**ChronoSort: Revealing Hidden Dynamics in AlphaFold3 Structure Predictions**

*Matthew Argyle¹, Will Heaps¹, Corbyn Cubalek¹, Spencer Gardiner¹, Dennis Della Corte¹,\**

¹ Department of Physics and Astronomy, Brigham Young University, Provo, Utah

\* corresponding author: [dennis.dellacorte@byu.edu](mailto:dennis.dellacorte@byu.edu)

**Abstract**

Protein function emerges from dynamic conformational changes that occur across multiple timescales, yet current structure prediction methods provide only static snapshots. While AlphaFold3 has revolutionized our understanding of protein structure, the potential for extracting dynamic information from its ensemble predictions has remained largely unexplored. Here, we demonstrate that AlphaFold3 structural ensembles contain substantial dynamic information that correlates remarkably well with molecular dynamics simulation data. We developed ChronoSort, a novel algorithm that organizes static structure predictions into temporally coherent trajectories by minimizing structural differences between neighboring frames. Through systematic analysis of four diverse protein targets—bacterial lipase A, adenylate kinase, HIV-1 protease, and onconase—we show that root mean square fluctuations derived from AlphaFold3 ensembles correlate strongly with those from explicit molecular dynamics simulations (r = 0.74-0.86). Principal component analysis reveals that AlphaFold3 predictions capture the same collective motion patterns observed in molecular dynamics trajectories, with cosine similarities between eigenvectors significantly exceeding those expected from random distributions. Visual comparison through porcupine plots confirms that AlphaFold3 ensembles correctly identify functional motion patterns, including lid domain movements in lipases, domain closure in kinases, and flap dynamics in proteases. The ChronoSort trajectories exhibit realistic RMSD evolution profiles comparable to molecular dynamics simulations, validating the biological relevance of the extracted temporal organization. These findings suggest that the boundary between static and dynamic structural biology may be less rigid than previously assumed, as modern AI-based structure prediction tools encode conformational flexibility information that can be systematically extracted and analyzed. We provide ChronoSort as an open-source software package to enable broad community adoption and further development of these approaches. This work represents a paradigm shift in extracting functional insights from structure prediction tools, potentially democratizing access to protein dynamics information without the computational expense of explicit molecular dynamics simulations. Our results have significant implications for protein engineering, drug discovery, and understanding structure-function relationships, offering a rapid and accessible method for assessing protein flexibility patterns across the proteome.

**Introduction**

Protein function is fundamentally linked to structural dynamics, with conformational changes enabling catalysis, binding, and regulatory processes that are essential to biological systems. The recent revolution in protein structure prediction, culminated by AlphaFold2 and its successor AlphaFold3, has provided unprecedented access to high-quality structural models for hundreds of millions of proteins¹,². AlphaFold3 represents a substantial advancement over previous versions, incorporating a diffusion-based architecture capable of predicting the joint structure of complexes including proteins, nucleic acids, small molecules, ions and modified residues with greatly improved accuracy over many previous specialized tools. However, these predictions fundamentally provide static snapshots of protein structures, while the proteins themselves exist as dynamic ensembles of conformations that fluctuate on timescales ranging from picoseconds to milliseconds³⁻⁵.

The challenge of bridging static structure predictions with dynamic protein behavior has become increasingly prominent in the post-AlphaFold era⁶,⁷. AlphaFold-based methods mainly rely on the co-evolutionary information contained in the sub-MSAs of the input model, so they are limited in capturing protein dynamics. Traditional approaches to understanding protein dynamics rely heavily on molecular dynamics (MD) simulations, nuclear magnetic resonance spectroscopy, and other experimental techniques that can probe conformational flexibility⁸⁻¹⁰. While MD simulations provide detailed atomic-level insights into protein motion, they are computationally expensive and require significant expertise to perform and analyze effectively¹¹,¹². This computational burden limits their application to high-throughput studies or proteome-wide analyses.

