Accelerated sampling of protein dynamics using BioEmu augmented molecular simulation

Soumendranath Bhakat*[1]
[1] AlloTec Bio Inc., St. Louis, MO, United States

Email: bhakatsoumendranath@gmail.com / sbhakat@allotec.bio

Abstract

Here, we introduce a workflow that combines BioEmu generated conformational ensemble with physics based molecular simulations and Markov State Model to capture biomolecular conformational dynamics. BioEmu augmented molecular simulation approach captures active-to-inactive transitions in serine-threonine kinases and resolves how disease-causing mutations lead to population shifts among distinct metastable states. Compared to AlphaFold2 based reduced multiple sequence alignment (rMSA-AF2) approaches, BioEmu generated ensemble covers a broader conformational space and, when integrated with molecular dynamics simulations, enable Boltzmann weighted sampling of rare conformational events. Overall, this protocol provides a straightforward framework for integrating generative AI based protein emulator with statistical physics to recover Boltzmann weighted conformational ensembles at scale.

Introduction

Protein dynamics are central to biological function, and exploiting conformational motion to develop therapeutics is a cornerstone of dynamics-driven drug discovery. Functional regulation in proteins is rarely governed by a single static structure; instead, it emerges from transitions between metastable conformational states that control cryptic pocket opening, protein–protein interactions, and allosteric signal propagation. While high-resolution structures obtained from X-ray crystallography or cryo-electron microscopy provide invaluable snapshots of terminal states such as the active and inactive forms of kinases, they often fail to capture the dynamic intermediates that encode mechanism, ligand selectivity, and mutation-induced population shifts(1–5).

Molecular dynamics (MD) simulations provide a principled, physics-based framework for probing the thermodynamics and kinetics of protein motion(6, 7). However, unbiased MD is fundamentally limited by the timescale problem: biologically relevant conformational transitions are often separated by large free-energy barriers, inaccessible on practical simulation timescales(8). Enhanced sampling strategies have been developed to address this limitation and broadly fall into two classes: methods that bias along predefined collective variables(9–12), and methods that globally modify the potential energy surface through temperature scaling(13) or history dependent bias potentials(14). While powerful, these approaches suffer from critical drawbacks, most notably their reliance on prior knowledge of reaction coordinates (collective

variables) and the fact that the resulting population distributions are not intrinsically physical, requiring careful reweighting and specialized protocols to recover unbiased thermodynamics.

An alternative class of approaches, such as adaptive sampling, seeks to remain unbiased by iteratively launching ensembles of short MD trajectories from diverse conformational states, often identified via clustering of initial simulation data. While conceptually appealing, adaptive sampling workflows typically depend on carefully tuned clustering strategies, state definitions, and feedback protocols, and require specialized software frameworks and expert intervention(15, 16). As a result, these methods often exhibit a steep learning curve and limited accessibility, particularly when applied to complex and biologically relevant conformational transitions.

Recent advances in generative AI based protein structure/ensemble prediction algorithms have raised the question of whether data driven models can bridge this sampling gap(17–21). Methods based on AlphaFold2 (AF2), including AF2-RAVE(22, 23), AF2-MSM(24) and CryoPhold(25), exploit reduced multiple sequence alignments (rMSA) to induce conformational diversity. The central idea behind rMSA-AF2 is to generate a heterogeneous set of initial structures that can seed downstream unbiased MD simulations, thereby accelerating conformational exploration. While this strategy can capture cryptic pockets and alternate conformational states, the resulting physics refined ensembles remain strongly biased toward the "*coverage*" of the initial ensemble. If the initial coverage of rMSA-AF2 fails to capture meaningful diversity, then it is pretty unlikely that seeding short MD simulations will drastically enhance such sampling.

BioEmu(26) offers a fundamentally different paradigm. Rather than perturbing the inputs to a static structure predictor, BioEmu is a generative diffusion model fine-tuned on molecular dynamics simulation data and trained to reproduce statistically independent equilibrium structural distributions. This enables substantially *broader coverage* of conformational space compared to rMSA-AF2. However, BioEmu generated ensembles alone lack explicit physical populations and kinetic information. To recover physical meaning, we integrate BioEmu augmented molecular simulation (MD) with Markov State Models (MSMs)(27–29), yielding a unified framework that provides quantitative state populations and statistically rigorous free energy landscapes of conformational transitions expressed along physically meaningful and human readable collective variables.

