastates and transition rates.

299 Latent-space FES from the cosine-annealed VAE consolidate this picture: native glycosylation
300 expands the conformational ensemble into multiple well-defined basins that are rapidly
301 interconnected, while α-methylation compresses the landscape into fewer, deeper minima,
302 indicative of restricted flexibility and kinetic trapping within rigidified conformers. Together,
303 these results indicate that α-methylation facilitated glycan-mediated packing around Ser18
304 kinetically "locks in" conformations and loss of conformational adaptability, while native
305 glycosylation does not apply such steric constraints and maintains kinetic connectivity among
306 functional conformers. Methodologically, the cosine-annealed VAE provided the most faithful
307 dynamical embedding (highest cross-validated VAMP2 scores), allowing accurate reconstruction
308 of state populations and transition pathways. The FES maps (SI Figure S12) and state-resolved
309 contact networks (SI Figure S14; Figure 4B and 5B) collectively validate these mechanistic
310 distinctions.

311 Overall, our MD-VAE-MSM framework reveals a modification-dependent reshaping of GccF's
312 conformational landscape: native glycosylation enforces structural order and kinetic persistence,
313 whereas α-methylation at Ser18 disrupts native interactions and restricts conformational flux.
314 These insights provide a structural-kinetic rationale for differences in functional behavior and
315 suggest concrete design principles centered around Ser18-proximal packing that endeavor to
316 preserve the native glycan-peptide contacts to tune activity in glycocin-like antimicrobial
317 peptides.

318

319 **Authors contribution**

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

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.
12. 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.
13. 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.
14. 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.
15. 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.
16. 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.
17. Kingma, D. P.; Welling, M. *Auto-Encoding variational Bayes*. arXiv.org. https://arxiv.org/abs/1312.6114.
18. 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.
19. 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.

20. 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.
21. 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.
22. 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.
23. 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.
24. 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.
25. 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
26. Frisch, M.J., Trucks, G.W., Schlegel, H.B., et al. Gaussian 16, **2016** Revision B.01. Gaussian, Inc., Wallingford.
27. 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.
28. 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.
29. 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.
30. 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.

31. 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.
32. 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.
33. 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.
34. 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.
35. 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.
36. 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.

37. 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.

38. 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.

39. 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.