ABSTRACT

Title of Dissertation: BRIDGING ARTIFICIAL INTELLIGENCE AND THERMODYNAMICS FOR ADVANCING CANCER IMMUNOTHERAPY AND DRUG DISCOVERY

Akashnathan Aranganathan
Doctor of Philosophy, 2026

Dissertation Directed by: Professor Pratyush Tiwary
Department of Chemistry and Biochemistry, UMD
Institute for Physical Science and Technology, UMD
University of Maryland Institute for Health Computing

Protein function is often discussed from the perspective of a single representative structure. This static view has led to the development of experimental structure determination techniques, including X-ray crystallography, NMR spectroscopy, and cryo-EM, and has also shaped the development of the highly accurate structure-prediction methods such as AlphaFold2 (AF2). However, proteins are dynamic molecules, and their biological functions are governed by conformational ensembles composed of interconverting metastable states. This ensemble perspective is especially important in drug discovery and immunotherapy, where ligand binding, receptor engagement, and downstream signaling can depend on those rare or transient conformations that are not apparent from a single structure. Therefore, a key challenge is to develop computational methods that can start from sequence, generate relevant conformational diversity, assign thermodynamic weights to the generated states, and connect these ensembles to biological function.

To address this, in this thesis, I develop and apply computational approaches that bridge Artificial Intelligence(AI)-based structure prediction with physics-based sampling and reweighting. First, I investigated how stochasticity can be introduced into AF2 through reduced multiple sequence alignment(MSA) approach, showing that AF2-generated ensembles can provide useful structural hypotheses but do not directly encode thermodynamic or Boltzmann weights. To this end, I developed AlphaFold2-RAVE (AF2-RAVE), a framework that combines AF2-derived structural diversity with AI-augmented enhanced sampling molecular dynamics to obtain thermodynamically ranked conformational ensembles. I apply AF2-RAVE to systems involving rotameric metastability, missense mutations, G-protein-coupled receptor motions, kinase DFG-loop and activation-loop conformational changes, and calcium-bound S100 homodimers. Together, these applications demonstrate that the post hoc integration of physics-based methods with generative-AI methods provides a framework that leverages the strengths of both approaches. Specifically, I show that AF2-RAVE type methods can use the conformational selection hypothesis to generate druggable metastable states for structure-based drug discovery in a computationally efficient manner.

I then extend the ensemble perspective to immunotherapy. I apply AF2-RAVE-based ideas to evaluate T-cell receptor - peptideMHC complexes generated by cofolding models. I also extend this approach to model the synaptic cleft formed by CAR T cells and develop a theory for CAR T-cell responsiveness based on the canonical T-cell immunological synapse. Finally, I develop a protocol that fits AI-generated structural ensembles to the Langevin Equation for Protein Dynamics (LE4PD), allowing residue-level relaxation timescales to be inferred from AF2 and other generative AI methods.

Overall, this thesis develops a “sequence–ensemble–function” framework for biomolecular

modeling. By combining generative AI, molecular dynamics, enhanced sampling, and coarse-grained dynamical models, the work presented here moves beyond static structure prediction toward a more complete description of protein function. This framework led to the development and application of several methods, including reduced-MSA approach for AF2, AF2-RAVE and its variants, CAR-BOOST, and an LE4PD-based approach for assigning dynamical information to AI-generated structural ensembles.

BRIDGING ARTIFICIAL INTELLIGENCE AND THERMODYNAMICS FOR ADVANCING CANCER IMMUNOTHERAPY AND DRUG DISCOVERY

by

Akashnathan Aranganathan

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2026

Advisory Committee:
Professor Pratyush Tiwary, Chair/Advisor
Dr. Grégoire Altan-Bonnet
Professor Jeffery B. Klauda
Professor Brian Pierce
Professor Sergei Sukharev

© Copyright by
Akashnathan Aranganathan
2026

Dedication

To my family – Anbu Selvi A., Aranganathan V.,

and Abhinayaa A. – and to the

loving memory of my grandmother, Sarada.

Thank you all for everything.

Acknowledgments

"Karmanyevadhikaraste Ma Phaleshu Kadachana,

Ma Karmaphalaheturbhurma Te Sangostvakarmani"

– Bhagavad Gita, Chapter 2, Verse 47

"... The woods are lovely, dark and deep,

But I have promises to keep,

And miles to go before I sleep,

And miles to go before I sleep."

– Stopping by Woods on a Snowy Evening, Robert Frost

I am deeply grateful to my advisor, Dr. Pratyush Tiwary, who has supported me both academically and personally throughout my Ph.D. From the first few weeks of my rotation, he recognized my strengths and interests and guided me toward projects that helped me grow with confidence. He gave me the freedom to explore new ideas, which led to Chapter 9 and part of Chapter 2 in this thesis, while also creating opportunities for me to mentor, write proposals, review manuscripts, and give talks. He also supported me strongly during my postdoc applications, both practically and emotionally. He has often reminded me of the first quote whenever I felt stressed about project or application process, and over time these two lines have become something I try to follow myself. I am extremely grateful for the five years I spent in the Tiwary

lab and for everything I learned from him. I really enjoyed working in the Tiwary lab under his guidance.

During my Ph.D., I was co-advised by Dr. Grégoire Altan-Bonnet starting from my second year via the NCI-UMD fellowship. He introduced me to many interesting problems in immunology where biophysical intuition could be useful, and he gave me confidence by trusting my models and supporting me as I presented my work to researchers from different areas of biology and immunology. I am very grateful to him for his patience, guidance, and trust in my work. I also had the opportunity to work with Dr. Naomi Taylor, who was very supportive and asked thoughtful questions that helped me think more broadly about what could be modeled. I am thankful to Dr. Grégoire Altan-Bonnet, Dr. Naomi Taylor, and my advisor Dr. Tiwary for this opportunity, where we were able to take a computational and theoretical model toward testing in mice. I am also grateful to the NCI-UMD fellowship program for making this collaboration possible.

I thank my committee members - Dr. Jeff Klauda, Dr. Sergei Sukharev, Dr. Brian Pierce - for their thoughtful and insightful feedback on my science and critical reading of my thesis. I thank Dr. Brian Pierce for suggesting TCR3d resource and providing data that helped strengthen Chapter 7 of this thesis. I also thank him for organizing a virtual computational biology journal club, which helped me stay connected to recent developments in this fast-growing field. I am also thankful to Dr. Klauda and Dr. Upadhyaya for always being accessible and patient with my many questions about the biophysics program. I was very fortunate to be part of the UMD Graduate School, where I took courses from amazing professors - John Weeks, Sergei Sukharev, Garegin A. Papoian, Maria Cameron, Soheil Feizi, and Christopher Jarzynski. Each of them taught their fields with great passion, and I enjoyed every moment in their classes. Special thanks to Christopher Jarzynski for allowing me to sit through some of his lectures in his nonequilibrium

statistical physics course a second time, which helped me understand the subject much better. Based on my experience at UMD, I hope to carry forward this passion for teaching, learning and asking curious questions wherever I go next. I would also like to thank the administrative people including Souad Nejjar, Jessica Crosby, Stacey L. Bremerman, Corrine Harvey (from NIH) for handling my concerns with care.

I am extremely grateful to the members of the Tiwary lab. I thank my mentors Dr. Bodhi Vani, Dr. Eric Beyerle, Dr. Zack Smith, and Dr. Shams Mehdi. Special thanks to Bodhi for mentoring me during my first two years, always being ready to answer my questions, and organizing a machine learning journal club with me so that we could learn the concepts from scratch. I thank Lukas, who joined the lab around the same time as me, worked through courses and homeworks with me, and helped me learn diffusion models. I also thank the other graduate students in the Tiwary lab, Yihang, Dedi, Connor, Luke, Vanessa, Suemin, Anjali, Ricky, Venkata, Preston, and Gokul for fun discussions about ongoing projects and for spending time with me outside work. I thank Anjali also for joining me in our small informal journal club, where we worked closely through papers. I also thank the postdocs in the Tiwary lab, Ruiyu, Xinyu, Da, Yunrui, Azamat, and Hale, from whom I learned about different subfields and different ways of approaching problems. I learned a lot from Xinyu, Da, and Yunrui, especially in the later stages of my PhD, and I also thank them for including me in our trips to the Institute for Health Computing. I also thank Xinyu, Azamat, and Mariadelia for allowing me to mentor them, an interesting experience from which I learned a lot. From Dr. Altan-Bonnet's group, I would like to give a huge thanks to Justin. We had a very productive collaboration, which led to Chapter 8 of this thesis.

I would also like to thank my undergraduate advisor, Dr. Athi N. Naganathan, who guided me and set me on this path.

I thank my past and present roommates, Shailesh, Ashwath, Ananth, Arunabh, Matthew, and Wei. I also thank my cohort and the friends I made at UMD, Vishavdeep, Anthony, Frank, Anna, Elissa, Smrithan, Sahana, Priyanka, Greeshma, Sashank, and Samyak. Together with my lab mates, they all made my PhD at UMD an enjoyable and memorable experience. I am especially grateful to my current roommates, Arunabh and Matthew, for making me feel at home. Whether we were attending concerts, going to the movies, or simply watching television series together at home, spending time with them helped me unwind from work. I would also like to thank my friends from undergraduate years, Kowshika and Pradeep. They have been like family to me since my first year of my undergraduate and have always supported me, irrespective of where we are on the globe.

Finally, and most importantly, I thank my family, Anbu Selvi, Aranganathan, and Abhinayaa. I could not have done any of this without your unconditional love and support. Thank you for everything!

Table of Contents

Dedication ii

Acknowledgments iii

Table of Contents vii

List of Tables xiii

List of Figures xiv

List of Abbreviations xxxiii

List of Publications xxxiv

Chapter 1: Introduction 1
- 1.1 Conceptual Background and History of Protein Biophysics . . . 2
  - 1.1.1 Proteins as Heteropolymers and Protein Folding Problem . . . 2
  - 1.1.2 Metastable States and the "Sequence-Ensemble-Function" Paradigm . . . 5
  - 1.1.3 Conformational Selection Hypothesis . . . 6
- 1.2 Methods . . . 8
  - 1.2.1 Static protein structure prediction . . . 8
  - 1.2.2 Protein dynamics . . . 11
- 1.3 Thesis descriptor and biological motivation . . . 22

Chapter 2: Understanding and Adding Stochasticity to AlphaFold2 24
- 2.1 From Potts Models to Masked Language Models to AlphaFold2 . . . 25
  - 2.1.1 AlphaFold2: Everything, Everywhere, All at Once . . . 29
  - 2.1.2 Looking into AF2 . . . 30
- 2.2 How much coevolution is required? . . . 32
- 2.3 Block Perturbation of the MSA . . . 34
- 2.4 Reduced MSA approach to AF2 (rMSA-AF2) . . . 35
- 2.5 Other related works . . . 39
- 2.6 Chapter Summary . . . 40

Chapter 3: AlphaFold2-RAVE: From sequence to Boltzmann ranking 42
- 3.1 Results . . . 44
  - 3.1.1 Establishing the AlphaFold2-RAVE Protocol . . . 44

3.1.2 Rotameric Metastability in Cold Shock Protein . . . 45
3.1.3 Disorder in the Mutant Form of Ubiquitin-Associated Protein . . . 46
3.1.4 Helical Motion in a G-Protein-Coupled Receptor . . . 47
3.2 Conclusion . . . 48

Chapter 4: Exploring kinase Asp-Phe-Gly (DFG) loop conformational stability with AlphaFold2-RAVE 50
4.1 Results and Discussion . . . 52
4.1.1 Learning a Bias Potential from AF2-RAVE on WT DDR1 . . . 54
4.1.2 Biased Dynamics on DDR1 Mutants . . . 57
4.2 Conclusion . . . 59

Chapter 5: Empowering AlphaFold2 for protein conformation selective drug discovery with AlphaFold2-RAVE 61
5.1 Results and discussion . . . 65
5.1.1 AF2 structures fail to dock with type II kinase inhibitors . . . 65
5.1.2 Holo-like metastable structures may be present among decoys generated from rMSA AF2 . . . 68
5.1.3 AF2-RAVE on DDR1 enriches holo-like classical DFG-out structures in rMSA AF2 decoys . . . 71
5.1.4 Transferable learning of holo-like structure for Abl1 and Src kinases with AF2-RAVE from DDR1 templates . . . 75
5.1.5 Discussion . . . 76
5.2 Conclusion . . . 79

Chapter 6: Hierarchical AF2-RAVE for Timescale-Resolved Holo-like Structure Generation 81
6.1 Methods . . . 83
6.1.1 Hierarchical AF2-RAVE funnel for multiconfiguration holo-like receptor generation . . . 83
6.1.2 Docking validation as a downstream test . . . 90
6.2 Results . . . 91
6.2.1 Global SPIB captured the backbone breathing of the hydrophobic pocket in Ca^{2+}-S100 . . . 91
6.2.2 Local SPIB captured metastable side-chain packing states, covering both the AF2-predicted apo and known holo conformations for the Ca^{2+}-S100 proteins . . . 92
6.2.3 Retrospective docking validation on AF2-RAVE structure set of Ca^{2+}-S100B . . . 95
6.2.4 Retrospective enrichment validation on AF2-RAVE structure set of Ca^{2+}-S100B . . . 97
6.3 Discussion . . . 98

Chapter 7: Extending AF2-RAVE for improved evaluation of T Cell Receptor - Peptide Major Histocompatibility Complexes from Multimer Models 102

7.1 TCR peptide MHC recognition . . . 103
7.2 Related Work . . . 104
7.3 Antigen Recognition and the HHAT Wild-Type/Neoantigen System . . . 105
7.4 Results . . . 106
7.4.1 Metadynamics ranks AlphaFold-Multimer binding poses . . . 106
7.4.2 AF2-RAVE for TCR-pMHC Complexes . . . 109
7.4.3 Structural validation of AF2-RAVE basins . . . 110
7.5 Conclusion and Future Work . . . 111

Chapter 8: CAR-BOOSTing T-Cell Responsiveness through Synaptic Cleft Engineering 113
8.1 Introduction . . . 113
8.2 Results . . . 114
8.2.1 The hinge region of the CD19 CAR construct regulates CAR T-cell responsiveness to $CD19^{low}$ leukemia . . . 114
8.2.2 Optimizing the CD8α-derived hinge for increased responsiveness of the tisa-cel-like CAR to $CD19^{low}$ leukemia . . . 116
8.2.3 Modeling CAR-generated synaptic clefts using an AlphaFold2RAVE (AF2RAVE)-based model. . . . 119
8.2.4 CAR T-cell responsiveness to a $CD19^{+}$ tumor is dependent on the intermembrane distance distribution of the CAR IS. . . . 121
8.2.5 CAR-BOOSTing model predicts maximally responsive hinges across multiple CAR targets . . . 123
8.3 Discussion . . . 125

Chapter 9: Applied causality to infer protein dynamics and kinetics 139
9.1 Results . . . 142
9.1.1 Overview: A Map from Sequence to Kinetics . . . 142
9.1.2 The Timescales of AlphaFold2-generated Ensembles are Inversely Correlated with MSA Depth . . . 142
9.1.3 MSA Depth is Inversely Correlated with the Span of the Slowest Collective Motions Generated by the AF2 Conformational Ensemble . . . 145
9.1.4 Quantification of the Timescales Spanned by AF2 Evaluated at Different MSA Depths . . . 147
9.1.5 Effect of Out-of-Set Sample Removal on the AF2-predicted Dynamics and Kinetics . . . 149
9.1.6 Comparison of AF2 Results to Other Generative Machine-learned Models 152
9.1.7 Dynamics of HIV-1 dimer from AlphaFold-Multimer . . . 157
9.2 Discussion and Conclusions . . . 160
9.3 Methods . . . 162
9.3.1 Using Conformational Ensembles to Parameterize a Langevin Equation to Approximate Conformational Dynamics . . . 162
9.3.2 Calculating Time Correlation Functions from Structural Ensembles . . . 163
9.3.3 Measuring the Distance between Time Correlation Functions . . . 164

Chapter 10: Conclusion and Future Works 167

10.1 Summary . . . 167
10.2 Future directions . . . 170
10.2.1 Interpretation of AF2 for a given MSA . . . 170
10.2.2 Better initial guesses with experimental hints . . . 171
10.2.3 Path-representing information bottleneck for non-overlapping initial trajectories . . . 172
10.2.4 SPIB for protein subsystems . . . 173
10.2.5 Context-dependent AF2-RAVE . . . 173
10.2.6 Application perspective: Family-specific AF2-RAVE . . . 174
10.2.7 Application perspective: Acceleration-based designing of peptides and quantifying antigenicity . . . 174

Chapter A: Supporting Information for Chapter 2 176
A.1 Evoformer Layer-Wise Interpretation . . . 176

Appendix B: Supporting Information for Chapter 3 177
B.1 Methods . . . 177
B.1.1 AlphaFold2 . . . 177
B.1.2 Metadynamics . . . 178
B.1.3 State predictive information bottleneck . . . 179
B.1.4 Other simulation details . . . 179
B.1.5 Collective variable choices . . . 179
B.2 Benchmarking the protocol using cold shock protein . . . 180
B.3 Conformational diversity in SSTR2 . . . 182
B.4 Functionality and structural changes in UBA2 . . . 183

Appendix C: Supporting Information for Chapter 4 187
C.1 Methods . . . 187
C.1.1 Training State predictive information bottleneck . . . 187
C.1.2 Simulation details . . . 187
C.2 Collective variables . . . 187
C.3 Comparison of AF2 results across kinases . . . 189
C.4 Sampling results . . . 192
C.5 Using GrASP on our results . . . 196
C.6 Code Availability . . . 196
C.7 Data Availability . . . 196

Appendix D: Supporting Information for Chapter 5 198
D.1 Methods . . . 198
D.1.1 Summary of systems tested and tools used . . . 198
D.1.2 Generation of rMSA AF2 ensemble, AF2-cluster ensemble and AF2-template structures . . . 198
D.1.3 AF2RAVE protocol . . . 199
D.1.4 Docking details . . . 200

Appendix E: Supplementary Figures and Data 201

Appendix F: Supporting Information for Chapter 6 215
F.1 MD simulations . . . 215
F.1.1 Unbiased MD simulations . . . 215
F.1.2 Restrained MD simulations . . . 216
F.2 SPIB setup . . . 217
F.2.1 Global SPIB . . . 217
F.2.2 Local/pocket SPIB . . . 217
F.3 Dataset for ground truth holo structures of Ca^{2+}-S100A1 and Ca^{2+}-S100B . . . 218
F.4 Docking setup . . . 218
F.4.1 Retrospective docking with Glide XP . . . 218
F.5 Detailed Docking and Enrichment Protocols . . . 219
F.5.1 Docking validation . . . 219

Appendix G: Supporting Information for Chapter 7 238
G.1 Methods . . . 238
G.1.1 Molecular dynamics setup . . . 238
G.1.2 Binding-pose metadynamics . . . 238

Appendix H: Supporting Information for Chapter 8 241
H.1 Synaptic Cleft Calculation . . . 241
H.2 Modelling Methods . . . 242
H.2.1 rMSA AF2/AF-multimer . . . 242
H.2.2 Molecular dynamics . . . 243
H.2.3 Enhanced sampling . . . 243
H.2.4 Defining lower and upper bounds . . . 244
H.2.5 TCR IS calculation using the CAR-BOOST approach . . . 244
H.2.6 Free energy calculation . . . 244
H.2.7 Intermembrane distance and synaptic cleft calculation . . . 245
H.2.8 ϕ value calculation . . . 245
H.2.9 Error calculation for the ϕ values . . . 245

Appendix I: Supporting Information for Chapter 9 272
I.1 Ensemble generating methods . . . 272
I.1.1 Machine-learned Protein Structural Ensemble Generation . . . 272
I.1.2 Generating Ground-Truth Conformational Ensembles Using Atomistic Molecular Dynamics . . . 273
I.2 Relaxation Timescales as a Function of Sequence . . . 273
I.3 Increasing the Number of Samples Has Little Effect in Predicted Timescales . . . 273
I.4 LE4PD Analysis for the Slow Modes of Each HIV-1 Protease Monomer Variant . 277
I.5 Conformational Diversity Plots for all HIV-1 Protease Sequences . . . 277
I.6 Dynamical Time Warping and Kolmogorov-Smirnov Distance Matrices . . . 279
I.7 Timescale Correlation Plots for the Six HIV-1 Protease Sequences . . . 279
I.8 Out-of-set-structure Analysis for All Six HIV-1 Protease Sequences . . . 286

I.9 Out-of-set-structure Analysis for BioEmu and AlphaFold2-Cluster 286
I.10 Analysis for a Wild-type Dimer and a Mutant Dimer 294
I.11 Out-of-set-structure Analysis of Dimer Ensembles Generated from Co-folding ML Methods 299