To date, the molecular simulation community has largely focused on variation of rMSA-AF2 seeded MD workflows, where conformational diversity is introduced by subsampling evolutionary information. In contrast, BioEmu augmented molecular simulation replaces heuristic input perturbations with generative sampling grounded in molecular simulation data, enabling direct comparison of these two paradigms. In this work, we systematically evaluate BioEmu augmented MD simulation against rMSA-AF2 augmented MD in their ability to capture functionally relevant conformational transitions. We focus on two serine threonine kinases, where transitions between the DFG-in (active) and DFG-out (inactive) states involve large

conformational barriers and are typically induced by ligand binding(30). This makes them an ideal system to assess whether such transitions can be sampled *a priori*, without ligands, biasing potentials, or prohibitively long simulations starting from a single experimental structure. We further demonstrate that BioEmu augmented molecular simulation captures mutation induced population shifts with significantly reduced cumulative simulation time compared to rMSA-AF2 based approaches(31). Together, our results establish BioEmu guided MD as a straightforward protocol for accelerating access to protein motion beyond the reach of traditional enhanced sampling and rMSA-AF2 seeded workflows, enabling practical exploration of functionally important protein conformations central to drug discovery.

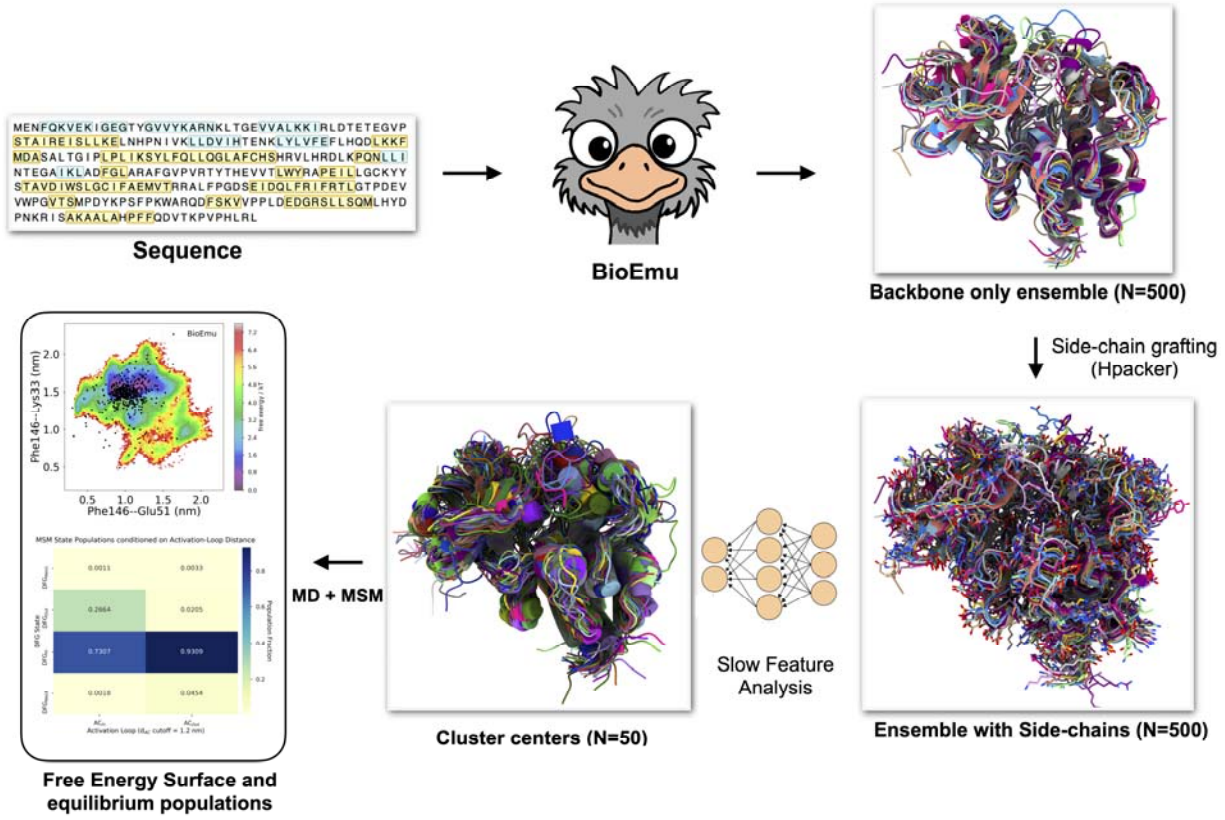

Figure 1. Schematic representation of BioEmu augmented molecular simulation**.** BioEmu takes a protein sequence as input and generates a backbone only conformational ensemble of a protein monomer. Hpacker(32) is then used to add side chains and produce all-atom representations of the ensemble. Slow feature analysis is performed on Cα–Cα distances, and *K means* clustering (N = 50) on the first two slow features is used to generate a latent structural ensemble. These structures are used to initiate 50 independent 100 ns MD simulations for each protein system studied in this manuscript. The combined simulation data are used to build an MSM, from which MSM weighted free-energy surfaces are projected along functionally relevant, human-readable collective variables, and the populations of macrostates and substates are computed from the equilibrium distribution.