Recent efforts have begun to explore whether static structure predictions might encode information about protein dynamics. Innovative approaches for predicting the relative populations of protein conformations using AlphaFold2 have emerged, showing that AI-powered methods can potentially capture aspects of conformational heterogeneity¹³,¹⁴. Additionally, investigations into AlphaFold's ability to predict structural ensembles of disordered proteins have shown promise, particularly for regions with low confidence scores¹⁵,¹⁶. However, these studies have primarily focused on highly flexible regions or intrinsically disordered proteins, leaving open the question of whether ensemble information from structured protein predictions contains meaningful dynamic signatures.

The confidence scores provided by AlphaFold models, particularly the per-residue Local Distance Difference Test (pLDDT) scores, have been interpreted primarily as measures of structural certainty rather than indicators of dynamic behavior¹⁷,¹⁸. While regions with low confidence scores often correspond to flexible loops or disordered regions, the relationship between confidence patterns and the specific modes of protein motion observed in MD simulations has not been systematically explored¹⁹,²⁰. This represents a significant gap in our understanding, as extracting dynamic information directly from static predictions could dramatically accelerate functional annotation and protein design efforts.

Molecular dynamics simulations reveal that protein motion is not random but follows specific patterns that can be characterized through principal component analysis (PCA) and normal mode analysis²¹⁻²³. These collective motions, often described as the first few principal components or normal modes, frequently correspond to functionally relevant conformational changes²⁴,²⁵. The question of whether similar collective motion patterns might be encoded within the structural variations of AlphaFold ensemble predictions has remained largely unexplored, despite its potential significance for structure-based drug design and protein engineering applications²⁶,²⁷.

Furthermore, the temporal organization of protein conformations—how structures transition between different states—represents a critical aspect of protein dynamics that traditional static predictions cannot capture²⁸,²⁹. Methods for organizing structural ensembles into meaningful temporal sequences could provide insights into folding pathways, conformational transitions, and functional mechanisms³⁰,³¹. The development of such approaches would represent a significant advancement in our ability to understand protein dynamics from computational predictions alone.

Here, we present a comprehensive investigation into the dynamic information content of AlphaFold3 ensemble predictions. We introduce ChronoSort, a novel computational approach that organizes static structural predictions into temporal trajectories by minimizing structural differences between neighboring frames, effectively creating pseudo-molecular dynamics trajectories from ensemble predictions. Through systematic comparison with explicit molecular dynamics simulations across diverse protein targets, we demonstrate that AlphaFold3 ensemble predictions contain substantial dynamic information that correlates strongly with experimentally-derived and simulation-based measures of protein flexibility.

**Results and Discussion**

**Conceptual Framework: Extracting Dynamics from Static Predictions**

The central premise of this work is that AlphaFold3 ensemble predictions contain latent dynamic information that can be systematically extracted and compared to traditional molecular dynamics simulations. This concept represents a paradigm shift from viewing structure prediction tools as sources of static snapshots to recognizing them as repositories of conformational heterogeneity that reflects the underlying dynamics of protein systems.

**A diagram of a computer program

AI-generated content may be incorrect.**

**Figure 1.** Graphical abstract illustrating the ChronoSort methodology. AlphaFold3 ensemble predictions (left) are processed through the ChronoSort algorithm to generate temporally ordered trajectories (center), enabling direct comparison with molecular dynamics simulations (right) for extraction of dynamic properties including RMSF and principal component modes.

Our approach leverages the observation that modern deep learning structure prediction models, while not explicitly trained on dynamic data, encode information about conformational flexibility through their sampling of sequence-structure relationships during training. Recent studies have suggested that AlphaFold2 models can indicate both structure and dynamics, with confidence scores correlating with protein flexibility. However, the systematic extraction and validation of dynamic information from AlphaFold3 ensembles has not been thoroughly explored, particularly through direct comparison with molecular dynamics simulations.