List of Tables

2.1 Conformational ensemble generation through perturbation-based approaches for AF2. 39

5.1 Comparing the IFD performance of various structure generation methods for docking type II kinase inhibitors 79

7.1 **Quantitative validation of AF2-RAVE-selected structures after MHC alignment.** RMSDs were calculated after aligning the MHC backbone to the corrected 6UK2 biological assembly. Q values were calculated using native heavy-atom contacts between the CDR loops and either the peptide or the MHC. 112

B.1 SSTR2 residues used to compute pairwise distances as inputs for RAVE. 180

C.1 Residues included in collective variable definitions 188
C.2 Distances used as input collective variables for SPIB 188

List of Figures

1.1 **Ensemble perspective and Boltzmann distribution.** A representative free-energy profile for a biomolecule illustrates how metastable conformational states contribute to biological function through their Boltzmann weights. A binding partner can shift the conformational population toward a functional state, providing a thermodynamic basis for conformational selection. This ensemble view is central to drug discovery and immunotherapy because rare states may contain selective pockets, altered signaling motifs, or functionally important binding geometries. . 5

2.1 **Coevolutionary information from multiple sequence alignment** Shows a representation of coevolutionary signal to structure hypothesis. 25

2.2 **AlphaFold2 architecture** a) Shows the architecture of AlphaFold2. The pipeline is drawn based on the architecture and algorithms presented by Jumper et al.[1]. I have expanded the initialization procedure presented in AF2. b) Using Abl1 Kinase as an example the heatmaps depict the role of recycle pre- and post- Evoformer block. 29

2.3 **Requirement of MSA** a) AF2 predictions using only sequence information, without MSA or template information, for ubiquitin, BPTI, and Abl1 kinase. These examples illustrate a simple way to assess whether AF2 requires MSA information for a given protein. b) AF2 predictions as MSA sequences are gradually added according to sequence similarity. c) Residue-blocking procedure used to coarse-grain the sequence representation for feature attribution. d) Schematic representation of the feature-attribution protocol in AF2. e) Feature-attribution results highlighting the important blocks in MSA that contribute to AF2 structure prediction. 31

2.4 **Reduced MSA approach to AF2** Structural description of the kinase domain using DDR1 kinase as an example. The right panel shows the key residues involved in the spatial classification of DFGin, DFGinter, and DFGout states. The distances between DFG-Phe-CZ and αC-Glu(+4)-CA, and between DFG-Phe-CZ and β3-Lys-CA, are used to define a two-dimensional space in which DFGin, DFGinter, and DFGout conformations are classified. b) Cartoon representation of sequence regions in kinase proteins. The blocks denote the sequence regions of interest for each kinase system. c) and d) Reduced MSA results for Abl1 and Src kinase domains. e) Reduced MSA results for kinase domains with experimentally known DFGout stable conformations. For all reduced-MSA studies, 640 structures were generated per hyperparameter. 36

3.1 **AlphaFold2RAVE pipeline:** The schematic diagram presents the AF2RAVE pipeline using a 66 residue long cold shock protein(CSP) as an example. The initial ensemble generation block obtains distinct conformations from clustering the structures generated by perturbing AF2. Unbiased MD is used as a data augmentation tool for information bottleneck based dimensionality reduction. Enhanced sampling is performed to reweight the conformational distribution and rank each conformations. 43

3.2 **Rotameric metastability in cold shock protein: a)** Results using AlphaFold2-RAVE on the 1HZB cold-shock protein (CSP). Structure of CSP with tryptophan shown in purple in its most stable rotameric form. Potentials of mean force (PMFs) projected along the two-dimensional SPIB coordinates σ_1 and σ_2 learned from my scheme: **b)** using structures obtained from reduced-MSA AlphaFold2, which provides additional conformational diversity compared with traditional AlphaFold2 but without thermodynamic reliability; **c)** using AlphaFold2-RAVE, showing both conformational diversity and thermodynamic ranking. 46

3.3 **On missense mutation and conformational changes: a)** SSTR2 structures highlighting helix movement in TM6, TM7, and TM5 via AlphaFold2-RAVE enhanced sampling, in agreement with experiments [2] and biological function. Low-energy structures are shown in color and overlaid on the native structure shown in gray. **b)** SSTR2 PMF using AlphaFold2-RAVE, showing multiple potential states. Black dots show structures obtained from reduced-MSA AF2, with structures marked as "junk" representing CV regions that cannot be sampled through transitions. **c)** UBA2 PMFs along total Q-value for WT (purple) and L355A (green). The WT shows a visibly higher barrier, whereas L355A has a wider and more stable disordered region. 47

4.1 **a)** Structural anatomy of the DDR1 kinase molecule showing the activation loop (A-loop) (red), Gly-rich P-loop (blue), αC-helix (purple), and the characteristic N-lobe β sheets (green). These motifs are relevant to the DFG-in to DFG-out transition. **b)** The conserved salt bridge between K57 in β_3 and E74 in the αC-helix that is crucial in ligand dissociation and basic kinase function [3]. 53

4.2 Representative structures from reduced-MSA AF2 for the DFG-in (purple), DFG-inter (blue), and DFG-out (green) states shown in two views and superimposed on the same structure. 54
4.3 Populations of active, inactive, and the known transition state, referred to as the "DFG-up" or "DFG-inter" state, for wild type and mutants D671N, Y755A, and Y759A. **a)** Populations obtained through reduced-MSA AF2, combining MSA depths 8 and 16. **b)** Populations obtained using my AF2-RAVE protocol. I clearly observe that AF2 alone is unable to distinguish between wild type and mutants, and in particular gives the wrong order of stability for the mutants. In contrast, AF2-RAVE is able to identify the reversal in stability upon point mutation and gives more thermodynamically representative populations for these states, in excellent agreement with benchmark calculations performed by Hanson et al. [4] using unbiased MD simulations. 57
4.4 An example trajectory for the DFG-out to DFG-in transition obtained using AF2-RAVE, with residues D (in red), F (in purple), and G (in blue). See Code Availability for details of an example video showing the transition. 58

5.1 A schematic of the AF2-RAVE Glide workflow: (i) decoy structures generated by reduced MSA AF2, (ii) regular space clustering and unbiased MD simulations starting from cluster centers, (iii) State Predictive Information Bottleneck model (SPIB, a RAVE variant) to learn reaction coordinates from unbiased MD, (iv) enhanced sampling runs to calculate free energy landscape, (v) distinguish holo-like structures from decoys in metastable states based on Boltzmann rank and conduct Glide or Induced-fit Docking (IFD) on holo-like structures for ligands targeting metastable states. (iv') and (v'): the decoy structure set and the learnt SPIB coordinates are transferable to homologous systems. 64
5.2 A)The NMR (Nuclear Magnetic Resonance) structures of Abl1 kinase overlay, comparing the activation loop (A-loop) in the active DFG-in state (red, PDB: 6XR6) with the classical DFG-out state (blue, PDB: 6XRG). Type I inhibitors target the active DFG-in state, where the DFG motif adopting the DFG-in configuration and the A-loop adopting the "extended" configuration, while type II inhibitors target the classical DFG-out state, where the DFG motif adopting the DFG-out configuration and the A-loop is "folded". The distance between CB atoms of residue N98 (grey bead) and residue R162 (red/blue bead) in Abl1 kinase serves as an order parameter here to illustrate the location of the A-loop. The dashed black block emphasizes the different configurations of the DFG motif in this two states. B) The Dunbrack definition for DFG motif configuration is employed here. The Dunbrack space is delineated by two order parameters: D1=dist(F158-CZ, M66-CA), D2=dist(F158-CZ, K47-CA). C) or D) The docking poses with the lowest ligand RMSD for 4 known kinase inhibitors targeting the Abl1 or DDR1 kinase AF2 structure, generated by Glide XP. Co-crystallized structures are shown as light-cyan cartoons (proteins), green sticks (ligand) and magenta sticks (DFG motif). Docking poses are shown as light-gray cartoons (proteins), gray sticks (ligand) and blue sticks (DFG motif). Comparing with type I inhibitors, AF2 structures of protein kinases fail to dock with type II inhibitors. 67

5.3 A) Upper panel: distribution of A-loop location for the reduced MSA AF2 structures of Abl1 kinase. 4 out of 1198 structures are A-loop folded. Lower panel: k-means clustering of 4 A-loop folded Abl1 structures in Dunbrack space, with the number of clusters (n cluster) set to 2. The structures in the cluster closest to the classical DFG-out state (red circles) fail to dock with type II inhibitors using induced-fit docking (IFD). B) Upper panel: distribution of A-loop location for the reduced MSA AF2 structures of DDR1 kinase. 124 out of 1147 structures are A-loop folded. Lower panel: k-means clustering of 124 A-loop folded DDR1 structures in Dunbrack space, with n cluster set to 5. Among the 15 structures in the cluster closest to the classical DFG-out state (red circles), one structure ("holo-model", highlighted by a red circle filled with green) demonstrates successful docking with type II inhibitors, showcasing a ligand RMSD < 2 Å, utilizing IFD or an extended-sampling version of IFD, IFD-trim. C) The docking poses with the lowest ligand RMSD for 2 type II kinase inhibitors targeting the DDR1 kinase "holo-model" structure, generated by IFD or IFD-trim. The color code is the same as Fig. 5.2. 70
5.4 A or B) The unbiased MD trajectories of DDR1 are projected onto the learnt SPIB latent space. In plot A), the colors of sample points represent the A-loop location, while in plot B), they depict the Dunbrack DFG state. The first SPIB coordinate, σ_1, correlates with the A-loop location, and the second SPIB coordinate, σ_2, correlates with configuration of the DFG motif. C) The reduced MSA AF2 structures of DDR1 are projected onto the latent space. Sample points are color-coded based on the A-loop location. Light green stars highlight the 15 classical DFG-out structures selected based on prior information in Fig. 5.3. D) Free energy profile in the A-loop folded region of the latent space, calculated from umbrella sampling simulations. The 15 classical DFG-out structures from reduced MSA AF2 are shown as red cross and circles (structures with free energy less than 1 kJ/mol). The "holo-model" structure is emphasized using a red circle filled with red. The embedding table shows the lowest ligand RMSD in IFD poses of the rMSA AF2 structure with ponatinib. The "holo-model" is among the 2 structures selected by AF2-RAVE (PMF < 1 kJ/mol). 74
5.5 A) The reduced MSA AF2 structures of Abl1 are projected onto the latent space. Sample points are color-coded based on the A-loop location. Light green stars highlight the 30 AF2-template Abl1 structures modelled from the 15 DDR1 classical DFG-out structures. B) Upper panel: the distribution of ligand RMSD for the IFD poses of Abl1 structures and two type II ligands. Lower panel: IFD poses with the lowest ligand RMSD for Abl1 AF2-template structures and two type II ligands. The color code is the same as Fig. 5.2. C) Free energy profile in the latent space, calculated from unbiased MD simulations. The 30 Abl1 classical DFG-out structures from AF2-template are shown as red cross and circles (structures with free energy less than 1 kJ/mol). The "holo-model" structures are emphasized using red circles filled with red. D) The table shows the lowest ligand RMSD in IFD poses of the AF2-template structures with two type II inhibitors. All the four "holo-models" are among the 8 structures selected by AF2-RAVE (PMF < 1 kJ/mol). 77

6.1 The hierarchical AF2-RAVE funnel for generating holo-like structures in metastable states. A) A diverse protein structural ensemble is generated from the sequence using reduced MSA AlphaFold2 (rAF2). B) Short unbiased MD simulations are performed, initiated from cluster centers identified in the rAF2 ensemble. C) Dimensionality reduction is applied to the MD trajectories (denoted trajs above) using the SPIB model[5] with different descriptors and timescales to decouple backbone motions from pocket side-chain dynamics. Guided by free-energy surfaces in both the global SPIB latent space and the pocket SPIB latent space, structures from AF2-based methods are further generated, refined, and selected. The final AF2-RAVE structure set consists of diverse, metastable, holo-like conformations. 85

6.2 Free-energy surfaces for Ca^{2+}-S100B (A) and Ca^{2+}-S100A1 (B) in a transferable global SPIB latent space. Two major global basins are identified, reflecting the breathing of the hydrophobic pocket of calcium-bound S100 proteins. Representative structures from unbiased MD for both basins and both Ca^{2+}-S100 systems are depicted in gray, with only one monomer shown for clarity. The projections of AF2 structures (red crosses) and experimentally determined holo structures (stars) indicate that these structures all reside within basin I, the native global basin. For Ca^{2+}-S100B, experimentally determined holo structures includes 24 published XRD structures in PDB (denoted as PDB24, projected as white stars) and 2 unpublished XRD structures with inhibitor SBi5345 or SBi5369 (projected as blue and orange stars). For Ca^{2+}-S100A1, two peptide-bound holo NMR structural ensembles were projected as blue and orange stars. . 92

6.3 Free-energy surfaces for Ca^{2+}-S100B (A) and Ca^{2+}-S100A1 (B) within their native global basins, represented in the local/pocket SPIB latent spaces. The left panel illustrates the most dominant features (whose weights are larger than 0.5 maximum weights) contributing to the learned SPIB coordinates, σ_1^L,σ_2^L. The color and width of the dashed lines indicate the sign and absolute magnitude of the coefficients in the learned linear combination. In both systems, slow conformational changes are primarily associated with side-chain rotamer transitions around the hinge loop region and the C-terminal, especially the FF-gate region. The projection of the rAF2 ensemble (middle panel, white stars) and AF2 structures (middle panel, red crosses) onto the pocket latent space reveals poor coverage of possible metastable basins for side-chain configurations. Notably, the experimentally determined holo structures (right panel, all stars) fall into metastable basins that are not well sampled by AF2-generated models. 95

6.4 A) Projection of XRD (Ca^{2+}-S100B with SBi5345 or SBi5369, shown as red stars), AF2 (red crosses), AF2-RAVE structure set (green circles) onto the local/pocket free-energy surfaces for Ca^{2+}-S100B with native global basin. B) Ligand RMSD distribution of the Glide XP docking poses generated from the three structure sets. The left panel corresponds to docking poses with inhibitor SBi5345, while the right panel corresponds to inhibitor SBi5369. C) Representative docking poses with the lowest ligand RMSD for each structure set. Docking poses (proteins in grey, ligands as grey sticks) are superimposed onto the corresponding X-ray holo structures (proteins in cyan, ligands as green sticks) for comparison. 96

6.5 A) The side-chain assignment of GLU50 is ambiguous in the crystal structure of Ca^{2+}-S100B/SBi5345. B) BEDROC profiles and C) enrichment profiles for different structural inputs. In both B) and C), the dotted black line represents the outward-flipped orientation of GLU50, while the solid black line denotes the inward-flipped orientation. DockPrep favors the outward-facing conformation, which yields lower enrichment performance compared to the inward-facing conformation. The light green line highlights the best-performing structure (from Basin 4) in AF2-RAVE structure set. 98

7.1 **Binding-pose metadynamics for complex filtering:** a) Interface RMSD, used as the biasing coordinate, plotted against simulation time for two selected initial binding poses out of 23. These two poses were selected as positive-control (left) and negative-control (right) examples. Each time series corresponds to one replicate. b) Conceptual representation of $c(t)$ between two consecutive bias-deposition steps. c) $c(t)$ versus RMSD plot depicting the average bias deposited during unbinding. d) Ranking based on the minimum value of $c(t)$. A structural comparison between the predicted peptide conformation (light violet) and the experimental conformation (black) is shown on the right. 107

7.2 **AF2-RAVE-selected conformations:** a) SPIB latent space for the TCR-pMHC complex. b) AF2-RAVE picked structures (2 per basin) with TCRs (silver) aligned and fixed. 109

7.3 **Validation of AF2-RAVE-selected TCR-pMHC structures against the crystal structure:** Representative structures from basin 0 (left) and 1 (right) are aligned to the reference MHC (PDB: 6UK2). The top MHC-aligned structures compares the global TCR-pMHC docking/positioning geometry, while the bottom highlights the interface interaction between the peptide, CDR loops and MHC. 111

8.1 The hinge region of the CD19 CAR construct regulates responsiveness to $CD19^{low}$ leukemia 130

8.2 Optimizing the CD8α hinge for increased responsiveness of tisa-cel-like and axi-cel-like CARs to $CD19^{low}$ leukemia. 132

8.3 Modeling synaptic clefts generated by CAR using an AlphaFold2RAVE(AF2RAVE)-based model 134

8.4 Cryo-ET links CD8α hinge–dependent synaptic distance to CD19 CAR T-cell responsiveness 136

8.5 The CAR-BOOST model predicts maximally responsive hinges across multiple CAR targets 138

9.1 Overview of our method. a) The protein sequence is input to the machine learning architecture of choice, generating an ensemble of all-atom configurations. This ensemble is coarse-grained at the C_α level and input to the Langevin Equation for Protein Dynamics (LE4PD), which predicts the relaxation timescales (kinetics) of each residues' structural fluctuations (dynamics). b) A representative structure of the HIV-1 protease monomer is shown, with the flap (residues 49-51) and N- and C-termini labeled. c) Residue-dependent autocorrelation functions as a function of MSA depth input to AF2 for the three labeled structural regions in (b). d) Predicted relaxation time of each residue as a function of residue index for one HIV-1 protease sequence, PDB ID: 1EBW. 141
9.2 Estimated timescales for the relaxation kinetics of each residue for the six indicated sequences of HIV-1 protease at MSA depths of a) 512, b) 64, and c) 8. Each sequence variant label on the right-hand side of the plot is printed with the same color as its integrated timescale curve. 143
9.3 Measuring the volume of the space of the slow dynamics spanned by the AlphaFold2 conformational ensembles. a) AlphaFold2-generated ensemble projected onto the free-energy surface spanned by the two slowest LE4PD modes parameterized by a one microsecond MD simulation of the HIV-1 protease sequence encoded by PDB ID: 1EBW. b) Comparing the relative variance in the space spanned by the two slowest LE4PD modes of the AlphaFold2-generated ensembles compared to the one-microsecond MD simulation. I also show in the bottom plot what fraction of structures generated by AlphaFold2 at the indicated MSA depth are outside the support of the one-microsecond dynamics. 145
9.4 Reporting the timescales measured by AlphaFold2 conformational ensembles generated using different MSA depths. a) Percent of each residue's ACF best measured by each of the reported MSAs at three different timescales : 10, 100, and 1000 ns. Colored circles denote the average across sequences at each MSA while grey circle denote the result for each sequence individually. b) Scatter plot and linear fits (colored, dashed lines) of the average integrated correlation time from the reference MD simulations (τ_{avg} MD sim.) and the LE4PD theory parameterized using the AlphaFold2 conformational ensembles at the given MSA depth (τ_{avg} rMSA AF2). Reported are the correlation times from MD and AlphaFold2 for the HIV-1 protease sequence encoded in PDB ID: 1EBW at the 1000 ns timescale. c) Pearson correlation coefficient between the timescales reported in b), but for all six HIV-1 protease sequences studied here and for all three timescales (10, 100, and 1000 ns). Also reported in the second column are the slopes of the linear regression of τ_{avg} MD sim. onto τ_{avg} rMSA AF2 for all six sequences (1Q9P, 2PC0, 3TTP, 1EBW, 4Z4X, 6P9A) and all three timescales (10, 100, and 1000 ns). 148

9.5 Effect of out-of-set structure removal on the predicted dynamics. All comparisons to MD are performed at the 1000 ns timescale. a) Root-mean-square fluctuations along the alpha-carbons predicted from the one microsecond MD simulation of HIV-1 protease with the sequence encoded by PDB ID: 1EBW (black) compared to the AlphaFold2 ensemble generated using an MSA depth of 8 and starting from the same sequence with (red, solid) and without (red, dashed) out-of-set structures in the ensemble input to the LE4PD theory. b) Correlation between the integrated correlation timescale predicted with the entire AlphaFold2 ensemble (τ_{ML}) and the ensemble with out-of-set structures removed (τ_{avg} ML (filtered) for all six sequences. The points are colored by the sequence given in the subplot's legend. c) Integrated timescales for 1EBW τ_{avg} from the LE4PD theory parameterized using the AF2 ensemble generated using an MSA depth of 8 with (blue) and without (red) out-of-set structures included in the ensemble. Including the data points from all six sequences, the Pearson correlation coefficient between the timescales calculated from the ensembles with and without out-of-set structures is 0.912. 151
9.6 Measuring the volume of the space of the slow dynamics spanned by the AF2 conformational ensembles and other competing generative ML methods. a) Free-energy surface spanned by the two slowest LE4PD modes calculated from a one microsecond MD simulation of HIV-1 protease encoded by the sequence PDB ID: 1EBW. Projected onto the surface are the samples from AFc, BioEmu, and AF2 at different MSA depth. b) The relative variance of the two slowest LE4PD modes spanned by ensembles calculated using the same set of methods as in (a) for the six HIV-1 protease sequences. Also plotted in the bottom row is the fraction of out-of-set structures generated using each method. DiG results are omitted because they span a nearly negligible volume in this space, so the relative variance is not visible on the scale provided and none of the structures are outside the support of the one microsecond simulation. 153
9.7 Reporting the timescales measured by alternative generative ML methods compared to AlphaFold2. a) Percent of residue ACFs best explained by generative ML methods for the HIV-1 protease sequence encoded by PDB ID: 1Q9P at the 10, 100, and 1000 ns timescales. Colored circles denote the average across sequences at each MSA while grey circle denote the result for each sequence individually. b) Correlation between the timescales(τ_{avg}) calculated from the ACFs from the one microsecond MD simulation and those calculated using the indicated ML ensembles input to the LE4PD theory. Dashed lines shown the best fit linear model to the data. c) Pearson correlation coefficients between the MD and ML integrated timescales for the three methods presented in (b) at the 10, 100, and 1000 ns timescales for the six HIV-1 protease sequences investigated here. Shown in the second column is the fit to the linear regression shown of τ_{avg} MD sim. onto τ_{avg} ML. The units for both τ_{avg} MD sim. and τ_{avg} ML are ps. 154