Results

BioEmu augmented molecular simulation sampled the equilibrium free energy landscapes of apo BRAF and CDK2, capturing conformational transitions between DFG-in and DFG-out states. rMSA-AF2 generated conformational ensembles predominantly restricted to the DFG-in state, whereas all-atom reconstruction of BioEmu generated ensembles populated a broader conformational space beyond DFG-in state (**Figure 1 A-D**). Markov State Models constructed using two key structural variables governing the DFG transition resolved not only terminal DFG-in and DFG-out states but also intermediate states (DFG_{Neo}; Figure S1 in Supporting Information) and additional substates, key features of local conformational dynamics. In apo CDK2, BioEmu augmented MD captured transitions between extended (AC_{out}) and folded (AC_{in}) activation loop conformations and quantified their populations within DFG-in and DFG-out macrostates (**Figure 2 A, B**). Transition from DFG-in to DFG-out shifts the population of AC_{out} towards AC_{in} substate, a key hallmark of active to inactive conformational transition in apo CDK2(33, 34).

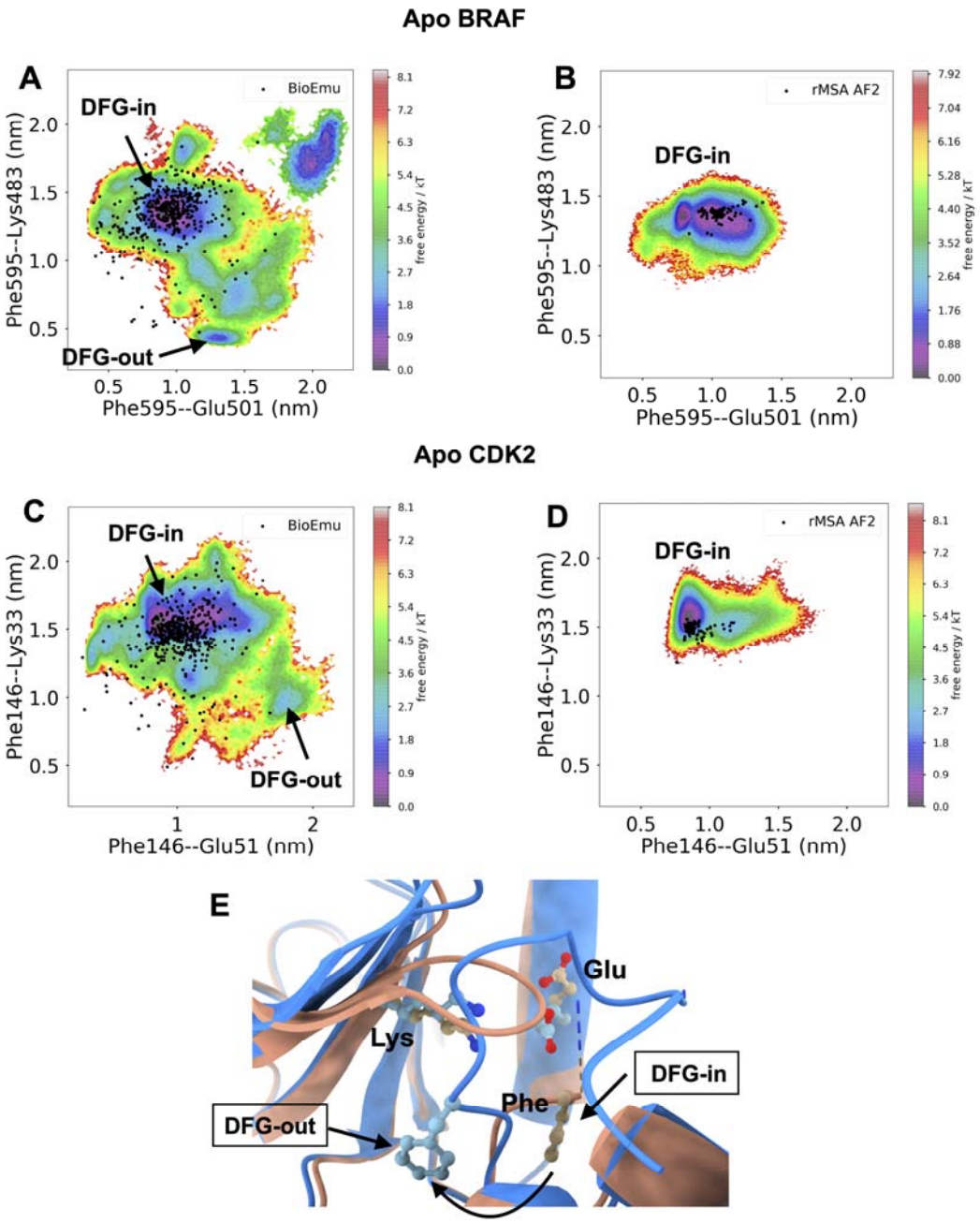

Figure 2. MSM weighted free-energy surfaces resolve DFG-in to DFG-out conformational transitions. MSM weighted free-energy surfaces projected along key inter-residue distances (Phe Cα–Glu Cα (x-axis) and Lys Cα–Phe Cζ (y-axis)) capture conformational transitions between DFG-in and DFG-out states. BioEmu augmented molecular simulations (total simulation time: 5 µs) sample the DFG-in to DFG-out transition in apo wild-type BRAF and CDK2 (A, C), whereas rMSA-AF2 augmented MD simulations (total simulation time: 8 µs) remain confined to the DFG-in state (B, D). Black dots denote conformational ensembles generated by BioEmu (n = 500) and rMSA-AF2 (n = 80), illustrating the broader conformational coverage achieved by BioEmu relative to rMSA-AF2. Representative structural orientations of the DFG motif are shown for the DFG-in (salmon) and DFG-out (cyan) states, highlighting the relative positioning of the DFG phenylalanine, lysine, and glutamate residues (E).

We next investigated whether BioEmu augmented MD simulations can capture mutation-induced population shifts between wild-type and V600E BRAF. BioEmu augmented molecular simulations successfully resolve a population shift of the DFG-Phe side chain from the Phe_{F1} rotamer ($\chi1 \approx -1$ rad) to the Phe_N rotamer ($\chi1 \approx +1$ rad) within the DFG-in macrostate (**Figure 2 C, D**), a hallmark conformational motion associated with V600E mediated BRAF activation(35, 36).