**Initial Discovery: Dynamic Signatures in NanoLuc Luciferase**

Our investigation began with the serendipitous observation of strong correlations between AlphaFold3 structural variations and molecular dynamics-derived flexibility patterns in NanoLuc luciferase, a 19 kDa enzyme that we were studying for protein engineering applications. The correlation between AlphaFold3-derived and MD-derived root mean square fluctuations was remarkably high (r = 0.82), suggesting that the structural ensemble predictions contained meaningful dynamic information.



**Figure 2.** Root mean square fluctuation comparison for NanoLuc luciferase. Per-residue RMSF values calculated from molecular dynamics simulations (blue line, averaged over three 250 ns replicates) show strong correlation with RMSF derived from AlphaFold3 structural ensemble (black line, n=100 structures). The correlation coefficient (r = 0.82, p < 0.001) indicates that AlphaFold3 predictions capture the major flexibility patterns observed in explicit molecular dynamics simulations. Secondary structure elements are indicated above the plot, with α-helices shown as cylinders and β-strands as arrows.

This initial finding was particularly striking because AlphaFold2 was neither designed nor trained to predict protein dynamics, yet the structural variations in its predictions appeared to encode information about conformational flexibility. The strong correlation observed in NanoLuc prompted us to investigate whether this phenomenon was generalizable across different protein families and structural motifs.

**Systematic Validation Across Diverse Protein Targets**

To assess the generalizability of our findings, we selected four additional protein targets representing diverse structural families and functional classes: bacterial lipase A, adenylate kinase, HIV-1 protease, and onconase. These proteins span different fold families and exhibit distinct types of functional motions, from lid movements in lipases to domain closure in kinases and flap dynamics in proteases.

**Figure 3.** Structural overview of protein targets and their predominant motions. 2×2 panel showing the four protein targets: (A) Bacterial lipase A (PDB: 1ISP) with lid domain motion indicated by red arrows; (B) Adenylate kinase (PDB: 4AKE) showing the characteristic domain closure motion; (C) HIV-1 protease (PDB: 1HPV) highlighting flap dynamics; (D) Onconase (PDB: 1ONC) with loop flexibility patterns. Red arrows indicate the primary axes of motion identified through principal component analysis of both molecular dynamics simulations and AlphaFold3 ensembles.

The selection of these targets was strategic, as each represents well-characterized systems with established functional motions that have been extensively studied through both experimental and computational approaches. Principal component analysis has become commonplace for revealing the most important motions in proteins, providing a robust framework for comparing the dynamic information extracted from different sources.

**ChronoSort Algorithm Validates Temporal Organization**

The ChronoSort algorithm successfully organized AlphaFold3 ensemble predictions into temporally coherent trajectories that exhibited RMSD profiles remarkably similar to those observed in molecular dynamics simulations. The algorithm's ability to extract meaningful temporal ordering from static predictions suggests that the structural variations in AlphaFold3 ensembles are not random but reflect biologically relevant conformational transitions.

**Figure 4.** RMSD analysis of ChronoSort trajectories compared to molecular dynamics simulations. 2×2 panel showing RMSD evolution for all four protein targets: (A) Lipase A, (B) Adenylate kinase, (C) HIV-1 protease, (D) Onconase. For each system, molecular dynamics trajectories (three replicates, colored lines) are compared with the ChronoSort trajectory derived from AlphaFold3 ensemble predictions (black line). The similar RMSD magnitudes and profiles indicate that ChronoSort captures realistic conformational excursions comparable to those observed in explicit molecular dynamics simulations.

The RMSD profiles generated by ChronoSort showed several important characteristics that validate the biological relevance of the extracted temporal organization. First, the magnitude of structural deviations was comparable to those observed in molecular dynamics simulations, suggesting that AlphaFold3 ensembles span a similar conformational space to that sampled by explicit dynamics. Second, the temporal evolution showed realistic patterns of structural relaxation and fluctuation, rather than monotonic divergence that might be expected from random structural ordering.