9.8 Applying my protocol for wildtype (1HHP) and a mutant (3UF3) HIV-1 protease dimer using AFmultimer. a) Projection of AFmultimer structures onto the free-energy surfaces spanned by the two slowest LE4PD modes predicted from a microsecond simulation started after equilibration of the structure deposited under PDB ID: 1HHP (left) or PDB ID: 3UF3 (right). b) Time correlation functions either calculated from MD (black curves) for the 'ear' and flap residues in 1HHP (top panels) or 3UF3 (bottom panels). c) The same time correlation as in b), but with the predictions from the AFmultimer ensembles calculated using the reduced ensemble remaining after removing out-of-set structures. d) Comparison between ACFs from MD and my protocol for both pre- and post- removal of out-of-set structures from AFmultimer. 156

10.1 A summary picture of this thesis 168

A.1 Evoformer layer-wise analysis was performed for three distinct protein systems: BPTI, a protein containing multiple cysteine residues; RfaH, a fold-switching protein; and the Abl1 protein kinase domain, a large multidomain protein. The Frobenius norm was calculated between the pair representation at each Evoformer layer and the final pair representation. Each curve corresponds to the representation before or after the triangle attention/update operation. For all three proteins, the profiles show a two-step transition, with one transition near layer $\sim$8 and another near layer $\sim$46. This suggests that the final few Evoformer layers undergo a sharp transition toward the final prediction. However, a larger-scale study is required to make a more general statement. Such information may eventually help guide improved fine-tuning strategies for AF2-like models. 176

B.1 Reduced-MSA AF2 structures projected on SPIB obtained from AF2-RAVE, used to compute average pLDDT scores. We use structures that do not overlap with regions we sample to compute these scores. 178

B.2 Representative structures for the six metastable states in Trp8 orientations for cold shock protein 1HZB, with Trp8 highlighted in purple. 181

B.3 PMFs along sidechain angles of the Trp8 residue which exhibits six metastable states. In (a) we show calculations from 20 ns long unbiased trajectory initialized from XRD structure. In (b) we show results of metadynamics performed on sidechain angles of the Trp8 residue. In (c) we show results from metadynamics performed using information bottleneck obtained without *a priori* knowledge of Trp8 metastability. Finally, in (d) we show a results from using our protocol *without* reduced-MSA AF2 inputs, learning a bottleneck with only the crystal structure initialized unbiased trajectory as an input. Colorbars for all 4 figures are shown with energies in units of $kcal/mol$. 182

B.4 Ergodicity enhancement in CSP residues. To examine the improvement in sampling of sidechain configurations, we compute the volume sampled in χ spaces for each residue and plot this for metadynamics performed by (i) directly biasing the sidechain dihedrals χ_1 and χ_2 of the Trp8 sidechains (purple) and (ii) biasing along the information bottleneck our methodology obtains (green). Both volumes are normalized by volume sampled by unbiased trajectories of the same length. . 183
B.5 The variation (represented by a periodic colormap) of some carefully chosen sidechain angles in our trajectory plotted on the SPIB coordinates. We see that the states are detecting rotameric transitions in several residues aside from Trp8 as well. 184
B.6 PMFs plotted on Q value-RMSD space to represent global structural changes found by AF2-RAVE for a) L355A and b) WT. Here, for reference structure, we use an equilibrated structure from classical AlphaFold2. 185
B.7 Fractional helicity for L355A and WT reweighted by metadynamics time-independent weights[6] : a) residues from the full protein structure, b), c), d) residues belonging to helix1 (H1), helix2 (H2) and helix3 (H3), respectively. Helices are defined by structure in the protein's wild type native fold. 186

C.1 a) Visualization of input CVs for SPIB in red dashed lines, detailed visualization of the CVs in b)-e) in black lines. b) CVs corresponding to the A-loop distance with N lobe αC helix and C lobe N173[7], c) CVs used for Dunbrack classification of DFG in/inter/out conformations that primarily focus on distances between DFG Phe and the conserved salt bridge between K57 and E74 present in the N lobe[8], d) CVs corresponding to A-loop DFG motif distances with to N lobe αC helix and C lobe R172. These CVs also represent the dissociation of the conserved salt bridge by including the salt bridge distance, as in c), and the structural integrity of the P-loop. e) CVs corresponding to A-loop DFG motif distance with N lobe αC helix and β3. 189
C.2 Collective variable distributions for reduced MSA AF2 structures for wild type and mutants D671N, Y755A, and Y759A. These CVs are described in Table 1, and represent a large number of important collective motions in the molecule. The distributions have been plotted with offsets since these are highly overlapping and similar, indicating that AF2 even with reduced MSA can not distinguish between conformational stabilities of these sequences. The zero lines are marked with dashed black lines. 190
C.3 Statistics of structural diversity obtained from AF2 reduced MSA on three kinases: DDR1, Abl, and Src, showing that AF2 does not really distinguish between them. 191
C.4 Bias potential learnt from metadynamics on SPIB learnt from unbiased AF2RAVE seeded trajectories. The potential is plotted on SPIB coordinates $\sigma_{0,1}$ 192
C.5 PMFs sampled for all four systems projected on Dunbrack co-ordinates, a) Wild type, b) Y759A, c) Y755A and d) D671N. The solid lines differentiate the DFG loop conformations. The top-left quadrant describes DFG-in, the bottom-left quadrant describes DFG-inter, the bottom-right quadrant describes DFG-out and the unassigned belongs to the top-right quadrant. 193

C.6 50ns unbiased trajectories seeded with initial structures from AlphaFold2 reduced-MSA . . . 194
C.7 Visualization of free energy differences for DFG-in to -out transition in bar chart form. . . . 195
C.8 Visualization of the E57-K74 salt bridge forming and helix upward motion observed during sampling. . . . 195
C.9 Cryptic pocket predictions for various free energy minimum structures in DFG-in and DFG-out conformations for wild type DDR1. . . . 197

E.1 The distributions of ligand RMSDs for Glide XP docking poses of DDR1 and type I/type II inhibitors. . . . 202
E.2 Full length multiple sequence alignment for Abl1, DDR1 and Src kinases. . . . 203
E.3 Filtering out unphysical structures in rMSA AF2 ensembles for Abl1, DDR1 and Src kinases. . . . 203
E.4 Filtering out unphysical structures in AF2-cluster ensembles for Abl1, DDR1 and Src kinases. . . . 203
E.5 IFD docking results for the AF2 structure or classical DFG-out structures in rMSA AF2 ensemble of DDR1 or Abl1 against type II inhibitors. . . . 204
E.6 Extended-sampling IFD: IFD-trim for the "holo-model" structure in DDR1 rMSA AF2 ensemble against type II inhibitors. . . . 205
E.7 DiffDock results for DDR1 AF2 structure or classical DFG-out structures in DDR1 rMSA AF2 ensemble against type II inhibitors. . . . 206
E.8 Regular space clustering for DDR1 rMSA AF2 ensemble. . . . 206
E.9 Distributions and their overlaps in the latent space for all umbrella sampling windows. . . . 207
E.10 DDR1 PMF calculated with all the umbrella sampling windows. . . . 208
E.11 PMF values and Boltzmann ranks fluctuate with the setup of the umbrella sampling simulations. . . . 209
E.12 DDR1 PMF calculated with unbiased MD simulations. . . . 210
E.13 The distribution overlap graph for the unbiased MD trajectories of DDR1 or Abl1. 210
E.14 IFD docking results for Abl1 tAF2 structures against type II inhibitors. . . . 211
E.15 The overlay between AF2RAVE PMF and AF2-generated ensembles in the latent space. . . . 211
E.16 The AF2-template structure for Src kinase in classical DFG-out state. . . . 212
E.17 The projection of AF2-generated ensembles for SrcK in the latent space. . . . 212
E.18 IFD docking results for SrcK tAF2 structure against type II inhibitors. . . . 213
E.19 Discarding umbrella sampling windows with αC helix broken for Abl1. . . . 214
E.20 Abl1 PMF calculated from umbrella sampling simulations. . . . 214

F.1 Pipeline for enrichment test using DOCK. Known inhibitors are diluted with a large set of property-matched random molecules in a ratio $>$ 50:1 (50 decoys per true inhibitor). A) AF2-RAVE allows capturing key side-chain fluctuations needed for DOCK to make the required interactions. B) Ligands from the ZINC database are used as decoys. C) 10 known inhibitors are used as true actives D) DOCK is used to score these ligands to the same pocket from diverse conformations of each representative structure E) This test evaluates how effectively DOCK scoring can distinguish true actives from decoys for a given structure. The AF2-RAVE ensemble demonstrates strong performance, achieving approximately 14-fold enrichment over random selection. 220
F.2 Structural quality assessment of (template) rMSA AF2 ensembles for S100A1 (left) and S100B (right) dimers. A) Backbone C_α RMSD values of reduced MSA AF2 ensembles generated with varying MSA depths, evaluated against the corresponding AF2 reference structure. B) Backbone C_α RMSD values of template reduced MSA AF2 ensembles, using the corresponding AF2 reference structure as a template. A 7 Å RMSD filter (retaining structure below black dashed line) is applied. C) Minimum distances for non-bonded heavy atom pairs in each structure from the template reduced MSA AF2 ensembles. A 0.7 Å cutoff (retaining structure above black dashed line) is applied to eliminate structures with steric clashes. The x-axes correspond to the pLDDT ranking within each ensemble. . . 222
F.3 Structural variability in long MD along the full amino acid sequence of S100 proteins, using a sliding window of 12 residues. The mean and variance of Q_{ref} values are represented by blue lines and error bars, respectively. The horizontal red line indicates the cutoff threshold set at 0.9. 223
F.4 A) The four rigid α-helices within a single chain of the Ca^{2+}-S100 dimer are shown in blue. The terminal C_α atoms of each helix, used as anchors for global feature construction, are highlighted as green spheres. Pairwise distances between these anchor atoms define the global features. Calcium ions are omitted for visual clarity. B) Side chains of pocket residues are displayed as sticks, with their distal atoms colored in green. Pocket features are defined by the pairwise distances between these distal atoms. 224
F.5 A-B) Distributions of distances between each oxygen atom in EF-hand residues and the corresponding Ca^{2+} ions from NMR structures of Ca^{2+}-S100A1 (PDB: 1ZFS, subplot A) and Ca^{2+}-S100B (PDB: 1MWN, subplot B). The N-terminal pseudo EF-hands are shown on the left and the C-terminal typical EF-hands on the right of each panel. Those oxygen–Ca^{2+} pairs with distance distributions entirely below the black dashed lines (marking the 5 Å threshold) were restrained in all the MD simulations. C) The format of the restraint potential applied to selected oxygen–Ca^{2+} pairs. D) The distribution of the number of unstable oxygen-Ca^{2+} coordinate bonds (defined as average distances $>$5 Å, despite restrained by potential in subplot C.) across 25 unbiased MD trajectories (10 for Ca^{2+}-S100A1 and 15 for Ca^{2+}-S100B). Only trajectories with two or fewer unstable coordinate bonds (to the left of the black dashed line) were retained for subsequent SPIB analysis. 224

F.6 Bar plots depict the linear weights assigned to each of the 28 global features in the learned global SPIB coordinates, using a time lag of 25 ns on the Ca^{2+}-S100B system. The right panel highlights the dominant features, defined as those with weights exceeding half of the maximum, contributing to each coordinate. The upper and lower panels correspond to σ_1^G and σ_2^G, respectively. The color and width of the dashed lines reflect the sign and absolute magnitude of the weights in the learned linear combinations. 225
F.7 A) Projection of reduced MSA AF2 (rAF2) structures which fall into the two major basins onto the global SPIB latent space, based on a free energy cutoff of 4 kJ/mol. B) The mean and variance of IB values from unbiased MD trajectories used for subsequent local/pocket SPIB analysis within each global basin. Left and right panels correspond to the Ca^{2+}-S100B and Ca^{2+}-S100A1 systems, respectively. No reduced MSA AF2 structures fall into basin 1 of the Ca^{2+}-S100A1 system, and only one unbiased trajectory met the criteria for local SPIB analysis in that basin. 226
F.8 Bar plots showing the linear weights assigned to each of the 53 local features in the learned local SPIB coordinates, on the Ca^{2+}-S100A1 native global basin (subplot A) and Ca^{2+}-S100B native global basin (subplot B), using a time lag of 300 ps. 227
F.9 A–C) Projection of structures used and generated during the AF2RAVE structure set generation pipeline on the local SPIB space corresponding to the native global basin of Ca^{2+}–S100B: A) MD frames used as templates for AF2-template (MD templates); B) the AF2-template structures (tAF2); C) final AF2RAVE structures after short, restrained MD (rMD). D) Projection of all three structure sets onto the global SPIB space for Ca^{2+}–S100B, illustrating that the global backbone configuration remains consistent throughout the structure generation process. . . 228
F.10 A-B) Distribution of PLIF similarity coefficients, Tanimoto coefficient (Tc, left panel) and McConnaughey coefficient (Mcc, right panel), for Glide XP docking poses of SBi5369/Ca^{2+}–S100B (subplot A) and SBi5345/Ca^{2+}–S100B (subplot B), using the corresponding X-ray holo structures as references. C) SBi5345/Ca^{2+}–S100B docking poses with the highest Tc values from each structure category. Docking poses (proteins in grey, ligands as grey sticks) are superimposed onto the corresponding X-ray holo structures (proteins in cyan, ligands as green sticks) for comparison. 229
F.11 Ligand RMSD values are plotted against PLIF similarity coefficients, Tanimoto coefficient (Tc, subplot A) and McConnaughey coefficient (Mcc, subplot B), for each Glide XP docking pose, using the X-ray holo structures as reference. The left and right panels represent docking poses for SBi5345/Ca^{2+}–S100B and SBi5369/Ca^{2+}–S100B, respectively. Docking poses that located in the upper left region indicate accurate atomic environment modeling of both ligand placement and protein-ligand interactions. AF2RAVE structures consistently outperform AF2 and perform comparably to experimental XRD structures. 230

F.12 The AF2RAVE structure set used for the enrichment test. We noticed equilibration trajectories without subsequent production runs in restrained MD provided slightly better coverage of the five local/pocket basins within the native global basin of Ca^{2+}–S100B. Consequently, this structure set was selected for the enrichment analysis. 231
F.13 Average similarity between known ligands was found to be very high. The similarity distribution (SD) of each ligand to its nearest neighbor in that sets is shown. (a) SD using all 87 ligands known to bind to S100B (b) SD using 10 selected ligands that were as distinct to each-other as possible 232
F.14 The average size of ZINC ligands is much lower than our set of known binders. Molecular weight distributions (in a.m.u) for (a) 2 million random ligands from ZINC database. The red line denotes the lowest molecular weight of a known binder (325 a.m.u) (b) Selected "heavy" subset as decoys. All decoys were selected to be heavier that the cut-off (red line in (a)) 232
F.15 Enrichment diversity from individual pockets. The average profiles (thick green) used to calculate enrichment are computed by first averaging the DOCK score from both pockets and then ranking all ligands by this average score. Profile for the crystal structure is shown for reference in each plot. (a-e) 5 basins as identified by local structure SPIB (f) 4 different pockets from 2 known crystal structures. Pocket 1 was omitted from subsequent calculations as key residue (PHE 88) was not modelled. The vertical dashed line is the value of α most commonly used in literature to compare these profiles (higher is better). 233
F.16 Comparing enrichment profiles from AF2 structures with pLDDT rank 1 (AF2 recommended) and pLDDT rank 2 (used to improve variance). However, diversity comes at a cost of poor structure quality as lower ranked AF2 structures show extremely poor enrichment. The dashed line shows the recommended literature value of α to use for comparing these profiles (higher is better) 234
F.17 Comparing the enrichment profiles from each basin. Key for both is shown in the inset of plot B. A) Direct enrichment profile showing number of true binders found when first K ligands are taken. B) BEDROC profiles constructed by changing α. Grey (bottom-most) represents Basin 0, which provides the least enrichment. Basin 4 (green) closely followed by Basin 2 (Purple) perform the best. The black profile is for the crystal structures (for reference) and the red profile is the vanilla AF2 structure. The dashed line shows the recommended literature value of α to use for comparing these profiles (higher is better) 235

F.18 A) X-ray structure of the SBi5345/Ca^{2+}–S100B complex (cyan, ligand hidden) and its corresponding protein pocket after DOCK protein preparation (green). The GLU50 residue, which exhibits two possible sidechain rotamer conformations in the X-ray structure, is highlighted with red circles. The default DOCK preparation selected the outward-facing rotamer for docking. B) Top-ranked compounds (magenta sticks) from the enrichment test using the DOCK-prepared X-ray structure of SBi5345/Ca^{2+}–S100B. Docking poses are aligned with the original X-ray structure (cyan). Ranking of active inhibitors are labeled in green, while ranking of decoys are labeled in red. Several top-ranked decoys exhibit potential steric clashes with the inward-facing GLU50 conformation, indicating that ambiguities in X-ray sidechain assignments allow these decoys to artificially achieve high docking scores, thereby compromising enrichment performance. GLU50 and neighboring residues are shown as sticks; calcium ions are omitted for visual clarity. 236
F.19 Variation in BEDROC scores seen by changing the crystal structures (each bound to a different ligand). Each curve represents one crystal structures (score averaged from both pockets). Structures that were monomeric, covalently modified, or bound to large molecules and peptides were omitted from the plot. All curves are labelled at their bottom-right by the PDB ID of the crystal structure used for docking. The black profile is for the crystal structures (for reference) and the red profile is the vanilla AF2 structure. The dashed line shows the recommended literature value of α to use for comparing these profiles (higher is better) 237

G.1 Template usage case study. The χ_1-χ_2 conformational space of the peptide position-6 tryptophan residue is shown. Violet dots represent generated structures, and the red star denotes the experimental structure. 239
G.2 Binding pose metadynamics results. 240

H.1 Cell surface expression and functional evaluation of modular CD19 CAR constructs. 247
H.2 Cell surface expression and functional evaluation of hinge-modified CD19 CAR T-cells. 249
H.3 Defining lower and upper geometric bounds of a functional immunological synapse. 251
H.3 Defining lower and upper geometric bounds of a functional immunological synapse. 253
H.4 Validation of fluctuation independence and synaptic bounds definitions using CAR-BOOST-like modeling of the canonical T-cell synapse. 256
H.5 Reduced-MSA AF2 performance on CD8α-derived hinge sequences. 258
H.6 AlphaFold2-RAVE as a distribution generator for CD8α-derived hinge sequences. 262
H.7 Correlated fluorescence light microscopy and cryo-ET images of CAR T-cell synapses. 265
H.8 Synaptic cleft distance calculation from cryo-ET images. 267
H.9 Target protein modeling using CAR-BOOST. 269
H.10 Modeling non-CD8α-derived hinges for CAR targeting CD19. 271

I.1 Timescales at each MSA depth studied in the main text as a function of sequence. Each subpanel is labeled according to the sequence whose timescales it describes. 274