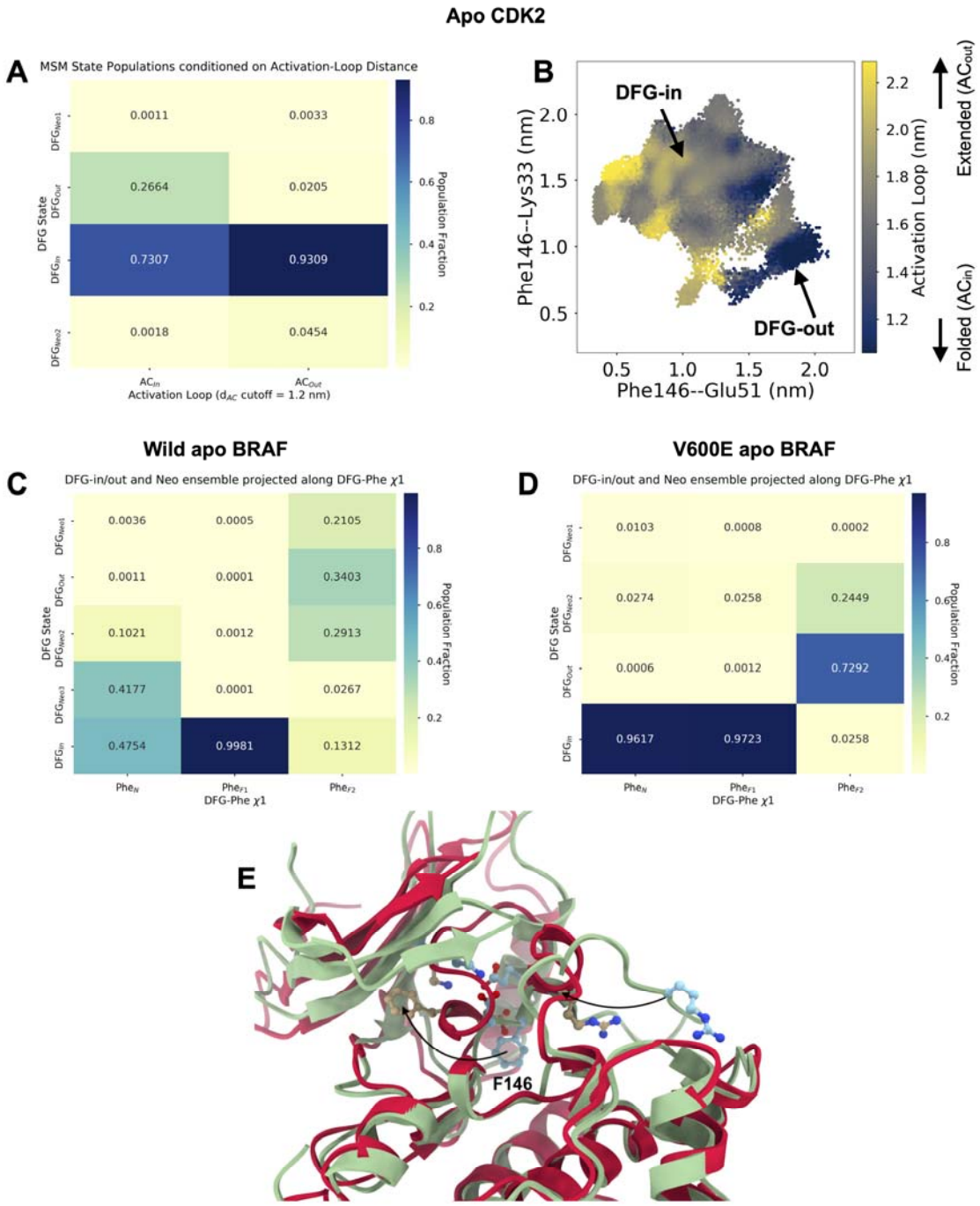

Figure 2. BioEmu augmented molecular simulations capture equilibrium populations of macrostates, their substates, and mutation-induced population shifts. In apo CDK2 (A, B), the simulations resolve the populations of extended (AC_{out}) and folded (AC_{in}) activation-loop conformations as a function of DFG state. The transition from extended to folded activation loop conformations is a hallmark of the active-to-inactive transition in CDK2. Structural representations comparing DFG-out/AC_{in} with DFG-in/AC_{out} highlight the conformational differences between inactive and active CDK2 states (E). The activation loop distance was calculated using D145-CA and R157-CA atoms. A cutoff of 1.2 nm was set to distinguish folded and extend states. In apo BRAF, BioEmu-augmented simulations capture a pronounced shift in the population fraction of the Phe_N rotamer within the DFG-in macrostate upon V600E mutation (WT: 0.4754, V600E: 0.9617), a defining feature of V600E mediated BRAF activation (C, D).

Conclusions

In this working paper, we demonstrate how BioEmu augmented molecular simulations (cumulative 5 µs per protein) resolve functionally important conformational states and mutation induced population shifts beyond what is captured by rMSA-AF2 augmented molecular simulations (cumulative 8 µs per protein). A key finding is that, for the serine-threonine kinases reported here, BioEmu generated conformational ensemble provide substantially broader conformational coverage than rMSA-AF2 derived ensemble, a critical factor governing accelerated sampling when integrated with physics based molecular simulation. However, conformations generated by BioEmu alone do not encode physical information, such as state populations or weights, making it challenging to prioritize functionally relevant states for downstream drug discovery workflows, such as population weighted ensemble docking(37). By incorporating physics based molecular simulations on top of BioEmu generated ensemble, we assign statistically meaningful weights to macro and microstates and capture mutation-induced population shifts, a hallmark of many human diseases. While initiating short MD simulations from approximately 500 BioEmu generated