**Quantitative Validation of Flexibility Patterns**

The most compelling evidence for the biological relevance of dynamic information in AlphaFold3 ensembles came from systematic comparison of root mean square fluctuation patterns. Across all four protein targets, we observed consistently high correlations between AlphaFold3-derived and MD-derived RMSF values, with correlation coefficients ranging from 0.74 to 0.86.

**Figure 5.** Root mean square fluctuation correlations across protein targets. 2×2 panel comparing MD-derived RMSF (blue lines, averaged over three replicates) with AlphaFold3-derived RMSF (black lines) for: (A) Lipase A (r = 0.78), (B) Adenylate kinase (r = 0.81), (C) HIV-1 protease (r = 0.74), (D) Onconase (r = 0.86). Shaded regions indicate standard deviation across MD replicates. The consistently high correlations demonstrate that AlphaFold3 structural ensembles capture per-residue flexibility patterns that closely match those observed in molecular dynamics simulations.

These correlations are particularly significant when considered in the context of the known challenges with reproducibility of principal component modes between equivalent molecular dynamics simulations. The fact that AlphaFold3-derived flexibility patterns show such consistent agreement with MD-derived patterns suggests that the static predictions are capturing fundamental aspects of protein dynamics that are robust across different simulation conditions and protocols.

The biological significance of these correlations is further supported by their consistency with known functional elements. For example, the highest RMSF values in adenylate kinase correspond to the lid domain that undergoes large-scale conformational changes during the catalytic cycle, while in HIV-1 protease, the peak flexibility occurs in the flap regions that are critical for substrate binding and product release.

**Principal Component Analysis Reveals Shared Motion Patterns**

To investigate whether AlphaFold3 ensembles capture not only the magnitude of flexibility but also the directional patterns of collective motion, we performed principal component analysis on both MD trajectories and AlphaFold3 structural ensembles. The results revealed striking similarities in the eigenvector patterns, with cosine similarity values significantly higher than those expected from random vectors of equivalent dimensionality.

**Figure 6.** Eigenvector similarity analysis across protein targets. Cosine similarity matrices comparing principal component eigenvectors derived from molecular dynamics simulations (MD1, MD2, MD3 representing three replicates) and AlphaFold3 ensembles (AF3). Color scale indicates cosine similarity values from 0 (orthogonal) to 1 (identical). White boxes indicate comparisons with randomly oriented vectors (control). The consistently high similarities between MD and AF3 eigenvectors, particularly for the first few principal components, demonstrate that AlphaFold3 ensembles capture the dominant collective motion patterns observed in molecular dynamics simulations.

The eigenvector similarity analysis provided several key insights. First, the dominant principal components (PC1-PC3) showed the highest similarities between MD and AlphaFold3 analyses, consistent with the expectation that the most collective motions would be most robustly captured across different analysis methods. Second, the similarities between AlphaFold3 and MD eigenvectors were consistently higher than the similarities between independent MD simulation replicates, suggesting that the static predictions may actually provide a more reproducible representation of collective motions than individual MD trajectories.

This finding has important implications for drug discovery applications where principal component analysis is used to understand protein-ligand interactions and conformational changes. The ability to extract reliable information about collective motions directly from structure predictions could significantly accelerate the identification of druggable conformational states and allosteric binding sites.

**Visual Validation Through Porcupine Plot Analysis**

The most intuitive validation of our findings came through direct visual comparison of eigenvector patterns using porcupine plots, which display the direction and magnitude of atomic displacements along each principal component. The visual similarity between MD-derived and AlphaFold3-derived motion patterns was striking, with nearly identical directional preferences and relative magnitudes across all four protein targets.

**Figure 7.** Porcupine plot comparison of principal motion patterns. 2×2 panel showing the dominant eigenvector (PC1) for each protein target derived from molecular dynamics simulations (blue vectors) and AlphaFold3 ensembles (red vectors): (A) Lipase A showing lid domain motion, (B) Adenylate kinase displaying domain closure, (C) HIV-1 protease highlighting flap dynamics, (D) Onconase showing loop flexibility. The near-perfect overlap of vector directions demonstrates that AlphaFold3 ensembles capture the same collective motion patterns identified in molecular dynamics simulations.