I.2 Integrated relaxation timescales for each of the six HIV-1 protease variants examined in the main text as function of MD simulation length. Specifically, I probe the relaxation timescales at lengths of a) 10 ns, b) 100 ns, and c) 1000 ns. 275
I.3 Change in the information contained by the ensemble as the number of samples increases. a) (Top) shows the change in von Neumann entropy in the mean-centered covariance matrix calculated for 5000(5k), 1000(1k), and 640 samples generated independently for the sequence of 4Z4X PDB. The change is calculated with the covariance matrix calculated for 320 samples as reference. (Bottom) Displays the backbone relaxation timescales calculated using LE4PD. b) shows the result for the sequence belonging to 2PC0 PDB. 276
I.4 Local mode lengthscale (LML) for the first a) and second b) LE4PD modes for each of the six sequence variants studied in the main text as a function of the MSA depth in AF2. Embedded ensembles show the LML projected onto the alpha-carbons for the MSA=8 ensemble for each sequence. The colorscale is red-white-blue, with red indicating no dynamics and blue indicating the most dynamic regions. The scaling is the same for each sequence for each mode, but the normalization of the scaling is set uniquely for each mode. Also given in each plot is the mode's timescale for the MSA=8 ensemble. The LML for each sequence is shown as the dashed, grey curve as a reference 277
I.5 Projection of AlphaFold2 structures at the indicated MSA depth onto the space spanned by the two slowest LE4PD modes, which are calculated individually for each protein from the respective one microsecond trajectory starting from the highest confidence AlphaFold two structure. Results for sequences a) 3TTP, b) 2PC0, c) 1Q9P, d) 1EBW, e) 4Z4X, and f) 6P9A. 278
I.6 Projection of AlphaFold2 structural ensembles generated at MSA depths of 8 and 64 as well as those generated from AlphaFold2-cluster, BioEmu, and DiG, onto the space spanned by the two slowest LE4PD modes, which are calculated individually for each protein from the respective one microsecond trajectory starting from the highest confidence AlphaFold2 structure. Results for sequences a) 3TTP, b) 2PC0, c) 1Q9P, d) 1EBW, e) 4Z4X, and f) 6P9A. 280
I.7 Projections of samples from the indicated ML methods and MD starting structures onto the free-energy surfaces spanned by the two slowest LE4PD modes. Besides the change in the number of contour lines present in the free-energy surfaces, the X-ray structures used as the initial conformations in MD simulations are projected onto the free-energy landscapes as yellow stars, except for 1Q9P, where the starting structure harvested from the NMR ensemble is indicated with a magenta star while the remaining nineteen structures in the ensemble are indicated as yellow stars. 281
I.8 Distance matrices between the brute-force calculated time correlation functions from the MD simulation and the analytical LE4PD time correlation functions parameterized using the AlphaFold2 structural ensembles generated at the indicated MSA depth for three simulation lengthscales indicated (10, 100, and 1000 ns) for each of the six HIV-1 protease sequences studied here: a) 3TTP, b) 2PC0, c) 1Q9P, d) 1EBW, e) 4Z4X, and f) 6P9A. The distance metric used is the dynamical time warping distance between the two time correlation functions. 282

I.9 Distance matrices between the brute-force calculated time correlation functions from the MD simulation and the analytical LE4PD time correlation functions parameterized using the AlphaFold2 structural ensembles generated at the indicated MSA depth for three simulation lengthscales indicated (10, 100, and 1000 ns) for each of the six HIV-1 protease sequences studied here: a) 3TTP, b) 2PC0, c) 1Q9P, d) 1EBW, e) 4Z4X, and f) 6P9A. To calculate this set of distance matrices, each time correlation function is converted to its respective cumulative distribution function. Then, the Kolmogorov-Smirnov test between the two cumulative distribution functions is calculated, and the mapping is performed to the simulated time correlation functions with the lowest value output from the Kolmogorov-Smirnov test. 283
I.10 Distance matrices between the brute-force calculated time correlation functions from the MD simulation and the analytical LE4PD time correlation functions parameterized using either AlphaFold2 or other machine-learned (AlphaFold2-cluster, BioEmu, or DiG) structural ensembles for three simulation lengthscales indicated (10, 100, and 1000 ns) for each of the six HIV-1 protease sequences studied here: a) 3TTP, b) 2PC0, c) 1Q9P, d) 1EBW, e) 4Z4X, and f) 6P9A. The distance metric used is the dynamical time warping distance between the two time correlation functions. 284
I.11 Distance matrices between the brute-force calculated time correlation functions from the MD simulation and the analytical LE4PD time correlation functions parameterized using either AlphaFold2 or other machine-learned (AlphaFold2-cluster, BioEmu, or DiG) structural ensembles for three simulation lengthscales indicated (10, 100, and 1000 ns) for each of the six HIV-1 protease sequences studied here: a) 3TTP, b) 2PC0, c) 1Q9P, d) 1EBW, e) 4Z4X, and f) 6P9A. To calculate this set of distance matrices, each time correlation function is converted to its respective cumulative distribution function. Then, the Kolmogorov-Smirnov test between the two cumulative distribution functions is calculated, and the mapping is performed to the simulated time correlation functions with the lowest value output from the Kolmogorov-Smirnov test. 285
I.12 Correlation between the integrated timescales from the one microsecond MD simulation starting from the most confidence structure predicted from AlphaFold2 for the given HIV-1 protease sequence and the integrated timescale from the analytical residue-residue time correlation functions predicted from the LE4PD theory parameterized using the AlphaFold2 structural ensemble generated at an MSA depth of 8. Results for sequences a) 3TTP, b) 2PC0, c) 1Q9P, d) 1EBW, e) 4Z4X, and f) 6P9A. 287
I.13 Correlation between the integrated timescales from the one microsecond MD simulation starting from the most confidence structure predicted from AlphaFold2 for the given HIV-1 protease sequence and the integrated timescale from the analytical residue-residue time correlation functions predicted from the LE4PD theory parameterized using either AlphaFold2-cluster-, BioEmu-, or the DiG-generated ensembles. Results for sequences a) 3TTP, b) 2PC0, c) 1Q9P, d) 1EBW, e) 4Z4X, and f) 6P9A. 288

I.14 Projection of structures from the AlphaFold2 ensemble with an MSA depth of 8 projected onto the free-energy surfaces spanned by the first two LE4PD modes for a) 1EBW and b) 2PC0. The mean structure of the distribution is labeled with a green triangle; the mode is labeled with a white triangle; outlying structures are colored yellow; finally, in-distribution structures (according to the microsecond MD) are colored blue. Comparison of the RMSF before (red) and after the removal of outlying structures (blue) are shown for c) 1EBW and d) 2PC0. The RMSF calculated from a microsecond MD of these two sequences is given by the solid, black curve. Finally, the integrated timescales with all structures kept in the ensemble (red) and with outlying structures removed (blue) are shown in e) for 1EBW and f) for 2PC0. 289
I.15 Projection of structures from the AlphaFold2 ensemble with an MSA depth of 8 projected onto the free-energy surfaces spanned by the first two LE4PD modes for a) 1Q9P and b) 3TTP. The mean structure of the distribution is labeled with a green triangle; the mode is labeled with a white triangle; outlying structures are colored yellow; finally, in-distribution structures (according to the microsecond MD) are colored blue. Comparison of the RMSF before (red) and after the removal of outlying structures (blue) are shown for c) 1Q9P and d) 3TTP. The RMSF calculated from a microsecond MD of these two sequences is given by the solid, black curve. Finally, the integrated timescales with all structures kept in the ensemble (red) and with outlying structures removed (blue) are shown in e) for 1Q9P and f) for 3TTP. 290
I.16 Projection of structures from the AlphaFold2 ensemble with an MSA depth of 8 projected onto the free-energy surfaces spanned by the first two LE4PD modes for a) 4Z4X and b) 6P9A. The mean structure of the distribution is labeled with a green triangle; the mode is labeled with a white triangle; outlying structures are colored yellow; finally, in-distribution structures (according to the microsecond MD) are colored blue. Comparison of the RMSF before (red) and after the removal of outlying structures (blue) are shown for c) 4Z4X and d) 6P9A. The RMSF calculated from a microsecond MD of these two sequences is given by the solid, black curve. Finally, the integrated timescales with all structures kept in the ensemble (red) and with outlying structures removed (blue) are shown in e) for 4Z4X and f) for 6P9A. 291
I.17 RMSF predicted from the BioEmu ensemble with (red) and without (blue) out-of-set structures included in the calculation for each of six HIV-1 protease variants: a) 3TTP; b) 2PC0; c) 1Q9P; d) 1EBW; e) 4Z4X; f) 6P9A. The RMSF is benchmarked by the RMSF calculated directly from a one microsecond equilibrium MD simulation (black). 292
I.18 Integrated timescales τ_{avg} predicted from the BioEmu ensemble with (red) and without (blue) out-of-set structures included in the calculation for each of six HIV-1 protease variants: a) 3TTP; b) 2PC0; c) 1Q9P; d) 1EBW; e) 4Z4X; f) 6P9A. 293

I.19 RMSF predicted from the BioEmu ensemble with (red) and without (blue) out-of-set structures included in the calculation for each of six HIV-1 protease variants: a) 3TTP; b) 2PC0; c) 1Q9P; d) 1EBW; e) 4Z4X; f) 6P9A. The RMSF is benchmarked by the RMSF calculated directly from a one microsecond equilibrium MD simulation (black). 294
I.20 Integrated timescales τ_{avg} predicted from the AlphaFold2-cluster ensemble with (red) and without (blue) out-of-set structures included in the calculation for each of six HIV-1 protease variants: a) 3TTP; b) 2PC0; c) 1Q9P; d) 1EBW; e) 4Z4X; f) 6P9A. 295
I.21 Histograms of a) the curling angle of the three residues of the flap tip, averaged over the two monomers; b) the flap-flap distance between monomers in the HIV-1 protease dimer; c) the I50-D25 distance for the first monomer in the dimer; and d) the I149-D124 distance for the second monomer in the dimer. 296
I.22 Local mode lengthscale (LML) for the six slowest LE4PD modes calculated from MD simulations of the wild-type 1hhp (black) and mutant 3uf3 (blue) HIV-1 protease dimer. The LMLs projected onto the structural ensembles from the microsecond MD simulations are colored using a red-white-blue colorscale, where red indicates no dynamics and blue indicates large scale conformational dynamics. Note that the colorscale is scaled to the LML of the dimer simulations, and it does not correspond to the same scale used in the analogous plot for the monomer simulations. Here, the LMLs are around an order of magnitude smaller compared to the monomer results. 297
I.23 Projection of the flap-flap distance (first row), flap tip-catalytic site distance (middle row), and the cosine of the flap tip curling angle (bottom row) onto the free-energy surface spanned by the first two LE4PD modes for the wildtype (1HHP; left column) and mutant (3UF3; right column) sequences. 298
I.24 Example C_{2h} symmetry structures for a) WT 1HHP sequence and b) MUT 3UF3 sequence predicted by AF2-multimer. 299
I.25 Overlap of ML-generated ensemble with MD ensemble. a) Spread of the conformational ensemble is shown for the WT (PDB ID:1HHP) and b) mutant (PDB ID:3UF3). c) Shows the percentage out-of-set samples present for both WT and mutant HIV-1 protease dimers. 300
I.26 Relaxation timescales for relevant residues. a) (Top) Shows the WT sequence's ACFs for the active site (residue 25), 'ear' (residue 25),flap tip (residue 25), and 'cheek' (residue 25). (Bottom) Shows the same ACFs calculated without the out-of-set distribution. b) Shows similar analysis for the mutant sequence. 301
I.27 DTW metric calculation. a) Shows the time warp distance between the MD ACF and my predicted ACF for the ensemble generated by the co-folding models. The top panel shows the calculated DTW matrix for the WT sequence, and the bottom panel shows that for the mutant. b) Shows the same analysis for the ACFs calculated without the out-of-set samples. 302

List of Abbreviations and Acronyms

AI	Artificial intelligence
ML	Machine learning
AF2	AlphaFold2
MSA	Multiple sequence alignment
rMSA AF2	Reduced MSA approach for AF2
lDDT	Local distance difference test
pLDDT	Predicted local distance difference test
PDB	Protein data bank
CV	Collective variable
RC	Reaction coordinate
RAVE	Reweighted autoencoded variational Bayes for enhanced sampling
AMINO	Automatic mutual information noise omission
SPIB	State predictive information bottleneck
RMSD	Root mean square deviation
MD	Molecular dynamics
MetaD	Metadynamics
WT-MetaD	Well-tempered metadynamics
US	Umbrella sampling
NVT	Constant number, volume, and temperature ensemble
NPT	Constant number, pressure, and temperature ensemble
AF2-RAVE	AlphaFold2-RAVE
LE4PD	Langevin equation for protein dynamics
CAR	Chimeric antigen receptor
TCR	T-cell receptor
pMHC	Peptide–major histocompatibility complex
IS	Immunological synapse
CAR-BOOST	Chimeric antigen receptor biophysically oriented optimization of synaptic topology

List of Publications

1. Da Teng, Yunrui Qiu, Gokulakannan Shakthivel, **Akashnathan Aranganathan**, Lukas Herron & Pratyush Tiwary. (2026). Detecting and quantifying overparametrization in RNA language models with REDIAL. *bioRχiv*. (**under submission**)

2. **Akashnathan Aranganathan***, Justin M. Mirazze*, Jaehyun Suh*, Xiang Chen*, Dongya Jia, Makoto Ando, Sooraj R. Achar, Christopher D. Chien, Marie Pouzolles, Jienyu Ding, Kniya DéDé, Anjali Chandroth, Philippe Youkharibache, Naoko Mizuno, Raul Cachau, Nirali N. Shah, Christopher D. Chien, Kylie J. Walters, Grégoire Altan-Bonnet, Naomi Taylor & Pratyush Tiwary. (2026). CAR-BOOSTing T-Cell Responsiveness through Synaptic Cleft Engineering. *Nature Chemical Biology*. (**under revision**)(*- *equal contribution*)

3. **Akashnathan Aranganathan*** & Eric Beyerle*. (2026). Applied Causality to Infer Protein Dynamics and Kinetics. *Journal of Chemical Information and Modeling*. Vol. 66, Issue 3, pp. 1661–1674. (*- *equal contribution*)

4. Xinyu Gu*, Venkata S. S. Adury*, **Akashnathan Aranganathan**, Xinhao Zhuang, Kristen M. Varney, David J. Weber & Pratyush Tiwary. (2025). Hierarchical AF2RAVE for Multiconformation Virtual Screening Targeting S100 Ca^{2+}-Binding Proteins. *Journal of Chemical Theory and Computation*.(*- *equal contribution*)

5. Da Teng*, Vanessa J. Meraz*, **Akashnathan Aranganathan**, Xinyu Gu, and Pratyush Tiwary. (2025). af2rave: protein ensemble generation with physics-based sampling. *Digital Discovery*. Vol. 4, Issue 8, pp. 2052-2061. DOI: 10.1039/D5DD00201J (*- *equal contribution*)

6. **Akashnathan Aranganathan***, Xinyu Gu*, Dedi Wang, Bodhi P. Vani and Pratyush Tiwary. (2025). Modeling Boltzmann weighted structural ensembles of proteins using AI based methods. *Current Opinion in Structural Biology*. Vol. 91, ISSN 0959-440X, pp. 103000. DOI: 10.1016/j.sbi.2025.103000 (*- *equal contribution*)

7. Xinyu Gu, **Akashnathan Aranganathan** and Pratyush Tiwary. (2024). Empowering AlphaFold2 for protein conformation selective drug discovery with AlphaFold2-RAVE. *eLife*. Vol. 13, pp RP99702. DOI: 10.7554/eLife.99702.1

8. Bodhi P. Vani, **Akashnathan Aranganathan** and Pratyush Tiwary. (2023). Exploring kinase DFG loop conformational stability with AlphaFold2-RAVE. *Journal of Chemical Information and Modeling*. Vol. 64, Issue 7, pp. 2789–2797. DOI: 10.1021/acs.jcim.3c01436

9. **Akashnathan Aranganathan***, Bodhi P. Vani*, Dedi Wang and Pratyush Tiwary. (2023). AlphaFold2-RAVE: From Sequence to Boltzmann Ranking. *Journal of Chemical Theory and Computation*. Vol. 19, Issue 14, pp. 4351–4354. DOI: 10.1021/acs.jctc.3c00290 (*- *equal contribution*)

Chapter 1: Introduction

Advances in immunotherapy and drug discovery primarily rely on understanding the biological functions of biomolecules such as globular proteins, intrinsically disordered proteins/regions (IDPs/IDRs), nucleic acids, and various associated small molecules. Traditionally, to assess the function of a biomolecule system, a "sequence-structure-function" paradigm was presented and has led to the development of X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, and, more recently, cryo-electron microscopy and other structural assignment experimental methods. These developments have led to an astonishing 180,000 biomolecule structures in the RCSB protein databank[9] by 2021. The accumulation of these structural data and protein sequence information led to the development of data-based tools like AlphaFold2 (AF2)[1], RosettaFold[10], ColabFold[11], OpenFold[12], ESMFold[13], Boltz-1[14] and AlphaFold3[15] (AF3).

However, this "sequence-structure-function" perspective is still incomplete, as biomolecules are inherently dynamic[16, 17] and primarily function through an ensemble of structures understood as representative structures with Boltzmann weights representing their probabilities(Fig. 1.1). Theoretically, it has already been discussed using simple statistical mechanics models like spin glass systems[18, 19] and lattice models[20] to describe the ensemble perspective of the proteins and the resultant protein folding funnel. With the increase in Artificial Intelli-

gence (AI)-based methods especially for better static protein structure prediction, the shift from "sequence-structure-function" to "sequence-ensemble-function" is being realized in the computational molecular biophysics field. I have written a brief overview of these new approaches including what we can learn from such ensemble predictors and where do these models are limited in performance or in principles. My PhD thesis is primarily centered on this thermodynamically corrected ensemble prediction and more importantly, how this approach can lead to improve immunotherapies and drug design. In this chapter I would like to give brief introduction to concepts and methods that lead up to the recent structure/ensemble prediction methods. I hope this chapter serves as a good road map to my thesis.

1.1 Conceptual Background and History of Protein Biophysics

1.1.1 Proteins as Heteropolymers and Protein Folding Problem

Proteins were not widely recognized as covalently linked polypeptides or heteropolymers until the early twentieth century, when peptide chemistry established that they are indeed chains of amino acids [21]. Further, the determination of the amino-acid sequence of insulin by Sanger and coworkers strengthened this polymer view, establishing proteins as heteropolymers with specific amino-acid sequences [22, 23](primary sequence). The structural, or more physical, view of protein molecules emerged only in the mid-twentieth century. This development was followed by the discovery of the α-helix [24], the identification of sterically allowed regions of backbone dihedral-angle space[25] (secondary structure), and, most importantly, the crystallographic structures of myoglobin and hemoglobin [26, 27]. These demonstrated that globular proteins possess specific three-dimensional structures(tertiary structure).

The "Thermodynamic Hypothesis", later proposed by Anfinsen, established the connection between a protein's amino-acid sequence, or primary sequence, and its biologically active "native fold". Anfinsen used renaturation experiments on ribonuclease A to show that this native fold enables the protein to perform its function. He also showed that, under denaturing conditions, the native fold is lost, forming largely inactive scrambled conformations. Thus, he proposed that the three-dimensional native fold of a protein in its physiological milieu is the conformation with the lowest Gibbs free energy, and that this native fold is encoded by its amino-acid sequence [28].

Together with advances in protein sequencing and crystallography, Anfinsen's thermodynamic hypothesis laid the foundation for the "sequence-structure-function" paradigm, in which the primary amino-acid sequence encodes the information required to form a three-dimensional native fold under physiological conditions, and this native fold provides the basis for function. However, this thermodynamic hypothesis raised a question about the timescale of folding. This problem, known as Levinthal's paradox, argued that a random search through all accessible conformations would require folding timescales much longer than those observed on biological timescales [29].

Altogether, this defines the classical protein folding problem: given a linear amino-acid sequence and a specified environment, how does the primary sequence determine its biologically active three-dimensional structure, and how does the chain reach this structure on biological timescales? This problem can be divided into three different problems [30]. The first is the folding-code, or thermodynamic, problem: what balance of intramolecular and solvent-mediated interactions determines the native fold or other relevant conformations? The second is the folding-mechanism, or kinetic, problem: how are these conformations reached on biological timescales? The third is the computational problem: can the native fold, or more generally the

relevant conformational ensemble, be predicted from sequence alone?

Levinthal's paradox is better understood by noting that proteins are not free homopolymers exploring an unbiased conformational space. A homopolymer can collapse into many compact conformations with similar free energies. In contrast, natural proteins are evolved heteropolymers whose sequences encode patterns of intramolecular and solvent-mediated interactions that stabilize biologically relevant native conformations under physiological conditions [31]. Borrowing ideas from spin-glass systems, Wolynes and coworkers [18] developed the principle of minimal frustration, which suggests that the native fold has relatively low frustration because the dominant interactions between amino acids are mutually compatible, thereby making the conformational space of natural proteins more funnel-like [32, 33]. Similar arguments were made from lattice-model studies by Dill, Shakhnovich, Karplus, and others [34, 35, 36], which showed that fast-folding sequences have a free-energy gap between the native structure and other compact non-native conformations, supporting the principle of minimal frustration. Purely from a kinetic perspective, Zwanzig et al. further showed that even a slight energetic penalty against locally unfavorable configurations drastically reduces the mean first-passage time to the native fold from a random conformation [37]. These statistical-mechanical models shifted the view of protein folding away from an unbiased conformational search for a native fold toward a free-energy landscape description in which folding proceeds on a rugged yet globally funneled surface biased toward the native state. In other words, the free energy decreases as the protein approaches native-like states.