structures could be computationally demanding, we mitigate this cost by training slow feature analysis (SFA)(37) on Cα–Cα distances derived from the BioEmu ensemble and clustering along the first two slow features, yielding a reduced latent ensemble of 50 representative structures. This reduction enables efficient parallel simulations on a standard GPU cluster. Moreover, the features identified by SFA can serve as collective variables for enhanced sampling, providing flexibility in the choice of sampling strategy. Although slow feature analysis is used here, the protocol is general and can be adapted to alternative dimensionality reduction approaches, such as time-lagged autoencoders(38, 39) or principal component analysis(12, 40), depending on the motions of interest.

A current limitation of this approach is the inability of BioEmu v1.0 to explicitly model protein-protein and protein-ligand interactions, which restricts our ability to directly assess how such interactions modulate the conformational landscape. In addition, BioEmu does not model side-chain orientations; instead, side chains are later grafted using external tools to generate all-atom representations of the conformational ensemble. Because side-chain dynamics play a critical role

in allosteric regulation across many members of protein families, limited side-chain conformational diversity in the reconstructed all-atom ensembles may reduce the extent to which subsequent physics-based molecular simulations accelerate sampling of functionally important alternative states. The impact of side-chain orientations on the efficiency and accuracy of BioEmu augmented molecular simulations remains an active area of our ongoing investigation.

Future efforts will focus on integrating BioEmu generated all-atom conformational ensembles with Bayesian reweighting(41) to incorporate experimental priors from methods such as CryoPhold(25) or DEERFold(42). This integration will enable refinement of generative AI derived ensembles using experimental constraints, including cryo-EM density maps and single-molecule spectroscopy data, resulting in a robust pipeline that combines generative AI, biophysical experiments, and molecular simulation to capture protein dynamics at scale(43).