The porcupine plots revealed that AlphaFold3 ensembles correctly identify the functional motion patterns that are characteristic of each protein family. In lipase A, both approaches identified the opening and closing motion of the lid domain that controls substrate access to the active site. For adenylate kinase, the dominant motion corresponded to the well-characterized domain closure that brings the ATP and AMP binding sites together during the catalytic cycle. In HIV-1 protease, the primary motion involved the flap regions that gate substrate binding, while in onconase, the dominant patterns reflected the flexibility of surface loops involved in RNA binding.

**Implications for Structure-Function Relationships**

The strong correlations observed across multiple metrics and protein systems suggest that AlphaFold3 ensemble predictions contain a wealth of dynamic information that has been largely overlooked in structure-function analyses. Recent work has begun to integrate AlphaFold confidence scores into enhanced protein flexibility simulations, but our results indicate that the full structural ensembles provide far richer dynamic information than confidence scores alone.

This finding has profound implications for understanding protein function. The traditional paradigm of using single static structures to infer function is increasingly recognized as insufficient, particularly for understanding allosteric regulation, conformational selection mechanisms, and the effects of mutations on protein dynamics. Our results suggest that much of this missing dynamic information can be extracted directly from structure prediction ensembles, potentially democratizing access to dynamic insights for the broader scientific community.

The ability to rapidly assess protein dynamics from structure predictions could transform several areas of biological research. In protein engineering, understanding flexibility patterns is crucial for designing mutations that preserve or modify specific functional motions. In drug discovery, knowledge of conformational dynamics is essential for identifying allosteric binding sites and understanding binding mechanisms. In structural biology, dynamic information helps interpret the functional significance of structural features and guides experimental design.

**Methodological Considerations and Limitations**

While our results are highly encouraging, several important limitations must be acknowledged. First, our analysis focused on relatively small, well-folded proteins, and the generalizability to larger, multidomain proteins or membrane proteins remains to be established. Recent work suggests that AlphaFold-Multimer can capture dynamics of intrinsically disordered regions, but systematic validation across diverse protein classes is needed.

Second, the ChronoSort algorithm, while effective at organizing structural ensembles into meaningful trajectories, represents only one possible approach to temporal organization. Alternative methods based on transition state theory or Markov state modeling might provide different insights into the conformational transitions encoded in static predictions. The current implementation also assumes that the ensemble predictions sample conformations along realistic transition pathways, which may not always be the case.

Third, our analysis was limited to backbone dynamics, while many functional motions involve side chain rearrangements that may not be well-captured in Cα-based analyses. Future extensions of this work should incorporate all-atom analyses and validation against experimental measures of side chain dynamics, such as NMR relaxation data.

Finally, while the correlations between AlphaFold3 and MD dynamics are impressive, they are not perfect, and understanding the sources of discrepancy could provide important insights into both the capabilities and limitations of structure prediction methods. Some discrepancies may reflect genuine sampling limitations in the MD simulations, while others may indicate fundamental limitations in the dynamic information encoded in structure predictions.

**Future Directions and Broader Impact**

The demonstration that AlphaFold3 ensemble predictions contain meaningful dynamic information opens several exciting avenues for future research. Extensions to protein-protein interaction dynamics, allosteric network analysis, and the effects of post-translational modifications on protein flexibility all represent high-impact applications of this methodology. Integration with experimental data, particularly NMR relaxation measurements and hydrogen-deuterium exchange experiments, could provide additional validation and refinement of the approach.

From a broader perspective, this work suggests that the artificial intelligence revolution in structural biology may have implications beyond static structure prediction. As AI models become increasingly sophisticated and are trained on larger and more diverse datasets, they may naturally evolve to encode more complex aspects of protein behavior, including dynamics, thermodynamics, and even functional mechanisms. Understanding how to extract and interpret this latent information will be crucial for maximizing the scientific impact of these powerful computational tools.