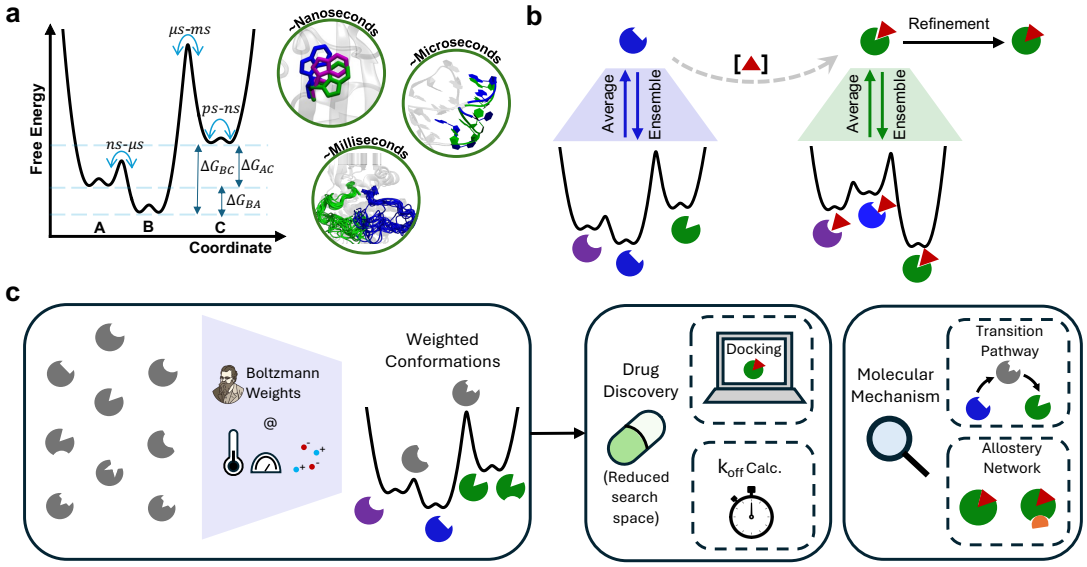


Figure 1.1: Ensemble perspective and Boltzmann distribution. A representative free-energy profile for a biomolecule illustrates how metastable conformational states contribute to biological function through their Boltzmann weights. A binding partner can shift the conformational population toward a functional state, providing a thermodynamic basis for conformational selection. This ensemble view is central to drug discovery and immunotherapy because rare states may contain selective pockets, altered signaling motifs, or functionally important binding geometries.

1.1.2 Metastable States and the "Sequence-Ensemble-Function" Paradigm

The folding funnel describes the free-energy separation between compact non-native or misfolded conformations and the native state, providing intuition for addressing Levinthal's paradox. However, this funnel is not perfectly smooth for a natural protein. While minimal frustration suggests that evolution selects sequences that are globally biased toward native architectures, residual frustration among competing local interactions can roughen the landscape [31, 33]. This ruggedness gives rise to conformational substates separated by kinetic barriers. This implies that proteins dynamically interconvert between these substates rather than behaving as single rigid structures [17]. These kinetic barriers appear across various timescales(Fig. 1.1a). A protein may retain its overall native-like architecture while undergoing conformational changes across multiple length scales and timescales. These changes can range from side-chain rotamer tran-

sitions and flexible-loop motions to binding-pocket opening and large-scale domain rearrangements. When such conformational substates have finite lifetimes, they are generally referred to as metastable states.

To define this more formally, a metastable state is a subset of conformational space in which the system equilibrates rapidly compared with the timescale required to escape from that subset [38, 39]. In other words, transitions among microstates within a metastable basin are fast, whereas transitions from one metastable basin to another are comparatively slow. This definition does not generally require a coordinate or a collective variable. However, as I will show in later chapters, a projected free-energy landscape along a set of coordinates can reveal these metastable states as local minima separated by barriers.

The existence of such metastable states motivates an ensemble perspective for understanding protein function. A protein's primary sequence does not simply encode a single static native structure. Instead, together with its environment, it shapes a free-energy landscape containing multiple metastable states with distinct populations and lifetimes. Thus, a protein can perform its function by transitioning between these states. This motivates a shift from the traditional "sequence-structure-function" paradigm to a "sequence-ensemble-function" paradigm, in which biologically relevant function is determined not only by the dominant native fold but also by the ensemble of accessible metastable conformations and their interconversion kinetics.

1.1.3 Conformational Selection Hypothesis

As hinted in the previous section, the metastable states need not be equally populated. At equilibrium, each metastable state has a probability determined by its free energy. If the free

energy of state i is G_i, its equilibrium population given by

$$p_i = \frac{e^{-\beta G_i}}{\sum_j e^{-\beta G_j}}, \tag{1.1}$$

where $\beta = 1/k_B T$. These probabilities are often referred to as Boltzmann weights. A key point to be noted here as that even conformations that are not dominant in the unbound (or apo) ensemble can still exist as low-population metastable states. Importantly, these Boltzmann weights are not fixed properties of the amino-acid sequence alone. They depend on the environment, including temperature, pH, ionic strength, solvent conditions, post-translational modifications, and the presence of binding partners. A change in environment can therefore reshape the free-energy landscape and shift the relative populations of metastable states.

This idea leads naturally to the conformational selection hypothesis. The concept was described in detailed by Nussinov and coworkers [40, 41, 42](Fig. 1.1b). In conformational selection, a protein is assumed to pre-exist as an ensemble of interconverting conformational states. A ligand or binding partner preferentially binds to one of these pre-existing states and stabilizes it, thereby shifting the ensemble population toward that state. In this view, binding does not necessarily create a new conformation from a single rigid structure. Instead, it selects and stabilizes a conformation that was already present in the ensemble.

Therefore, the conformational selection hypothesis provides a direct functional interpretation of the "sequence-ensemble-function" paradigm. These concepts are essentially the backbone of my thesis. For practical use of these concepts, in my thesis, I apply the conformational selection hypothesis to metastable states and assume that, upon introducing a new binder, the populations of different metastable states change, while the definitions of each metastable state

remain largely unchanged. In other words, the change in the kinetic barrier's location is assumed to be small. This is a crucial assumption that allows us to study a protein in detail in its apo form and leverage the results for drug discovery or protein binders, etc(Fig. 1.1c).

1.2 Methods

I first provide a brief overview of static protein structure prediction methods, followed by methods for studying protein dynamics. Since my thesis lies at the interface of bioinformatics and biophysics, aided by AI-based models, the following methods are used throughout this work.

1.2.1 Static protein structure prediction

As a part of the protein folding problem, static structure prediction works from the "sequence-structure-function" framework. Traditionally, the structure prediction methods can be divided into template-based and non-template-based approaches. Template-based modeling uses experimentally determined structures of related proteins. Non-template, or ab initio, modeling attempts to predict structure without relying on a full structural template. Recent deep-learning methods blur this boundary. In the following subsections, I will be discussing about the strutcure prediction method from bioinformatics perspective alone.

Template-based modeling: homology modeling and threading

Homology modeling is based on the empirical observation that protein structure is often more conserved than sequence. That is, while there exists a vast diversity of protein sequences, the number of distinct native folds are comparatively less[43]. Leveraging this, for a given query

sequence a homolol with a known structure is searched for and chosen to serve a template. A typical homology-modeling workflow contains following steps: identify suitable templates, align the query sequence to the template sequence, build a structural model that satisfies spatial restraints implied by the backbone of the template, refine for the side chain rotamers and score the resulting structural model (e.g., MODELLER)[44, 45]. This approach gives a more accurate structural model depending in the sequence identity between the query and template sequence.

Threading, or fold recognition, extends template-based modeling to cases where sequence similarity is weak. In this approach, a target sequence is aligned against known structural folds and scored according to how compatible each residue is with the local structural environment of the template [46]. This approach can 1) comment on if a query sequence can fold to a specific fold using an energy function, and 2) by threading to a library of folds, perhaps can fid a tertiary structure for a given seqeuence.

Ab initio methods: fragment assembly and coevolution

The fragment-based ab initio methods, most notably the Rosetta framework developed by Baker and coworkers, reduces the conformational search problem of protein folding, by assembling short backbone fragments from known protein structures and scoring the resulting structural models with knowledge-based and physics-inspired energy functions[47, 48]. The framework also uses a bayesian score approach to pick a fragment based on local sequence similarity.

Coevolution-based approaches form another important class of ab initio methods structure prediction methods. These methods use sequence information from a query sequence's homologous sequences to infer residue-residue contacts that constrain the tertiary structure. I will briefly

discuss this idea in the following sections.

Coevolution, Potts models, and direct coupling analysis

Like homology modeling, coevolution-based modeling is motivated by the contrast between the number of distinct protein sequences and the comparatively limited number of distinct native folds[43]. The coevolution hypothesis suggests that three-dimensional constraints can be inferred from correlated mutations observed across homologous sequences[49]. This idea motivated several methods for predicting residue-residue contacts or structural constraints, including correlated-mutation analysis, mutual-information (MI)-based approaches, and direct coupling analysis (DCA) methods such as mean-field DCA and pseudo-likelihood maximization DCA[50, 51]. The key ingredient for these approaches is a multiple sequence alignment (MSA), in which homologous sequences are aligned so that each column approximately corresponds to an evolutionarily equivalent residue position. The performance of these methods depends strongly on the quality and depth of the MSA: too few effective sequences provide insufficient information to infer reliable couplings, whereas highly redundant sequences can distort the estimated correlations.

To understand this framework, let us consider a query protein sequence of length L. The sequence can be represented as a one-dimensional lattice of length L, where each lattice position corresponds to a residue position. Each position i can be occupied by an amino-acid state, or spin, s_i. In practice, each spin can take one of 21 possible values, corresponding to the 20 standard amino acids and an additional gap state. Thus,

$$s_i \in \{A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W, Y, -\}. \tag{1.2}$$

An MSA containing N homologous sequences can then be written as a matrix $M \in \mathbf{S}^{N\times L}$, where $\mathbf{S}$ is the alphabet of possible spin states. Each row of this matrix corresponds to one aligned sequence, and each column corresponds to one residue position. The rows of the MSA can therefore be treated as samples from an evolutionary sequence distribution.

This representation naturally connects coevolution-based modeling to the Potts model. In a maximum-entropy framework, the probability of observing an aligned sequence $\mathbf{s} = (s_1, s_2, \ldots, s_L)$ is written as

$$P(\mathbf{s}) = \frac{1}{Z} \exp\left[\sum_{i=1}^{L} h_i(s_i) + \sum_{1\le i<j\le L} J_{ij}(s_i, s_j)\right], \tag{1.3}$$

where $h_i(s_i)$ are site-specific fields, $J_{ij}(s_i, s_j)$ are pairwise couplings between positions i and j, and Z is the normalization constant. The fields capture the conservation preferences at individual positions, whereas the couplings capture statistical dependencies between pairs of positions. The inverse problem is to infer the parameters $\mathbf{h}$ and $\mathbf{J}$ from the MSA such that the model reproduces the observed sequence statistics. In this context, DCA aims to distinguish direct residue-residue couplings from indirect correlations that arise through chains of interacting residues. The resulting coupling matrix $\mathbf{J}$ contains information about residue pairs that are often close in the tertiary structure, commonly measured using C_β–C_β contacts (or C_α–C_α contacts for glycine). The first implementation used mean-field approximations[50]; later methods improved inference through pseudo-likelihood maximization[51].

1.2.2 Protein dynamics

"Imagine that an alien lands on Earth, hears about something called a 'bicycle,' and wants to understand how it works, how to ride it, and how to fix it when it breaks. Figuring this out

given just a picture of a bicycle would be challenging. Watching a movie of someone riding a bicycle would help. Even better, the alien would experiment with an actual bicycle—for example, by turning a pedal and seeing how the wheels respond." - Hollingsworth et al., 2018[52]

In this section, I introduce briefly the main computational approaches I have used to study protein dynamics in my thesis.

Molecular Dynamics from a Statistical Mechanics Perspective

Molecular dynamics (MD) simulation is a computational method used to describe how a molecular system evolves in time. It can also be viewed as a physics-based sampling method that generates a sequence of time-ordered configurations. In all-atom MD, atoms are treated as classical particles that evolve according to Newton's equations of motion,

$$m_i \frac{d^2\mathbf{r}_i}{dt^2} = -\nabla_i U(\mathbf{R}), \tag{1.4}$$

where m_i and $\mathbf{r}_i$ are the mass and position of atom i, respectively, $U(\mathbf{R})$ is the potential energy of the system, and $\mathbf{R}$ denotes the coordinates of all atoms. The resulting trajectory, $\{\ \mathbf{R}(0), \mathbf{R}(\Delta t), \mathbf{R}(2\Delta t), \ldots\}$, contains structural and dynamical information[53, 54, 55].

The potential energy function $U(\mathbf{R})$ is defined by a molecular mechanics force field. A force field consists of an analytical function with a set of parameters used to approximate bonded and non-bonded interatomic interactions. These parameters are usually calibrated against quantum-mechanical calculations and experimental data[56, 57].

In practice, an MD simulation is usually started from an initial structure $\mathbf{R}(0)$, which is solvated with water molecules and ions. The equations of motion in Eq. 1.4 are then integrated

numerically using a timestep, typically on the order of femtoseconds, so that fast bond vibrations involving hydrogen atoms can be resolved. Thermostats and barostats are often used to control temperature and pressure, allowing simulations to sample approximate canonical (NVT) or isothermal-isobaric (NPT) ensembles[58, 59].

From a statistical mechanics perspective, MD provides samples from an equilibrium distribution. In the canonical ensemble, the configurational probability of observing a configuration $\mathbf{R}$ is

$$P(\mathbf{R}) = \frac{1}{Z} e^{-\beta U(\mathbf{R})}, \tag{1.5}$$

where $\beta = 1/k_B T$ and Z is the configurational partition function. Note that this is only the configurational distribution as we typically integrate out the momentum degrees of freedom since configurational/conformational changes are more interesting. If the simulation is sufficiently long, invoking the ergodicity condition, ensemble averages can be estimated from time averages along the trajectory. For an observable $A(\mathbf{R})$,

$$\langle A \rangle \approx \frac{1}{T} \int_0^T A(\mathbf{R}(t))\, dt. \tag{1.6}$$

Despite its usefulness, conventional MD has two major limitations. First, because the integration timestep is on the order of femtoseconds, many biologically relevant conformational changes, such as loop rearrangements, binding-pocket opening, and domain motions, are difficult to sample directly. Second, the accuracy of the simulation depends on the quality of the force field. Nevertheless, long-timescale MD simulations have successfully captured folding and unfolding events in small proteins and reproduced several experimental observables, demonstrating

that MD is a powerful tool for studying molecular mechanisms, metastable states, and transition pathways [60, 61].

Calculating free energies from molecular dynamics simulations

Calculating free energy is necessary to understand such conformational changes as they account for the ensemble's enthalpic and entropic contributions. Following the Eq.1.5, we can calculate the free energy surface for a conformational coordinate or collective variable, denoted as $\mathbf{s} = s(\mathbf{R})$, using equations 1.7 and 1.8.

$$P(\mathbf{s}) = \frac{\int_0^t dt' \delta[\mathbf{s} - s(\mathbf{R}(\mathbf{t}'))]}{\int_0^t dt'}, \tag{1.7}$$

and

$$F(\mathbf{s}) = -k_B T \ln P(\mathbf{s}) + C, \tag{1.8}$$

where C is an arbitrary constant. The reliability of this estimate depends on both the amount of sampling and the accuracy of the force field. In principle, if the trajectory is sufficiently long and ergodic, the time average in Eq. 1.6 approaches the equilibrium ensemble average. In practice, however, biomolecular free-energy landscapes often contain metastable basins separated by kinetic barriers. As a result, an unbiased MD simulation may remain trapped in one basin and fail to sample other relevant conformations within accessible simulation times.

This sampling limitation directly affects the accuracy of calculated free-energy surfaces. Therefore, improving conformational sampling is essential for reliable free-energy calculations from MD simulations. This motivates the use of enhanced sampling methods, which are designed

to accelerate transitions between metastable states and recover the underlying equilibrium free-energy landscape.

Enhanced Sampling

Enhanced sampling methods accelerate rare events by changing how the system explores phase space such that the desired equilibrium distribution can be recovered later. Broadly, they can be divided into ensemble-based methods and collective-variable-based methods.

Ensemble-based methods run multiple simulations under modified thermodynamic or Hamiltonian conditions. Temperature replica exchange molecular dynamics (T-REMD) simulates replicas at different temperatures and exchanges configurations to help low-temperature replicas cross barriers [62]. Hamiltonian replica exchange modifies parts of the potential energy to enhance sampling of selected interactions [63]. These methods can be powerful, but they often scale poorly with system size and require many replicas.

Collective-variable-based methods bias one or more coordinates $\mathbf{s} = s(\mathbf{R})$. This approach falls under the importance sampling family [64, 65], where sampling is performed to estimate a target distribution by biasing or stratifying the sampled distribution such that the ratio between the sampled and target distributions is either known beforehand or calculated during the simulation. Schematically, this can be written as

$$P^{\text{unbiased}}(\mathbf{s}) \propto P^{\text{biased}}(\mathbf{s})\, e^{w(\mathbf{s})}, \tag{1.9}$$

where P^{unbiased} and P^{biased} are the target and sampled distributions, respectively, and $e^{w(\mathbf{s})}$ is the reweighting factor. Thus, the unbiased distribution can be recovered by reweighting configura-

tions according to the bias applied during sampling. In this sense, enhanced sampling increases the probability of visiting rare but important regions of conformational space, while reweighting restores the equilibrium probabilities of the target unbiased ensemble.

Umbrella Sampling

Umbrella sampling (US) is a classic enhanced sampling method for calculating free-energy profiles along a chosen collective variable $\mathbf{s}$ [66]. Instead of relying on one long unbiased trajectory to sample all relevant metastable states, umbrella sampling divides the collective-variable space into a series of windows, or umbrellas. In each window k, a static biasing potential, commonly a harmonic restraint,

$$V_k(\mathbf{s}) = \frac{1}{2} K_k (\mathbf{s} - \mathbf{s}_k)^2, \tag{1.10}$$

is applied to keep the system near a chosen value $\mathbf{s}_k$. Independent restrained MD simulations are then performed in each window, often in parallel. In this way, the bias potential is predefined before the simulation.

A crucial point is that orthogonal degrees of freedom should be adequately sampled within each window. If these orthogonal degrees of freedom contain long-lived metastable states, then the simulation can suffer from the same limited-timescale problem encountered in unbiased MD. The importance sampling form of US is

$$P_k^{\text{unbiased}}(\mathbf{s}) = P_k^{\text{biased}}(\mathbf{s}) \exp[\beta V_k(\mathbf{s})] \left\langle \exp[-\beta V_k(\mathbf{s})] \right\rangle_{\text{unbiased}}, \tag{1.11}$$

where $\langle \cdot \rangle_{\text{unbiased}}$ denotes an ensemble average over the unbiased distribution. In this case, P_k^{biased}

can be calculated from the biased simulation using Eq. 1.7, and V_k is known from the beginning of the simulation. The ensemble average is the only term that is not trivial to calculate. However, if the distributions across neighboring windows have sufficient overlap, this average term can be treated as a normalization constant because it is independent of $\mathbf{s}$, and can be calculated using the weighted histogram analysis method (WHAM) [67, 68]. Thus, one can obtain the unbiased relative free energy among the metastable states present along $\mathbf{s}$.

Well-Tempered Metadynamics

Metadynamics is an importance sampling method that adds biasing weights as the system visits distinct regions in the chosen collective-variable space, thereby encouraging the system to sample a broader distribution. Practically, this is achieved by adding a biasing potential term $V(s(\mathbf{R}))$, or equivalently $V(\mathbf{s})$, to the force field $U(\mathbf{R})$. By construction, once the target uniform distribution is achieved in standard metadynamics, the accumulated biasing potential approaches the negative free-energy surface along the collective variable, $-F(\mathbf{s})$ [69].

Barducci, Bussi, and Parrinello proposed a more controlled version of this scheme called well-tempered metadynamics (WTMetaD) [70]. In this scheme, the rate of bias deposition decreases with time. The scheme can be implemented using the following stochastic iteration:

$$V_n(\mathbf{s}) = V_{n-1}(\mathbf{s}) + g(\mathbf{s}, \mathbf{s}_n) \exp\left[-\frac{\beta V_{n-1}(\mathbf{s}_n)}{\gamma - 1}\right], \tag{1.12}$$

where $\beta = 1/k_B T$, n denotes the bias-deposition step, and $g(\mathbf{s}, \mathbf{s}_n)$ is the biasing kernel, commonly taken to be a short-ranged Gaussian kernel over the CV space. The biasing potential is updated periodically after every N_g steps, where $\tau_g = N_g\, dt$ and dt is the MD timestep. During

the time between updates, the system evolves under the potential $U(\mathbf{R}) + V_n(\mathbf{s}(\mathbf{R}))$.