Methods

Ensemble generation

BioEmu(26) v1.0 (GitHub: https://github.com/microsoft/bioemu) was used to generate conformational ensembles comprising 500 structures for apo wild type CDK2, apo BRAF, and apo V600E BRAF. All-atom models were constructed by grafting side chains using H-packer(32) (GitHub: https://github.com/gvisani/hpacker) software.

Feature selection and slow feature analysis

All Cα–Cα distances from the all-atom BioEmu generated conformational ensemble were converted into smooth, differentiable features using a continuous contact switching function (see *Bhakat et al*.(36)),

$$S(r) = \frac{1}{1 + \left(\frac{r}{r_{\text{decay}}}\right)^n}$$

where r is the *Cα–Cα* distance, r_{decay} controls the transition length scale, and n sets the sharpness of the transition. Short distances ($r \ll r_{\text{decay}}$) yield values close to 1 (formed contact), while long distances ($r \gg r_{\text{decay}}$) yield values close to 0 (broken contact), providing a noise-robust, continuous measure of contact strength. We used $r_{\text{decay}} = 0.90$ nm and $n = 6$.

The Cα–Cα contacts were transformed using this function, yielding datapoints of $S(r) \in [0,1]$. To retain only dynamically informative contacts, we computed the minimum and maximum $S(r)$ values for each contact across all trajectories and retained contacts satisfying

$$min(S) < 0.4 \; and \; max\;(S) > 0.6.$$

This filtering removes contacts that remain persistently formed or broken and selects contacts that undergo meaningful formation and disruption in the all-atom BioEmu ensemble. For interpretability, $S(r) \geq 0.6$ corresponds to formed contacts, $S(r) \leq 0.4$ to broken contacts, and intermediate values reflect transient contact fluctuations. The filtered contact set was used for subsequent dimensionality reduction.

The switching function transformed filtered distances were then used as input to Slow Feature Analysis (SFA), a dimensionality reduction technique for biomolecular simulations introduced by *Vats et al.(37)* in context of biomolecular data. SFA transforms the J-dimensional input signal $\mathbf{c}(t)$ into output signals $y_k(t) = g_k(\mathbf{c}(t))$ by minimizing the temporal variation,

$$\Delta y_k = \langle \dot{y}_k^2 \rangle,$$

thereby extracting features that vary slowly over time. Applied to the filtered contact features, SFA identifies first two linear combinations of slowly fluctuating distance pairs that capture conformational hotspots from the all-atom BioEmu generated ensemble. K-means clustering (N=50) was subsequently performed on the first two slow features and subsequent cluster centers were saved in a PDB format for molecular simulation. SFA and clustering were performed using the MDML software package (GitHub: https://github.com/svats73/mdml/tree/main)

Molecular simulation

Representative structures from the SFA derived posterior ensemble were selected as starting points for unbiased molecular dynamics simulations. Each structure was prepared using the *tleap* module in Amber2022(44, 45) following the protocols described by *Meller et al.(24)* Proteins were parameterized with the AMBER ff14SB(46) force field, neutralized with counterions, and solvated in a truncated-octahedron TIP3P(47) water box with a minimum buffer of 10 Å from any protein atom to the box boundary. Energy minimization was performed in two stages: solvent and ions were first minimized with harmonic restraints on protein atoms (100 kcal mol□¹ Å□²), followed by unrestrained minimization of the full system.

Amber topologies were converted to GROMACS format using ACPYPE(48), and all subsequent simulations were carried out in GROMACS 2022(49). Systems were heated from 0 to 300 K over 500 ps in the NVT ensemble with backbone heavy-atom restraints (500 kJ mol□¹ nm□²), followed by 200 ps of unrestrained equilibration in the NPT ensemble at 300 K and 1 bar. Temperature was controlled using a velocity-rescale thermostat and pressure using a *Parrinello–Rahman*(50) barostat. Production simulations were performed in the NPT ensemble using a 2 fs

time step, a 1.0 nm cutoff for nonbonded interactions, particle-mesh Ewald(51) electrostatics, and LINCS(52) constraints on bonds involving hydrogen atoms.

Unbiased production simulations of 100 ns were initiated from the SFA selected cluster centers, with trajectories saved every 10 ps, resulting in a cumulative simulation time of 5 μs for each protein system. rMSA-AF2 augmented molecular simulation data (total 8 μs for each protein system) were repurposed from Bhakat *et al(36).* Markov state models were constructed from the resulting trajectories using the PyEMMA(53) package.

Competing interests

Author declares no conflict of interests.