The development of the ChronoSort methodology and its implementation as an open-source software package represents an important step toward democratizing access to protein dynamics information. By enabling researchers without extensive computational resources or molecular dynamics expertise to rapidly assess protein flexibility patterns, this approach could accelerate discovery across numerous areas of biological research and therapeutic development.

**Methods**

**Protein Target Preparation**

Four diverse protein targets were selected to represent different structural families and functional classes: NanoLuc luciferase (PDB: 5IBO), bacterial lipase A from *Bacillus subtilis* (PDB: 1ISP), adenylate kinase from *Escherichia coli* (PDB: 4AKE), HIV-1 protease (PDB: 1HPV), and onconase from *Rana pipiens* (PDB: 1ONC). Amino acid sequences were retrieved directly from the Protein Data Bank entries, with signal peptides and non-canonical residues removed where appropriate to ensure compatibility with both AlphaFold3 prediction protocols and molecular dynamics simulation force fields.

**AlphaFold3 Ensemble Generation**

AlphaFold3 predictions were generated using the official DeepMind server with default parameters. To create structural ensembles, we performed 100 independent predictions for each target protein using different random number seeds, following established protocols for ensemble generation from deep learning structure prediction models. Each prediction was carried out with identical input sequences but unique initialization states to sample the conformational heterogeneity encoded in the AlphaFold3 model. All predicted structures were downloaded in PDB format and subjected to quality control checks to ensure proper chain assignment and coordinate integrity.

**Molecular Dynamics Simulations**

All molecular dynamics simulations were performed using GROMACS 2022.4 with the CHARMM36m force field. Initial structures were prepared by taking the first model from each target's crystallographic structure, with missing hydrogen atoms added using the pdb2gmx tool. Each protein was solvated in a rectangular box with TIP3P water molecules, maintaining a minimum distance of 1.0 nm between the protein and box boundaries. The systems were neutralized by adding appropriate counter-ions (Na⁺ and Cl⁻) to achieve physiological ionic strength (0.15 M NaCl).

Energy minimization was performed using the steepest descent algorithm until the maximum force fell below 1000 kJ mol⁻¹ nm⁻¹. Subsequently, the systems underwent NVT equilibration at 300 K for 100 ps using the V-rescale thermostat, followed by NPT equilibration at 300 K and 1 bar for 100 ps using the Parrinello-Rahman barostat. Production runs were conducted for 250 ns in triplicate for each target, with coordinates saved every 10 ps, resulting in 25,000 frames per trajectory. All bond lengths involving hydrogen atoms were constrained using the LINCS algorithm, allowing for a 2 fs integration time step.

**ChronoSort Algorithm**

The ChronoSort algorithm was developed to organize AlphaFold3 ensemble predictions into temporally coherent trajectories by minimizing cumulative structural differences between consecutive frames. First, pairwise Cα root-mean-square deviation (RMSD) values were calculated between all predicted structures in the ensemble after optimal superposition using the Kabsch algorithm. This generated a symmetric 100×100 distance matrix D, where D[i,j] represents the RMSD between structures i and j.

The optimal temporal ordering was determined by solving a variant of the traveling salesman problem to find the path through the distance matrix that minimizes the total RMSD between consecutive structures. We employed a greedy nearest-neighbor heuristic followed by 2-opt optimization to efficiently identify near-optimal solutions. The algorithm begins with a random structure and iteratively selects the nearest unvisited structure until all structures are incorporated into the temporal sequence. The resulting ChronoSort trajectory was then analyzed using identical protocols as the molecular dynamics simulations.