One of the most important features of this stochastic iteration scheme is that the bias evolution can be described by an ordinary differential equation (ODE) [64, 70]. This ODE behavior has been shown in different ways and provides a foundation for understanding and leveraging well-tempered metadynamics. Additionally, the asymptotic solution for the ODE at long times is

$$V(\mathbf{s}, t) \to -\left(1 - \frac{1}{\gamma}\right) F(\mathbf{s}) + c(t), \tag{1.13}$$

where $c(t)$ is a time-dependent offset independent of $\mathbf{s}$, given by

$$c(t) = k_B T \ln \frac{\int d\mathbf{s}\, e^{-\beta F(\mathbf{s})}}{\int d\mathbf{s}\, e^{-\beta [F(\mathbf{s}) + V(\mathbf{s}, t)]}}. \tag{1.14}$$

The function $c(t)$ is independent of the collective variable, and while calculating the relative free-energy ranks of states in this collective variable, $c(t)$ cancels out; thus, it can be ignored for that specific purpose. Now, notice the influence of γ on this expression. When $\gamma \to 1$, the sampling tends to an unbiased simulation. When $\gamma \to \infty$, $V(\mathbf{s})$ tends to $-F(\mathbf{s})$, and we recover standard non-tempered metadynamics, where the target distribution is uniform. This suggests that γ quantifies the amount of fluctuation enhancement in this scheme. Thus, the chosen CV $\mathbf{s}$ must contain information about a protein's biologically relevant and essential fluctuations. Generally, γ is also defined as $(T + \Delta T)/T$, providing a temperature-based intuition for how barriers in the collective variable are modified.

The importance sampling form for WTMetaD is given by

$$P^{\text{unbiased}}(\mathbf{s}) = P^{\text{biased}}(\mathbf{s}, t) \exp\{\beta[V(\mathbf{s}, t) - c(t)]\}, \tag{1.15}$$

and $c(t)$, which is an estimator for the reversible work done by the biasing potential, can be calculated by

$$c(t) = k_B T \ln \frac{\int d\mathbf{s} \, \exp\left[\frac{\gamma}{\gamma-1}\beta V(\mathbf{s}, t)\right]}{\int d\mathbf{s} \, \exp\left[\frac{1}{\gamma-1}\beta V(\mathbf{s}, t)\right]}. \tag{1.16}$$

Unlike US, WTMetaD updates the weights on the fly in a single trajectory. This implies that if the trajectory performs multiple back-and-forth motions in $\mathbf{s}$, the reweighting scheme can provide a better estimate of the free energy along $\mathbf{s}$.

Practically, from a simulation, the free energy for any coordinate $\mathbf{q}$ can be calculated by

$$P(\mathbf{q}) = \frac{\int_0^t dt' \, \delta[\mathbf{q} - q(\mathbf{R}(t'))] \, \exp\{\beta[V(\mathbf{s}(\mathbf{R}(t')), t') - c(t')]\}}{\int_0^t dt' \, \exp\{\beta[V(\mathbf{s}(\mathbf{R}(t')), t') - c(t')]\}}. \tag{1.17}$$

Under sufficient sampling and adequate back-and-forth transitions in both the biasing coordinate and the desired coordinate, the $F(\mathbf{q})$ defined using Eq. 1.17 and Eq. 1.8 is an estimate of the unbiased free energy. Further, WTMetaD also allows us to check convergence by comparing $F(\mathbf{s})$ estimated from the biasing potential with $F(\mathbf{s})$ obtained from reweighting.

Low-dimensional representation of protein dynamics

As already hinted, when describing ensembles of biomolecules, specifically proteins, not all the $\mathbb{R}^{3N}$ degrees of freedom, where N is the number of atoms, are required [71]. Especially while dealing with all-atomistic simulation data of these biomolecules, the number of atoms

roughly ranges from 10^3 to 10^7. Typically, system-specific coordinates are used to visualize or interpret different conformations. From a folding perspective, RMSD and Q values, or the fraction of native contacts, are commonly used as metrics to check how close a conformational state is to a known reference conformation [72].

$$\mathrm{RMSD} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} \|\mathbf{r}_i - \mathbf{r}_i^0\|^2}, \tag{1.18}$$

where N is the number of atoms, $\mathbf{r}_i$ is the position vector of atom i in the query structure, and $\mathbf{r}_i^0$ is the position vector of atom i in the reference structure.

$$Q = \frac{1}{|S|}\sum_{(i,j)\in S} \frac{1}{1+\exp\left[\beta\left(r_{ij} - \lambda r_{ij}^0\right)\right]}, \tag{1.19}$$

where S is the set of atom pairs, r_{ij} is the distance between atoms i and j in the query structure, and r_{ij}^0 is the corresponding distance in the reference structure. β is the smoothing factor for the sigmoid function, and λ is an adjustable factor that allows some flexibility.

While these system-agnostic coordinates define how far away a protein conformation is from a given reference state, distinguishing between different conformations based on dynamics is required to understand the system. Since we also aim to perform WTMetaD, we require a collective variable whose fluctuation enhancement will lead to better sampling [73]. This requirement means that the variable should be able to capture the essential dynamics and relevant metastable states. Typically, the coordinate that best describes the dynamics and the mechanism of conformational changes is the reaction coordinate (RC), the most informative degree of freedom [74].

State predictive information bottleneck (SPIB)[5]

SPIB is one such dimensionality reduction method that performs dimensionality reduction while simultaneously identifying the number of metastable states. This is achieved by leveraging the information bottleneck approach, where the neural network attempts to find the minimum amount of information required, $\mathbf{z}$, from a given set of input features, $\mathbf{X}$, to predict the target $\mathbf{y}$.

$$\mathcal{L}_{IB} = \mathcal{I}(\mathbf{y}, \mathbf{z}) - \beta\mathcal{I}(\mathbf{z}, \mathbf{X}), \tag{1.20}$$

where $\mathcal{I}(x, y)$ denotes the mutual information between x and y.

The SPIB loss function uses a variational approach for the information bottleneck, Eq. 1.20, to predict the future state, $\mathbf{X}^{t+\Delta t}$, of a system given its current information through a set of collective variables, $\mathbf{X}^t$. The following equation presents the SPIB loss function:

$$\mathcal{L}_{SPIB} = \frac{1}{M}\sum_{n=1}^{M}\left[\ln q_\theta(\mathbf{y}^{n+s}|\mathbf{z}^n) - \ln\left(\frac{p_\theta(\mathbf{z}^n|\mathbf{X}^n)}{r_\theta(\mathbf{z}^n)}\right)\right]. \tag{1.21}$$

Here, M denotes the total number of frames present in an MD simulation, $\mathbf{X}$ denotes the input collective variables, $\{\mathbf{X}^n\}_{n=1}^M$, $\mathbf{z}$ is the latent variable, or RC, and $\mathbf{y}$ denotes the one-hot encoded state labels, $\{\mathbf{y}^n\}_{n=1}^M$. The time interval between $\mathbf{X}^n$ and $\mathbf{X}^{n+s}$ is Δt.

Maximizing the $\mathcal{L}_{SPIB}$ function leads to a latent space that incorporates dynamics through the time-lag parameter and forms an information-dense space required to differentiate the states. In each iteration of maximizing the loss, SPIB predicts the probability that the system starting from $\mathbf{X}$ will be found in state i after a time delay of Δt. This probability denotes the state

transition probability. At the core of SPIB is an iterative scheme where, based on the state transition probabilities for a given Δt, SPIB updates the state labels for the system starting from $\mathbf{X}$. This on-the-fly update also allows SPIB to predict the number of metastable states with an escape time, Section 1.1.2, greater than Δt. Moreover, the boundaries between the SPIB-predicted states correspond to a state transition probability of 0.5, resulting in what is known as the committor surface. Thus, the SPIB latent space presents itself as a reasonable reaction coordinate [74, 75, 76, 77, 78, 79, 80, 81].

1.3 Thesis descriptor and biological motivation

The big picture problem in this thesis is to develop methodologies that work in "sequence-ensemble-function" regime. To achieve this goal, in this thesis the conformational selection hypothesis is used. In this thesis, I use a conformational metastable state selection hypothesis: ligand or binder association mainly shifts the populations of metastable states that already exist in the apo ensemble. The major apo metastable states are therefore assumed to remain approximately conserved, while their relative Boltzmann weights change upon binding. This assumption allows the apo landscape to be characterized first and then used to guide drug discovery, protein binder design, and related applications(Fig. 1.1).

With this in mind, this thesis develops and applies computational strategies that bridge AF2-like structure prediction with MD-based thermodynamic reweighting. The core methodological idea is to use predicted structural diversity as an initial hypothesis for possible metastable states, learn low-dimensional coordinates that separate those states, and recover Boltzmann-ranked ensembles through enhanced sampling. This framework is designed to answer questions

such as:

(i) Which metastable conformations are accessible to a protein sequence under a specified condition?

(ii) Which of these conformations are thermodynamically important?

(iii) Can conformational information learned for one protein be transferred to its homologs?

(iv) Can rare but druggable or immunologically relevant conformations be identified before ligand binding or receptor engagement?

(v) Can ensemble-based descriptors improve the design of immunotherapies, especially systems where disorder, flexibility, and population shifts are functionally important?

(vi) Can AI generated structural ensembles be interpreted in terms of their dynamical timescales and functional relevance?

The following chapters build on this foundation. The first part of the thesis focuses on combining data-based models with physics-based sampling to obtain Boltzmann-ranked conformational ensembles. The second part applies this idea to conformation-selective drug discovery, where metastable states provide docking targets that may not be apparent from a dominant static structure. The third part extends the ensemble perspective to cancer immunotherapy, where conformational distributions can influence downstream signaling. The final part fits a Langevin equation to AI-generated structural ensembles to estimate their dynamical timescales.

Chapter 2: Understanding and Adding Stochasticity to AlphaFold2

In recent years, since the development of AlphaFold2 (AF2)[1], The computational study of protein molecular biophysics has undergone an extreme change. In the Critical Assessment of protein Structure Prediction round 14 (CASP), AF2 outperformed other methods, achieving near-experimental accuracy on many CASP 14 targets. The subsequent release of easy-to-access implementations, particularly ColabFold[11], further lowered the barrier for the broader scientific community. Since then, AF2 has been used not only as a structure prediction tool but also as a platform for method development in molecular biophysics, structural biology, protein design, docking, mutational analysis, and conformational sampling. Apart from the structural biology community, AF2 has also been used across many areas of biology as a structural hypotheses for interpreting molecular mechanisms at atomistic resolution.

Interestingly, many top-performing methods in CASP14 substantially improved over the best methods from CASP13. While the availability of large structural and sequence databases was essential, one of the key ingredients behind this progress is the use of large attention-based neural networks that integrate sequence, evolutionary, structural template, and geometric information. However, despite its empirical success, several questions remain: what has AF2 learned about the sequence-to-structure relation, how reliable are its predictions, how strongly does it depend on MSA information, and to what extent do its predictions reflect generalizable structural principles

rather than memorized patterns from training data?

In this chapter, I focus on AF2's architecture, the flow of information from its inputs, possible strategies for feature attribution, and, most importantly, ways to introduce stochasticity into the network. The overall goal of this chapter is to motivate how AF2 can be used.

2.1 From Potts Models to Masked Language Models to AlphaFold2

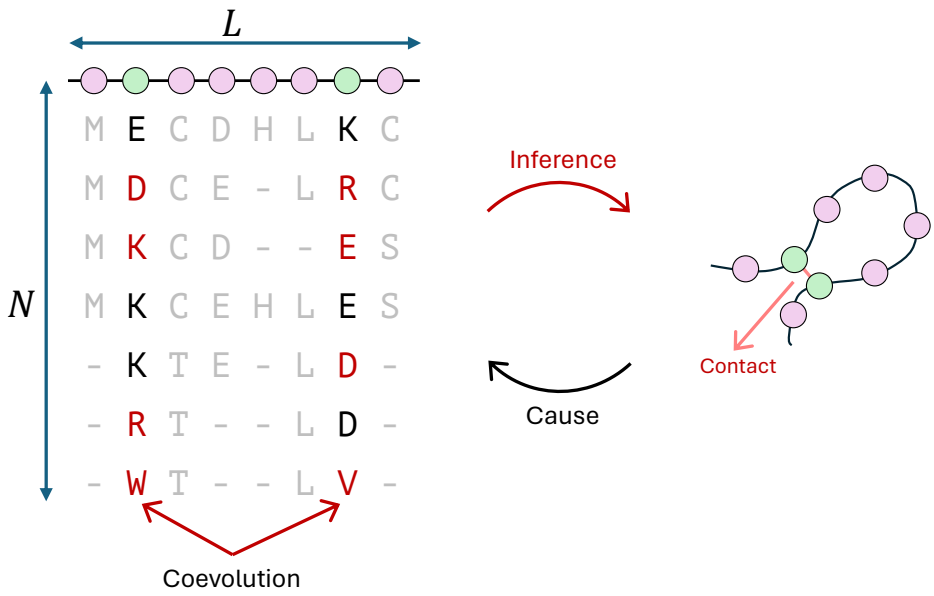


Figure 2.1: Coevolutionary information from multiple sequence alignment Shows a representation of coevolutionary signal to structure hypothesis.

In the previous chapter I discussed briefly about the static structure prediction methods involving template based and non template based. One of the key discovery is the coevolutionary information for *ab initio* contact prediction. As mentioned in the previous chapter, coevolution information was developed by inferring that occurrence of certain pairs of mutations are coupled due to their interaction in the protein's three-dimensional structure(Fig. 2.1). This is purely a correlation-based method development and thus requires a better classes of methods that can dif-

ferentiate between noise and coevolutionary signals. Typically, a multiple sequence alignment is performed to search and align sequences belonging to the family or similar to the query sequence. This gives an MSA matrix $M \in \mathbf{S}^{N\times L}$, where $\mathbf{S}$ is the set of naturally occurring amino acids together with gap ('-'), N is the number of sequences in the MSA, and L is the length of the sequence, as described in the first chapter. Each column of this matrix provides information about the frequency of amino acids at a given site in the sequence, while pairs of columns provide information about the joint frequency of amino-acid pairs at two sites. Using these quantities, the mutual information (MI) between two sites can be used to estimate the correlation between them. Formally, for an amino-acid state $a \in \mathbf{S}$ at position i, the single-site frequency is

$$f_i(a) = \frac{1}{N}\sum_{m=1}^{N} \delta_{s_i^m,a}, \tag{2.1}$$

where s_i^m is the amino-acid state at position i in sequence m, and $\delta(\cdot,\cdot)$ is the Kronecker delta. Similarly, the joint frequency of observing amino-acid states a and b at positions i and j is

$$f_{ij}(a,b) = \frac{1}{N}\sum_{m=1}^{N} \delta_{s_i^m,a}\ \delta_{s_j^m,b}. \tag{2.2}$$

The mutual information between positions i and j is then given by

$$MI_{ij} = \sum_{a\in\mathbf{S}}\sum_{b\in\mathbf{S}} f_{ij}(a,b)\ln\left[\frac{f_{ij}(a,b)}{f_i(a)f_j(b)}\right]. \tag{2.3}$$

Here, MI_{ij} quantifies how strongly the amino-acid distributions at positions i and j deviate from statistical independence. A high mutual information value suggests that the two sites are correlated across the MSA. However, these correlations can arise from direct coupling, indirect

coupling through other residues, phylogenetic effects, or sampling noise. Thus, an additional average product correction (APC) is commonly used to reduce background correlation effects and refine the contact signal [82]. Later, direct information was later suggested to disentangle the direct and indirect couplings through a maximum entropy modeling resulting in a Potts model built on the MSA. As discussed in Chapter1, this leads to a Potts model over aligned sequences, where site-specific fields capture conservation and pairwise couplings capture direct statistical dependencies between residue positions. From the inferred pairwise coupling parameters, one can define the direct information (DI) between two sites as

$$DI_{ij} = \sum_{a,b} P_{ij}^{\text{dir}}(a,b) \ln \left[\frac{P_{ij}^{\text{dir}}(a,b)}{f_i(a) f_j(b)} \right], \tag{2.4}$$

where $P_{ij}^{\text{dir}}(a,b)$ is the direct pair distribution constructed from the inferred coupling $J_{ij}(a,b)$ and constrained to reproduce the single-site frequencies $f_i(a)$ and $f_j(b)$.[50].

However, this inference problem framework is difficult because the number of parameters is large and the partition function is generally intractable. A Pseudolikelihood maximization approach was then developed by replacing the total likelihood with a product of conditional probabilities. That is,

$$P(s_i^m \mid \mathbf{s}_{\setminus i}^m) = \frac{\exp\left[h_i(s_i^m) + \sum_{j \neq i} J_{ij}(s_i^m, s_j^m)\right]}{\sum_{a \in \mathbf{s}} \exp\left[h_i(a) + \sum_{j \neq i} J_{ij}(a, s_j^m)\right]}. \tag{2.5}$$

Then, the pseudolikelihood objective is

$$\mathcal{L}_{\text{PL}} = \sum_{m=1}^{N} \sum_{i=1}^{L} \log P(s_i^{(m)} \mid \mathbf{s}_{\setminus i}^{(m)}). \tag{2.6}$$

where N is the number of sequences in the MSA, L is the sequence length, $s_i^{(m)}$ is the amino-acid state at position i, and $\mathbf{s}_{\setminus i}^{(m)}$ denotes all positions except position i. This conditional view is closely related to masked language modeling (MLM) in BERT-like transformer architectures, where masked tokens are predicted from the surrounding sequence context [83].

Bhattacharya et al. [84] showed that a simplified attention-based model can recover contact information comparable to Potts models, and that a factored-attention construction can be interpreted as a parameter-efficient way to model protein family interactions. In this formulation, the key idea is to decouple positional interactions from amino-acid interaction features. In other words, the positional encoding and amino-acid state information are separated between the attention matrix and the value vectors. Later work analytically connected factored attention to generalized Potts models, showing that an attention-based architecture with a single-head layer can recover the coupling matrix of a Potts model with fewer parameters [85].

In a similar spirit, MSA Transformer was introduced as a protein language model that takes an MSA as input and is trained using a masked language modeling objective. By using attention across both sequence, or row-wise, and alignment, or column-wise, dimensions, MSA Transformer learns correlations between residue positions and predicts contact maps from coevolutionary information in the MSA [86].

AF2 builds on related ideas through the Evoformer module, which processes both an MSA representation and a pair representation. The pair representation can be viewed as carrying residue–residue information analogous to a learned J_{ij}-like representation, while the additional triangle-attention operation refines pairwise signals by enforcing consistency among residue triplets [1]. Thus, the progression from Potts models to masked language models provides useful insight into a part of AF2's mechanism.

2.1.1 AlphaFold2: Everything, Everywhere, All at Once

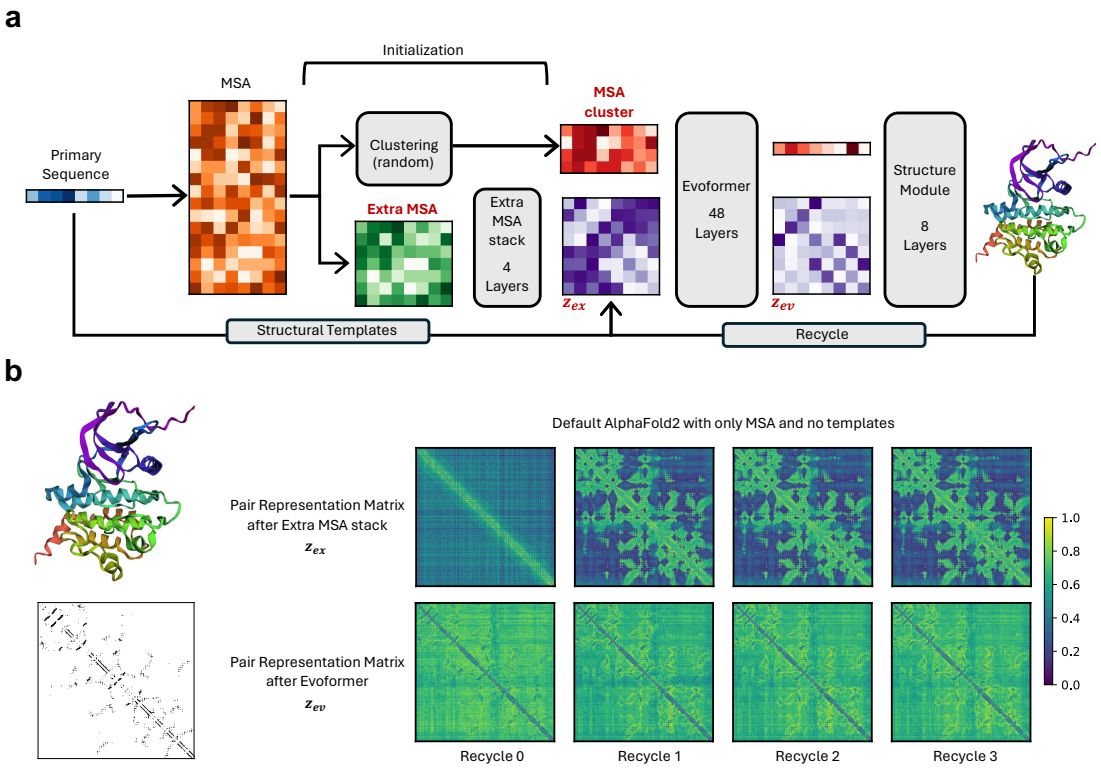


Figure 2.2: AlphaFold2 architecture a) Shows the architecture of AlphaFold2. The pipeline is drawn based on the architecture and algorithms presented by Jumper et al.[1]. I have expanded the initialization procedure presented in AF2. b) Using Abl1 Kinase as an example the heatmaps depict the role of recycle pre- and post- Evoformer block.