References

1. D. R. Livesay, Protein dynamics: dancing on an ever-changing free energy stage. *Curr Opin Pharmacol* [Preprint] (2010).
2. A. P. Kornev, S. S. Taylor, Dynamics-Driven Allostery in Protein Kinases. *Trends Biochem Sci* 40, 628–647 (2015).
3. J. Guo, H.-X. Zhou, Protein Allostery and Conformational Dynamics. *Chem Rev* 116, 6503–6515 (2016).
4. A. Meller, D. Kelly, L. G. Smith, G. R. Bowman, Toward physics-based precision medicine: Exploiting protein dynamics to design new therapeutics and interpret variants. *Protein Science* 33, e4902 (2024).
5. G. R. Bowman, AlphaFold and Protein Folding: Not Dead Yet! The Frontier Is Conformational Ensembles. *Annu Rev Biomed Data Sci* 7, 51–57 (2024).
6. S. A. Hollingsworth, R. O. Dror, Molecular Dynamics Simulation for All. *Neuron* 99, 1129–1143 (2018).
7. J. L. Klepeis, K. Lindorff-Larsen, R. O. Dror, D. E. Shaw, Long-timescale molecular dynamics simulations of protein structure and function. *Curr Opin Struct Biol* 19, 120–127 (2009).
8. R. C. Bernardi, M. C. R. Melo, K. Schulten, Enhanced sampling techniques in molecular dynamics simulations of biological systems. *Biochimica et Biophysica Acta (BBA) - General Subjects* 1850, 872–877 (2015).
9. A. Barducci, M. Bonomi, M. Parrinello, Metadynamics. *WIREs Computational Molecular Science* 1, 826–843 (2011).
10. A. Barducci, G. Bussi, M. Parrinello, Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys Rev Lett* 100, 20603 (2008).
11. P. Tiwary, M. Parrinello, From Metadynamics to Dynamics. *Phys Rev Lett* 111, 230602 (2013).

12. S. Bhakat, Collective variable discovery in the age of machine learning: reality, hype and everything in between. *RSC Adv* 12, 25010–25024 (2022).
13. P. Liu, B. Kim, R. A. Friesner, B. J. Berne, Replica exchange with solute tempering: A method for sampling biological systems in explicit water. *Proceedings of the National Academy of Sciences* 102, 13749–13754 (2005).
14. Y. Miao, V. A. Feher, J. A. McCammon, Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation. *J Chem Theory Comput* 11, 3584–3595 (2015).
15. M. I. Zimmerman, G. R. Bowman, FAST Conformational Searches by Balancing Exploration/Exploitation Trade-Offs. *J Chem Theory Comput* 11, 5747–5757 (2015).
16. D. M. Zuckerman, L. T. Chong, Weighted Ensemble Simulation: Review of Methodology, Applications, and Software. *Annu Rev Biophys* 46, 43–57 (2017).
17. J. Abramson, *et al*., Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 630, 493–500 (2024).
18. B. Jing, B. Berger, T. Jaakkola, AlphaFold Meets Flow Matching for Generating Protein Ensembles. *ArXiv* (2024). https://doi.org/https://doi.org/10.48550/arXiv.2402.04845.
19. T. Saldaño, *et al.*, Impact of protein conformational diversity on AlphaFold predictions. *Bioinformatics* 38, 2742–2748 (2022).
20. J. Jumper, *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589 (2021).
21. Y. Jin, *et al.*, P2DFlow: A Protein Ensemble Generative Model with SE(3) Flow Matching. *J Chem Theory Comput* 21, 3288–3296 (2025).
22. B. P. Vani, A. Aranganathan, D. Wang, P. Tiwary, AlphaFold2-RAVE: From Sequence to Boltzmann Ranking. *J Chem Theory Comput* 19, 4351–4354 (2023).
23. B. P. Vani, A. Aranganathan, P. Tiwary, Exploring Kinase Asp-Phe-Gly (DFG) Loop Conformational Stability with AlphaFold2-RAVE. *J Chem Inf Model* 64, 2789–2797 (2024).
24. A. Meller, S. Bhakat, S. Solieva, G. R. Bowman, Accelerating Cryptic Pocket Discovery Using AlphaFold. *J Chem Theory Comput* 19, 4355–4363 (2023).
25. S. Bhakat, S. Vats, A. Mardt, E. M. Strauch, CryoPhold: CryoEM meets AlphaFold and molecular simulation to reveal protein dynamics. *bioRxiv* 2025.09.12.675912 (2025). https://doi.org/10.1101/2025.09.12.675912.
26. S. Lewis, *et al.*, Scalable emulation of protein equilibrium ensembles with generative deep learning. *Science (1979)* 389, eadv9817 (2026).
27. B. E. Husic, V. S. Pande, Markov State Models: From an Art to a Science. *J Am Chem Soc* 140, 2386–2396 (2018).
28. G. Pérez-Hernández, F. Paul, T. Giorgino, G. De Fabritiis, F. Noé, Identification of slow molecular order parameters for Markov model construction. *J Chem Phys* 139, 15102 (2013).
29. V. S. Pande, K. Beauchamp, G. R. Bowman, Everything you wanted to know about Markov State Models but were afraid to ask. *Methods* 52, 99–105 (2010).
30. J. Gizzio, A. Thakur, A. Haldane, R. M. Levy, Evolutionary divergence in the conformational landscapes of tyrosine vs serine/threonine kinases. *Elife* 11, e83368 (2022).