**RMSD Analysis**

Root-mean-square deviation calculations were performed on Cα atoms after least-squares superposition to the initial frame of each trajectory. For molecular dynamics trajectories, RMSD was calculated as a function of simulation time. For ChronoSort trajectories, frames were assigned artificial time points by dividing the 250 ns simulation time window by 100 frames, yielding a temporal resolution of 2.5 ns per frame to enable direct comparison with MD trajectories. RMSD analysis was performed using GROMACS analysis tools, with results visualized using matplotlib and compared across the three MD replicates and the single ChroSort trajectory for each target.

**Root Mean Square Fluctuation Calculations**

Root-mean-square fluctuation (RMSF) values were calculated for Cα atoms after center-of-mass alignment of all trajectory frames. For molecular dynamics simulations, RMSF was computed using the standard definition:

RMSF\_i = √(⟨(r\_i - ⟨r\_i⟩)²⟩)

where r\_i is the position vector of atom i and ⟨r\_i⟩ is the time-averaged position. For AlphaFold3 ensembles, RMSF was calculated as the root-mean-square deviation of each Cα atom from its mean predicted position across the ensemble. RMSF values were computed using GROMACS gmx rmsf tool for MD trajectories and custom Python scripts for AlphaFold3 ensembles, with results compared using Pearson correlation coefficients.

**Principal Component Analysis**

Principal component analysis was performed on Cα coordinates after removal of translational and rotational motions. Covariance matrices were constructed from the coordinate fluctuations, and eigenvalue decomposition was performed to obtain principal components (eigenvectors) and their associated eigenvalues. The number of eigenvectors retained for analysis was determined by the cumulative eigenvalue criterion, typically encompassing 80% of the total variance, corresponding to approximately the first 9 eigenvectors for most systems.

Cosine similarity between eigenvectors was calculated using the dot product of normalized vectors. For validation, we first computed pairwise cosine similarities between eigenvectors from independent MD simulation replicates to establish the baseline reproducibility of principal modes. Subsequently, cosine similarities were calculated between MD-derived and AlphaFold3-derived eigenvectors. Statistical significance was assessed by comparing observed similarities to null distributions generated from randomly oriented vectors of equivalent dimensionality.

**Eigenvector Visualization**

Principal component eigenvectors were visualized using porcupine plots (also known as hedgehog plots), an established method for displaying collective motions in proteins. These plots display the protein backbone as a ribbon representation with vectors emanating from each Cα position, indicating the direction and magnitude of atomic displacements along each principal component. The most similar eigenvector pairs between MD and AlphaFold3 analyses, as determined by cosine similarity calculations, were selected for comparative visualization. Porcupine plots were generated using PyMOL with custom scripts to overlay MD and AlphaFold3 eigenvectors in different colors, allowing direct visual comparison of the dominant motion patterns captured by each approach.

**Statistical Analysis**

All correlation analyses were performed using Pearson correlation coefficients, with significance assessed using two-tailed t-tests. Confidence intervals were calculated using bootstrap resampling (n=1000). For eigenvector similarity analysis, null distributions were generated by computing cosine similarities between 10,000 pairs of randomly oriented unit vectors in 3N-dimensional space (where N is the number of Cα atoms). P-values were calculated as the fraction of random similarities exceeding the observed values. All statistical analyses were performed using SciPy and NumPy in Python 3.9.