Given this brief overview, AF2 can be viewed as a deep-learning framework that integrates many ideas from earlier static structure prediction methods. Fig. 2.2a shows the overall architecture of the model. For a given query sequence, AF2 searches sequence and structural databases to construct an MSA and identify structural templates. Because attention over a long MSA is computationally expensive, AF2 separates the MSA information into an extra MSA representation, M_{extra}, and a clustered MSA representation, M_{cluster}. The extra MSA stack is used to initialize the pair representation matrix, z, while template information is also incorporated into the initial pair representation.

The clustered MSA representation, together with the pair representation, is then passed through the Evoformer module and subsequently to the structure module to generate the final three-dimensional structure. In this chapter, I denote the pair representation before the Evoformer as z_{ex} and the pair representation after the Evoformer as z_{ev}. The structure module initializes residues in a common reference frame and uses the learned sequence and pair representations to predict rotations and translations for each residue. Finally, the structure module outputs a three-dimensional structure, a residue-wise confidence metric - predicted local distance difference test (pLDDT) - and a distogram (distance histogram) for each residue pair. Thus, from a viewpoint of classical structure prediction, AF2 combines several of the methodological traditions:

- Like homology modeling and threading, it can use structural templates when available.
- Like coevolution-based methods, it uses MSAs to infer constraints between residues.
- Like refinement-based methods, it iteratively improves internal representations and coordinates through recycling.

2.1.2 Looking into AF2

Using Abl1 kinase as an example, Fig. 2.2b shows how the two different pair representation matrices evolve over recycling. With no recycling, z_{ev} already shows contacts similar to the ground-truth contact map shown below the structure. However, after the first recycle, it is noticeable that z_{ex} contains information about regions where contacts are absent. That is, a lower signal represents regions with no contacts, whereas a higher signal represents long-range contacts. In contrast, the lower signal in z_{ev} corresponds to regions where short-range contacts are present.

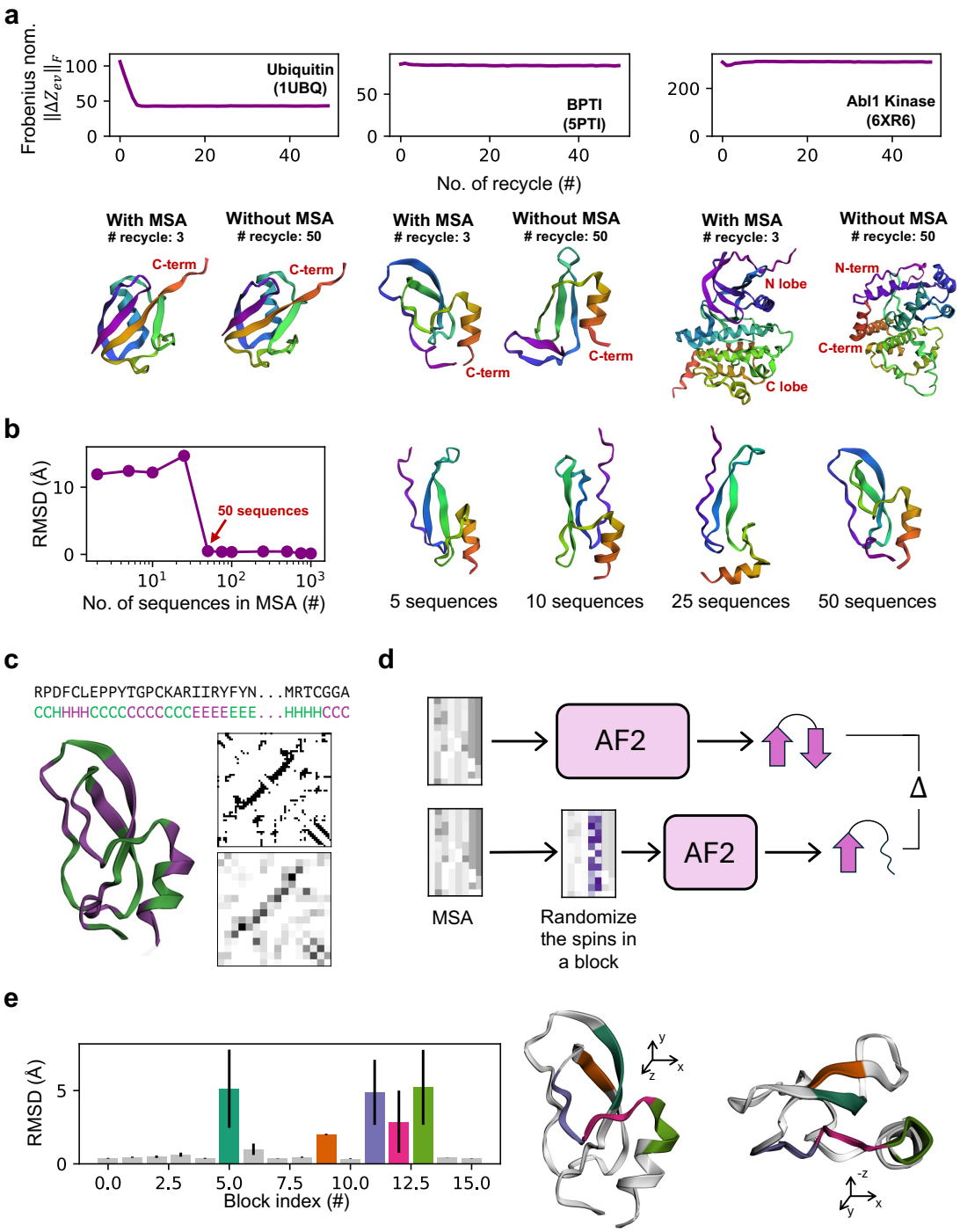


Figure 2.3: Requirement of MSA a) AF2 predictions using only sequence information, without MSA or template information, for ubiquitin, BPTI, and Abl1 kinase. These examples illustrate a simple way to assess whether AF2 requires MSA information for a given protein. b) AF2 predictions as MSA sequences are gradually added according to sequence similarity. c) Residue-blocking procedure used to coarse-grain the sequence representation for feature attribution. d) Schematic representation of the feature-attribution protocol in AF2. e) Feature-attribution results highlighting the important blocks in MSA that contribute to AF2 structure prediction.

This distinction between long-range, or global, and short-range, or local, information was made by Jumper et al. when comparing the roles of the Extra MSA stack and the Evoformer.

Using Abl1 kinase as an example the Fig. 2.2b shows how the two different pair representation matrix evolve over recycles. It is clear that with no recycle the z_{ev} shows contacts similar to the groundtruth contact shown below the structure. However, its easily noticable that after the first recycle, the z_{ex} contains where the contacts are not present. That is, a lower signal represents no contact region and a higher signal represent long range contacts. On the contrary, the lower signal in z_{ev} contains where the short range contacts are present. This long range(global) and short range(local) distinction is made by Jumper et al. when comparing the algorithms between Extra MSA stack and Evoformer.(Additional layer-wise analysis on Evoformer is presented in Fig. A.1)

2.2 How much coevolution is required?

One of the first question addressed in this chapter is how strongly AF2 depends on the MSA for different protein systems. One simple test is to remove the MSA input and increase the number of recycling steps. If a protein can still be predicted accurately without MSA information, this suggests that AF2 has learned a strong sequence-to-structure prior for that fold family or structural pattern. In such cases, recycling may allow the model to iteratively refine a plausible fold even when explicit covariation information is missing. This approach is similar to the one prescribed by Chang et al.[87]. However, it is different in the problem addressed. Chang et al. tried to predict folding pathway by using this approach whereas in this chapter we try to use this approach to see if MSA information required(Fig. 2.3a).

However, Fig. 2.3a shows the Frobenius norm of the difference between z_{ev} predicted by single-sequence AF2 at each recycling step and z_{ev} predicted with full MSA information. Systems such as ubiquitin, which is highly represented in the PDB, show little requirement for MSA information. As expected, larger molecules such as Abl1 kinase suffer from the lack of coevolutionary information. Interestingly, a smaller protein such as BPTI also suffers from the lack of coevolutionary information. Furthermore, in all these cases, while the tertiary structure predictions are not consistent, the secondary structure information remains consistent. This suggests that long-range residue-residue constraints may be required for better AF2 performance. These constraints can come from structural templates or from MSA information.

To further analyze how much sequence information in MSA aids AF2 in improving structure prediction, sequences were added to the MSA in their original order at different length, and structures were predicted using the resulting MSA matrices. Fig. 2.3b shows the RMSD between the predicted structure and the reference structure at different MSA depths. It is clear that approximately 50 sequences are sufficient for AF2 to predict the BPTI's structure accurately. This is also consistent with the study performed by Jumper et al. The predicted structure at each MSA depth up to 50 sequences is shown in Fig. 2.3b. Once again, while the secondary-structure information is preserved, long-range interactions are missing when a lower number of sequences is used.

Thus, an interesting question is which set of sequences, or what information within the MSA, is required for AF2 to predict the structure accurately. A related question is whether such feature attribution can help explain AF2's performance.

2.3 Block Perturbation of the MSA

To identify which parts of the sequence or MSA most strongly influence AF2's prediction, one can use feature-attribution tests commonly used in the explainable AI community. Methods such as LIME[88], TERP[89], and SHAP[90] perform local perturbations to the input and use surrogate models to explain the resulting variation in the output. In the same spirit, I propose a block-perturbation approach for MSA signal attribution. In this approach, the protein sequence is divided into blocks according to secondary-structure information from a structure predicted using the full MSA. This blocking procedure coarse-grains the sequence so that the perturbation effects are more pronounced. Fig. 2.3c shows an example of the blocking procedure for the BPTI system. For the BPTI system, each block was restricted to a size of three to four residues. Fig. 2.3d displays the schematic for the block-perturbation algorithm applied to BPTI. Practically, for a chosen sequence block, the corresponding columns in the MSA matrix are perturbed by randomly selecting rows(50%) and replacing their amino-acid identities with samples drawn uniformly from the set $\mathbf{S}$.

This experiment can be viewed as an evolutionary-information ablation test. If perturbing a block causes a large structural change, then AF2 is sensitive to the evolutionary constraints contained in that block. If the output remains nearly unchanged, then the block may not be essential for the predicted structure, or its structural information may be redundantly encoded elsewhere in the MSA or in AF2's learned prior.

Fig. 2.3e shows the result for BPTI. The root mean squared deviation(RMSD) changes caused by perturbation clearly show that certain blocks contain most of the relevant information. Based on this analysis, the selected blocks correspond to part of an antiparallel β-sheet

and a coil region spanning across the structure. Inspection of the sequence composition shows that these blocks include most of the hydrophobic core of BPTI. This suggests that AF2 requires MSA-derived information about the hydrophobic core for accurate structure prediction. This observation is also consistent with the previous hypothesis that, rather than requiring MSA information for secondary-structure prediction, AF2 requires MSA information mainly for long-range or tertiary contacts.

This interpretation should be made cautiously. Block perturbation does not directly identify thermodynamic importance; rather, it identifies model sensitivity, namely which parts of the evolutionary input most strongly affect AF2's output. Nevertheless, this analysis can be highly useful. In the current implementation, MSA signal attribution is performed as an ablation experiment. A more systematic approach would be to use linear surrogate models to quantify the importance of different sequence blocks. Additionally, instead of perturbing single blocks independently, pairwise block perturbations based on contact information may provide a more meaningful way to probe coupled structural signals.

2.4 Reduced MSA approach to AF2 (rMSA-AF2)

So far, I have focused on how much MSA information is required by AF2 and what kinds of MSA signals are useful for structure prediction. Based on the empirical results described above, MSA information appears to play an important role in resolving long-range tertiary contacts. However, to address the broader question raised in Chapter 1, we need a way to obtain an ensemble of structural hypotheses rather than a single static structure. The perturbation experiments suggest that changing the MSA input can lead to different tertiary structures. To address

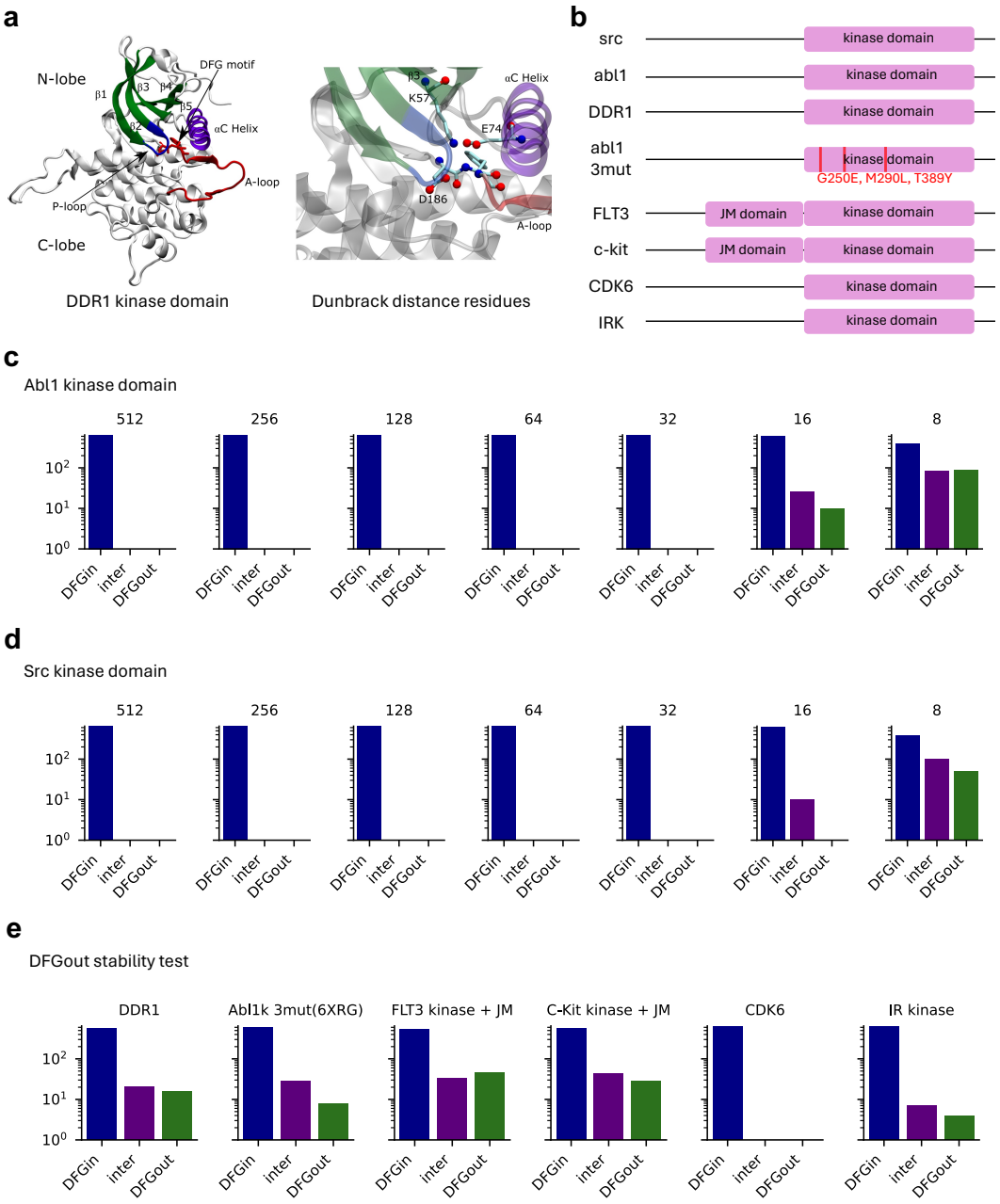


Figure 2.4: Reduced MSA approach to AF2 Structural description of the kinase domain using DDR1 kinase as an example. The right panel shows the key residues involved in the spatial classification of DFGin, DFGinter, and DFGout states. The distances between DFG-Phe-CZ and αC-Glu(+4)-CA, and between DFG-Phe-CZ and β3-Lys-CA, are used to define a two-dimensional space in which DFGin, DFGinter, and DFGout conformations are classified. b) Cartoon representation of sequence regions in kinase proteins. The blocks denote the sequence regions of interest for each kinase system. c) and d) Reduced MSA results for Abl1 and Src kinase domains. e) Reduced MSA results for kinase domains with experimentally known DFGout stable conformations. For all reduced-MSA studies, 640 structures were generated per hyperparameter.

this, I developed a reduced-MSA approach, similar in spirit to the MSA subsampling strategy introduced by del Alamo et al. around the same time [91]. In this approach, instead of directly perturbing the amino-acid content of the MSA, I reduce the number of MSA sequences used in the AF2 architecture. Specifically, the depths of M_{extra} and M_{cluster} are reduced during initialization. In the standard AF2 protocol, M_{cluster} has a depth of 512, while M_{extra} has a depth of 5120 in the original AF2 implementation. To observe conformational changes, these MSA depths must be reduced substantially.

I use kinase domains as examples to demonstrate the effect of the reduced-MSA approach. Kinase domains are important drug targets and are known to adopt different DFG conformational states, including DFGin, DFGinter, and DFGout states [8, 92](Fig. 2.4a). These states are discussed in more detail in Chapter 4. In this chapter, I focus on how reduced-MSA AF2 affects the predicted kinase conformations using DFG spatial classification (Fig. 2.4a). The reduced-MSA test was performed on two tyrosine kinase domains: Src kinase and Abl1 kinase. Fig. 2.4c and 2.4d show that alternative DFG states appear only after reducing the MSA depth to approximately 16 sequences. Here, MSA depth refers to the number of sequences retained in M_{cluster}. These results also show that AF2 preferentially generates DFGin-like kinase conformations compared with other DFG states, possibly reflecting the overrepresentation of DFGin conformations in the PDB database. Interestingly, irrespective of sequence identity, both Abl1 and Src kinase domains show similar populations at MSA depths of 16 and 8. However, the relative populations inferred from these reduced-MSA predictions should not be interpreted as thermodynamic populations.

To further examine AF2's apparent bias toward the DFGin state, I searched the PDB database for kinase domains with experimentally observed DFGout-stable structures. DDR1 kinase is one example of a kinase domain that prefers a DFGout conformation in the apo form. In

Abl1, three key mutations are known to stabilize the DFGout state, as observed in PDB structure 6XRG. In addition, the juxtamembrane domain located N-terminal to the kinase domain is known to autoinhibit some kinases and stabilize DFGout conformations. Examples of this category include FLT3 kinase (PDB: 1RJB) and the receptor protein tyrosine kinase c-Kit. A deeper search of the PDB also showed that CDK6 kinase and the insulin receptor kinase domain (PDB: 1IRK) were among the early deposited apo DFGout kinase structures. Therefore, I selected these kinase systems to cover multiple cases in which the DFGout conformation is experimentally stabilized.

Fig. 2.4e shows the reduced-MSA AF2 results for these kinase systems at an MSA depth of 16. Even for systems with experimentally observed DFGout-stable structures, the DFGin state remains highly populated in the AF2 predictions. Some kinase domains, such as CDK6, do not produce a DFGout conformation even at this reduced MSA depth. These results suggest that reduced-MSA AF2 can expose alternative conformational hypotheses, but the method is not an unbiased thermodynamic sampler. Instead, it should be interpreted as a way to perturb the evolutionary context provided to AF2 and test which conformational states are accessible to the network under reduced sequence information.

Overall, the reduced-MSA AF2 approach provides a simple way to introduce controlled stochasticity into AF2 predictions. By reducing the amount of evolutionary information or constraints passed through the MSA representations, AF2 can be encouraged to generate alternative structural hypotheses that may correspond to low-population or functionally relevant conformations. A recent study [93] has proposed that AF2 has indeed learned an energy surface, not a free energy surface, for protein folding in its transformer weights. They propose and provide significant evidence for the idea that the MSA pair representation matrix simply initializes close to the "correct minimum", while the transformer architecture performs an optimization step on

this energy surface. This also suggests that the MSA reduction protocol simply initializes the transformer closer to a different local minimum, possibly a non-native for the learned energy surface. However, the resulting structures should not be interpreted directly as Boltzmann-weighted ensembles.