31. I. Taneja, *et al.*, AlphaFold-RandomWalk and AlphaFold-Ensemble: Sampling Alternative Protein Conformations with Perturbed Versions of AlphaFold. *J Chem Inf Model* (2025). https://doi.org/10.1021/acs.jcim.5c02004.
32. G. M. Visani, W. Galvin, M. N. Pun, A. Nourmohammad, H-Packer: Holographic Rotationally Equivariant Convolutional Neural Network for Protein Side-Chain Packing. *ArXiv* (2023).
33. T. G. Davies, *et al.*, Inhibitor Binding to Active and Inactive CDK2: The Crystal Structure of CDK2-Cyclin A/Indirubin-5-Sulphonate. *Structure* 9, 389–397 (2001).
34. A. Majumdar, *et al.*, Allostery governs Cdk2 activation and differential recognition of CDK inhibitors. *Nat Chem Biol* 17, 456–464 (2021).
35. R. C. Maloney, M. Zhang, H. Jang, R. Nussinov, The mechanism of activation of monomeric B-Raf V600E. *Comput Struct Biotechnol J* 19, 3349–3363 (2021).
36. S. Bhakat, S. Vats, A. Mardt, A. Degterev, Generalizable Protein Dynamics in Kinases: Physics is the key. *bioRxiv* 2025.03.06.641878 (2025). https://doi.org/10.1101/2025.03.06.641878.
37. S. Vats, R. Bobrovs, P. Söderhjelm, S. Bhakat, AlphaFold-SFA: Accelerated sampling of cryptic pocket opening, protein-ligand binding and allostery by AlphaFold, slow feature analysis and metadynamics. *PLoS One* 19, e0307226 (2024).
38. C. Wehmeyer, F. Noé, Time-lagged autoencoders: Deep learning of slow collective variables for molecular kinetics. *J Chem Phys* 148, 241703 (2018).
39. J. M. L. Ribeiro, P. Bravo, Y. Wang, P. Tiwary, Reweighted autoencoded variational Bayes for enhanced sampling (RAVE). *J Chem Phys* 149, 72301 (2018).
40. T. Gökdemir, J. Rydzewski, Machine learning of slow collective variables and enhanced sampling via spatial techniques. *Chemical Physics Reviews* 6, 11304 (2025).
41. J. Köfinger, *et al.*, Efficient Ensemble Refinement by Reweighting. *J Chem Theory Comput* 15, 3390–3401 (2019).
42. T. Wu, R. A. Stein, T.-Y. Kao, B. Brown, H. S. Mchaourab, Modeling protein conformational ensembles by guiding AlphaFold2 with Double Electron Electron Resonance (DEER) distance distributions. *Nat Commun* 16, 7107 (2025).
43. S. Bottaro, K. Lindorff-Larsen, Biophysical experiments and biomolecular simulations: A perfect match? *Science (1979)* 361, 355–360 (2018).
44. D. A. Case, *et al.*, Amber 2022, University of California, San Francisco. (2022).
45. R. Salomon-Ferrer, D. A. Case, R. C. Walker, An overview of the Amber biomolecular simulation package. *WIREs Computational Molecular Science* 3, 198–210 (2013).
46. J. A. Maier, *et al.*, ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J Chem Theory Comput* 11, 3696–3713 (2015).
47. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79, 926–935 (1983).
48. A. W. Sousa da Silva, W. F. Vranken, ACPYPE - AnteChamber PYthon Parser interfacE. *BMC Res Notes* 5, 367 (2012).
49. M. J. Abraham, *et al.*, GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1–2, 19–25 (2015).
50. M. Parrinello, A. Rahman, Polymorphic transitions in single crystals: A new molecular dynamics method. *J Appl Phys* 52, 7182–7190 (1981).

51. T. Darden, D. York, L. Pedersen, Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J Chem Phys* 98, 10089–10092 (1993).
52. B. Hess, H. Bekker, H. J. C. Berendsen, J. G. E. M. Fraaije, LINCS: A linear constraint solver for molecular simulations. *J Comput Chem* 18, 1463–1472 (1997).