**References**

1. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589 (2021).
2. Abramson, J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature 630, 493–500 (2024).
3. Henzler-Wildman, K. & Kern, D. Dynamic personalities of proteins. Nature 450, 964–972 (2007).
4. Boehr, D. D., Nussinov, R. & Wright, P. E. The role of dynamic conformational ensembles in biomolecular recognition. Nat. Chem. Biol. 5, 789–796 (2009).
5. Motlagh, H. N., Wrabl, J. O., Li, J. & Hilser, V. J. The ensemble nature of allostery. Nature 508, 331–339 (2014).
6. Hollingsworth, S. A. & Dror, R. O. Molecular dynamics simulation for all. Neuron 99, 1129–1143 (2018).
7. Shaw, D. E. et al. Anton 2: raising the bar for performance and programmability in a special-purpose molecular dynamics supercomputer. Proceedings of SC14 41–53 (2014).
8. Palmer, A. G. NMR characterization of the dynamics of biomacromolecules. Chem. Rev. 104, 3623–3640 (2004).
9. Kay, L. E. NMR studies of protein structure and dynamics. J. Magn. Reson. 173, 193–207 (2005).
10. Mittermaier, A. K. & Kay, L. E. Observing biological dynamics at atomic resolution using NMR. Trends Biochem. Sci. 34, 601–611 (2009).
11. Karplus, M. & McCammon, J. A. Molecular dynamics simulations of biomolecules. Nat. Struct. Biol. 9, 646–652 (2002).
12. Hospital, A., Goñi, J. R., Orozco, M. & Gelpí, J. L. Molecular dynamics simulations: advances and applications. Adv. Appl. Bioinform. Chem. 8, 37–47 (2015).
13. Wayment-Steele, H. K. et al. High-throughput prediction of protein conformational distributions with subsampled AlphaFold2. Nat. Commun. 15, 2464 (2024).
14. Monteiro da Silva, G. et al. Machine learning/molecular dynamic protein structure prediction approach to investigate the protein conformational ensemble. Sci. Rep. 12, 10018 (2022).
15. Ruff, K. M. & Pappu, R. V. AlphaFold and implications for intrinsically disordered proteins. J. Mol. Biol. 433, 167208 (2021).
16. Alderson, T. R. & Pritišanac, I. AlphaFold prediction of structural ensembles of disordered proteins. Nat. Commun. 16, 1096 (2025).
17. Tunyasuvunakool, K. et al. Highly accurate protein structure prediction for the human proteome. Nature 596, 590–596 (2021).
18. Varadi, M. et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res. 50, D439–D444 (2022).
19. Mariani, V., Biasini, M., Barbato, A. & Schwede, T. lDDT: a local superposition-free score for comparing protein structures. Bioinformatics 29, 2722–2728 (2013).
20. Pereira, J. et al. High-accuracy protein structure prediction in CASP14. Proteins 89, 1687–1699 (2021).
21. Amadei, A., Linssen, A. B. M. & Berendsen, H. J. C. Essential dynamics of proteins. Proteins 17, 412–425 (1993).
22. Garcia, A. E. Large-amplitude nonlinear motions in proteins. Phys. Rev. Lett. 68, 2696–2699 (1992).
23. Ichiye, T. & Karplus, M. Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. Proteins 11, 205–217 (1991).
24. Bahar, I., Lezon, T. R., Yang, L.-W. & Eyal, E. Global dynamics of proteins: bridging between structure and function. Annu. Rev. Biophys. 39, 23–42 (2010).
25. Zheng, W. & Brooks, B. R. Identification of dynamical correlation networks in proteins. J. Chem. Phys. 143, 044108 (2015).
26. Dror, R. O., Dirks, R. M., Grossman, J. P., Xu, H. & Shaw, D. E. Biomolecular simulation: a computational microscope for molecular biology. Annu. Rev. Biophys. 41, 429–452 (2012).
27. De Vivo, M., Masetti, M., Bottegoni, G. & Cavalli, A. Role of molecular dynamics and related methods in drug discovery. J. Med. Chem. 59, 4035–4061 (2016).
28. Bowman, G. R., Pande, V. S. & Noé, F. An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation (Springer, 2014).
29. Plattner, N. & Noé, F. Protein conformational plasticity and complex ligand-binding kinetics explored by atomistic simulations and Markov models. Nat. Commun. 6, 7653 (2015).
30. Husic, B. E. & Pande, V. S. Markov state models: from an art to a science. J. Am. Chem. Soc. 140, 2386–2396 (2018).
31. Scherer, M. K. et al. PyEMMA 2: a software package for estimation, validation, and analysis of Markov models. J. Chem. Theory Comput. 11, 5525–5542 (2015).