2.5 Other related works

Based on the same general idea, several other approaches have been developed to introduce stochasticity into AF2. In some methods, the initial MSA is clustered into smaller MSAs based on sequence similarity, or randomly subsampled, to generate alternative conformations. In other approaches, information in the MSA matrix is perturbed by substituting alanine residues at selected sequence positions across different MSA sequences. The methods are summarized in Table 2.1.

Table 2.1: Conformational ensemble generation through perturbation-based approaches for AF2.

S.No.	Method	Description
1	rMSA-AF2[91, 94]	Reduces the depths of the internal MSA representations, M_{cluster} and M_{extra}.
2	AF-Cluster[95]	Clusters MSA sequences based on sequence-similarity metrics before passing them to the AF2 architecture.
3	CF Random[96]	Performs reduced MSA sampling across multiple MSA depths.
4	CF Random-expert[97]	Updates the MSA matrix with gaps based on prior structural or biological knowledge.
5	SPEACH_AF[98]	Perturbs the MSA matrix by substituting alanine residues before passing it to the AF2 architecture.
6	Frustratometer-based perturbation[99]	Clusters or selects sequences based on frustration scores obtained from the Frustratometer.

These approaches, especially the reduced MSA approach, rely on AF2's learned internal representations and structure module to generate structures that satisfy microscopic and geomet-

ric constraints beyond inter-residue contacts, such as chain connectivity, excluded volume, and physically reasonable dihedral angles. However, this is not always guaranteed. When the MSA cluster embedding or initialization contains limited information, the output structure can violate some of these requirements. Thus, geometry-based post-processing filters are essential when using these approaches.

2.6 Chapter Summary

This chapter connects Potts-model coevolution, masked-language-like objectives, and AF2-based structure prediction. Pseudolikelihood maximization in Potts models can be viewed as predicting each amino acid from its sequence context, making it conceptually similar to masked language modeling. Recent work on factored attention formalizes this relationship by showing how simplified attention mechanisms can recover Potts-like conditional distributions. AF2 extends this logic into an end-to-end structure prediction framework that integrates MSA information, structural templates, recycling, and geometric reasoning.

To conclude, the key architectural contributions of AF2 can be summarized as follows:

- AF2 uses row-wise and column-wise attention on the MSA representation, together with triangle attention on the pair representation, to refine residue-residue information and predict distograms.
- The structure module uses invariant point attention to predict three-dimensional atomic coordinates while preserving covalent bonds within an amino acid.
- The recycling procedure feeds predicted representations back into the network, allowing AF2 to iteratively refine its predictions.

This chapter then addresses the importance of MSA-derived signals for structure prediction. Although the feature-attribution approach introduced here is still at an early stage, it provides a useful way to probe which regions of the MSA influence AF2 predictions. A more systematic extension would involve linear surrogate models or related explainability methods to quantify signal importance more rigorously. Most importantly, this chapter also discusses how stochasticity can be introduced into AF2 through MSA reduction to generate diverse structural hypotheses. However, these reduced MSA AF2 structures should not be interpreted as Boltzmann-weighted ensembles.

Although they are not Boltzmann-weighted ensembles, reduced MSA AF2 predictions can produce structures with substantial conformational variance. In some cases, these structural differences may correspond to slow-timescale conformational changes or alternative metastable states. Therefore, rather than relying directly on AF2 confidence scores or AF2-derived ensemble weights, these structures can be used as structural hypotheses for adaptive sampling workflows designed to explore metastable states using molecular dynamics simulations. In this way, AF2 can be leveraged as a hypothesis-generating model whose outputs are refined and reweighted using physics-based methods. While any of the perturbation-based approaches discussed in this chapter may be useful, the following chapters primarily use the rMSA-AF2 approach. This chapter therefore provides the conceptual and methodological foundation for the post hoc physics-based sampling strategies developed later in the thesis.

Chapter 3: AlphaFold2-RAVE: From sequence to Boltzmann ranking

In this chapter, I introduce a method I developed to bridge AlphaFold2 (AF2)-based structural ensemble prediction with physics-based approaches for identifying and ranking metastable states. As discussed in Chapter 2, AF2 can generate structurally diverse ensembles when its MSA input is modified, but these structures do not provide relative thermodynamic probabilities. Moreover, MSA reduction can introduce thermodynamically unstable or physically improbable structures. Another limitation discussed in this chapter is AF2's limited ability to predict structural changes induced by missense mutations[100]. Assigning Boltzmann weights to AF2-generated conformations can help filter unphysical models, rank the remaining structures by thermodynamic stability, and address these limitations. As discussed in Chapter 1, Boltzmann-weighted metastable states can also aid in understanding molecular mechanisms and in structure-based drug discovery. In principle, sufficiently long unbiased molecular dynamics (MD) simulations should directly characterize the conformational diversity and thermodynamics of proteins without assistance from AI-based structure prediction. In practice, however, MD simulations are limited by accessible timescales, making it difficult to sample diverse protein conformations with statistical reliability [17]. To combine the strengths of both approaches, I propose a protocol that starts from a protein sequence and produces an ensemble of Boltzmann-weighted structures, where each structure is assigned a relative thermodynamic stability. I combine AF2-generated

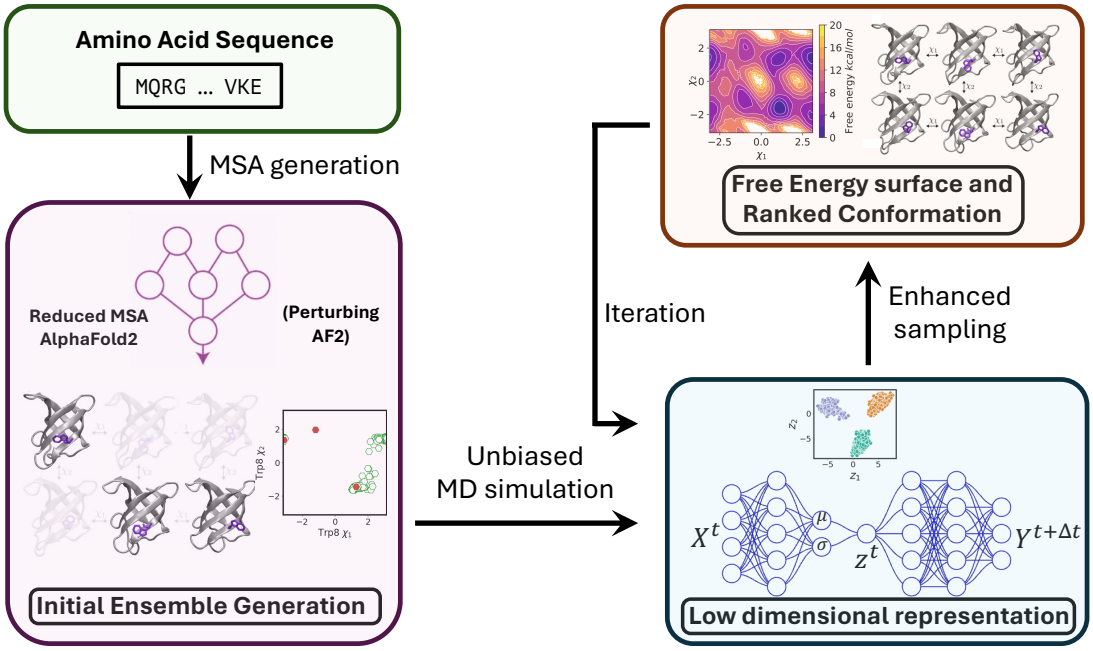


Figure 3.1: AlphaFold2RAVE pipeline: The schematic diagram presents the AF2RAVE pipeline using a 66 residue long cold shock protein(CSP) as an example. The initial ensemble generation block obtains distinct conformations from clustering the structures generated by perturbing AF2. Unbiased MD is used as a data augmentation tool for information bottleneck based dimensionality reduction. Enhanced sampling is performed to reweight the conformational distribution and rank each conformations.

structural hypotheses with the AI-augmented MD method "Reweighted Autoencoded Variational Bayes for Enhanced Sampling" (RAVE) [101], and refer to the final protocol as AlphaFold2-RAVE (AF2RAVE).

In this chapter, I apply AF2RAVE to proteins with distinct structural and functional challenges: the 1HZB cold shock protein (CSP) for side-chain rotameric metastability [102], the ubiquitin-binding protein UBA2 for disorder induced by missense mutation [100, 103, 104, 105], and the somatostatin receptor SSTR2, a G-protein-coupled receptor (GPCR), for conformational diversity [91]. For these systems, AF2 alone, even with reduced MSA input, is insufficient to provide thermodynamically ranked conformational ensembles. AF2RAVE reproduces known CSP benchmarks from experiments and specialized simulations, while also producing results for UBA2 and SSTR2 that are consistent with experimental observations and biological function.

In the following sections, I first describe the general AF2RAVE protocol using cold shock protein as a model system. I then show that AF2RAVE can identify and rank alternative conformational states. Finally, I apply the protocol to systems involving missense mutation, partial disorder, and larger-scale conformational changes.

3.1 Results

3.1.1 Establishing the AlphaFold2-RAVE Protocol

The key steps involved in the AlphaFold2-RAVE protocol are as follows: The central idea of AlphaFold2-RAVE is to use reduced-MSA AF2 to generate diverse structural hypotheses, and then use RAVE to learn reaction coordinates that separate metastable states. These coordinates are used for enhanced sampling, allowing the AF2-generated structures to be refined and ranked by their Boltzmann weights(Fig. 3.1).

1. Generate a conformationally diverse ensemble using the reduced MSA approach, or another perturbative approach described in Table 2.1.

2. Identify distinct representative structures from this ensemble and perform short MD simulations.

3. Use State-Predictive-Information-Bottleneck(SPIB) to determine a reaction coordinate that separates biologically relevant metastable states.

4. Perform enhanced sampling along the SPIB coordinate using WTMetaD or umbrella sampling.

5. Use the relative weights to rank the conformations.

Steps 1 and 2 are crucial because the subsequent SPIB learning and enhanced sampling steps depend strongly on the quality and diversity of the initial structures and trajectories. I use regular-space clustering to select representative structures from the AF2-generated ensemble. This procedure performs approximately uniform clustering in a high-dimensional conformational space. Since alternative conformations often appear with low frequency in the AF2-generated ensemble, regular-space clustering helps select these less-populated structures for downstream MD simulations. In Step 3, SPIB is used to learn a reaction coordinate that best separates the metastable states observed in the short MD trajectories. This SPIB latent space can then be used as the collective variable for enhanced sampling in Step 4. Because the SPIB coordinate is learned from short trajectories initialized from distinct AF2-generated conformations, the coordinate may not be fully converged after the first round. Therefore, iterating Steps 3 and 4 can improve both the learned coordinate and the sampling of metastable states.

3.1.2 Rotameric Metastability in Cold Shock Protein

Fig. 3.2 shows results for the 66-residue CSP (PDB: 1HZB). CSP has known rotameric metastability in its eighth residue, Trp8 (Fig. 3.2a), exhibiting six states in its χ_1 and χ_2 torsions through fluorescence spectroscopy [102]. To demonstrate the power of AlphaFold2-RAVE (Fig. 3.1), I assume no prior knowledge of any such special residues. Fig. 3.2b and Fig. 3.2c show probability distributions of conformations obtained from AlphaFold2 with reduced MSA length and from AlphaFold2-RAVE, respectively, projected in the space of SPIB coordinates. The enhancement in sampling quality is evident. Further analysis is provided in SI Section II,

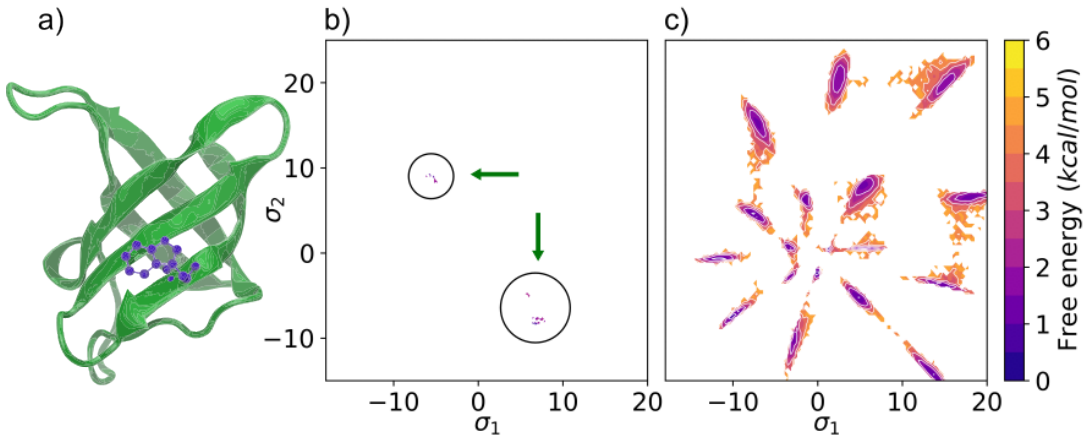


Figure 3.2: Rotameric metastability in cold shock protein: a) Results using AlphaFold2-RAVE on the 1HZB cold-shock protein (CSP). Structure of CSP with tryptophan shown in purple in its most stable rotameric form. Potentials of mean force (PMFs) projected along the two-dimensional SPIB coordinates σ_1 and σ_2 learned from my scheme: **b)** using structures obtained from reduced-MSA AlphaFold2, which provides additional conformational diversity compared with traditional AlphaFold2 but without thermodynamic reliability; **c)** using AlphaFold2-RAVE, showing both conformational diversity and thermodynamic ranking.

establishing that the enhanced sampling here conforms to the Boltzmann distribution (Figs. S1 and S2), and that my protocol is more general than focused sampling (Fig.S3).

3.1.3 Disorder in the Mutant Form of Ubiquitin-Associated Protein

Fig. 3.3c shows PMFs from AlphaFold2-RAVE for both WT and L355A UBA2. AlphaFold2 was recently shown to be unable to capture changes in native-state stability of a partially helical disordered structure caused by the missense mutation L355A [100]. For the WT, I find that the unfolded–folded state energy difference is -0.3 kcal/mol, while for the mutant the same difference is 1.2 kcal/mol. Calculation and further analysis can be found in SI Section IV. This shows that the mutation increases the disordered nature of the protein.

3.1.4 Helical Motion in a G-Protein-Coupled Receptor

The final system I study is the medically relevant [106] G-protein-coupled receptor (GPCR), for which AlphaFold2 is unable to capture conformational diversity [105]. Specifically, I explore conformational shifts in the somatostatin receptor SSTR2. Once again, I find that AlphaFold2-RAVE detects several local changes in side-chain positions, as well as larger-scale helical motions. This corresponds well with known evidence of GPCR structural diversity [2]. In Fig. 3.3a, I show examples of the large-scale helical shifts observed using this protocol. Fig. 3.3b shows the PMF projected onto the SPIB coordinates. Details on collective variables and further discussions can be found in SI Sections I-E and III, respectively.

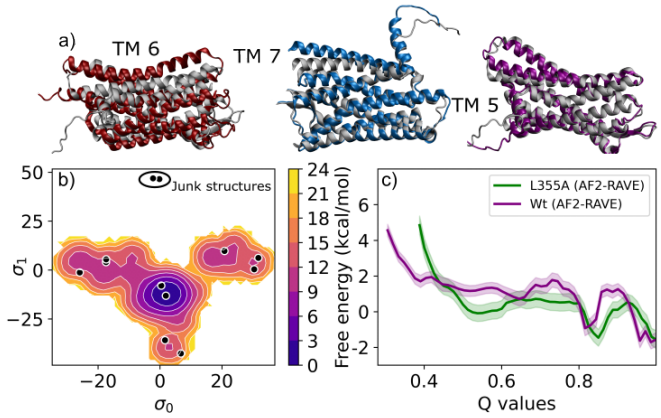


Figure 3.3: On missense mutation and conformational changes: a) SSTR2 structures highlighting helix movement in TM6, TM7, and TM5 via AlphaFold2-RAVE enhanced sampling, in agreement with experiments [2] and biological function. Low-energy structures are shown in color and overlaid on the native structure shown in gray. **b)** SSTR2 PMF using AlphaFold2-RAVE, showing multiple potential states. Black dots show structures obtained from reduced-MSA AF2, with structures marked as "junk" representing CV regions that cannot be sampled through transitions. **c)** UBA2 PMFs along total Q-value for WT (purple) and L355A (green). The WT shows a visibly higher barrier, whereas L355A has a wider and more stable disordered region.

3.2 Conclusion

To conclude, I propose the AlphaFold2-RAVE method, which combines the strengths of AlphaFold2 with the all-atom enhanced-sampling capabilities of RAVE [101]. This combination provides an effective reaction coordinate, learned through SPIB, for a given protein sequence and uses it to discover and differentiate biologically relevant conformational ensembles. A converged enhanced-sampling method, such as metadynamics along the SPIB coordinate, can then rank these conformations with their correct Boltzmann weights. Even before full convergence, significant information can be gained about the relative importance of metastable structures, which can help refine inputs for structure-based drug design methods [107, 108]. To some extent, this approach relies on the ability of my modified version of AlphaFold2 to capture structural diversity. For rare proteins with no evolutionary homologs exhibiting structural diversity, this diversity may not appear in the AlphaFold2 output. However, I find that the learned information bottleneck can still guide sampling toward regions that were not directly represented in the AlphaFold2 output, as observed in both CSP and UBA2.

While any sampling protocol could in principle be implemented here, I use RAVE, specifically the SPIB variant, because it requires minimal hand-tuning and limited prior knowledge of metastable states or latent spaces for driving sampling. RAVE itself can be combined with many rare-event sampling methods, each with its own advantages and disadvantages. Here, I demonstrate the protocol using metadynamics because of its relatively straightforward implementation. Alternatively, this protocol could incorporate other collective-variable-based methods, such as umbrella sampling [66], weighted ensemble [109], NEUS [110], or the string method [111]. Similarly, instead of AlphaFold2, one could use other experimental [112] or computational [10, 113]

approaches to generate initial dictionaries of possible conformations. I use AlphaFold2 because of its simplicity, accuracy, and relative ease of use.

Overall, I present a protocol that combines three components, sequence-based structure prediction, machine-learning-based collective-variable discovery, and molecular dynamics, to learn more than what each component can provide alone. This pipeline now can further be used on immunologically important systems such as kinase molecules. In the next couple of chapter I go back to the kinase domain protein molecule to improve structure based drug design by including ensemble perspective of protein molecules.

Chapter 4: Exploring kinase Asp-Phe-Gly (DFG) loop conformational stability with AlphaFold2-RAVE

*[This chapter is based on: Bodhi P. Vani, **Akashnathan Aranganathan** and Pratyush Tiwary. (2023). Exploring kinase DFG loop conformational stability with AlphaFold2-RAVE. Journal of Chemical Information and Modeling. Vol. 64, Issue 7, pp. 2789–2797. DOI: 10.1021/acs.jcim.3c01436 [114]. I contributed in the structure generation, SPIB calculation including simulations and feature dropout, and methodology development. I provided the codes for AF2-RAVE]*

In this chapter, I return to the kinase domain proteins introduced in Chapter 2, focusing specifically on DDR1 kinase and its mutants. I apply the AlphaFold2-RAVE framework established in chapter 3 to the DFG-loop conformational landscape of kinase domains. Kinases are among the most important therapeutic targets in structure-based drug discovery because they regulate many cellular signaling processes and are frequently implicated in cancer [115, 116]. Although several effective cancer therapies already function by targeting and inhibiting kinases [116], kinase-based therapeutics remain far from fully explored [117].

In their active state, protein kinases catalyze the phosphorylation of substrate proteins through the transfer of the γ-phosphate group from adenosine triphosphate (ATP) or guanosine triphosphate (GTP). Often, the substrate protein is another in a cascade of kinases required for cell signalling[118, 119]. While there are over 500 kinases in the human kinome, making them

a challenging class to study, the protein kinome universally has some highly conserved structural motifs with a structurally well characterized active state. One key motif for this characterization is the Asp-Phe-Gly (DFG) motif in the activation loop(Fig. 4.1). This motif has two structural conformers, one with the Asp pointing into the loop, the DFG-in or active conformation, and one with it pointing out into the solvent, the DFG-out or inactive conformation. The ATP binding, and hence phosphorylation catalysis can only occur in the DFG-in conformation. Most drugs targeting kinases are "ATP competitive", i.e. they bind to prohibit ATP binding. These ATP-competitive drugs themselves are classified mainly in two types, either binding to the active site in DFG-in conformation, hence inhibiting the binding of ATP, or binding to the DFG-out conformation and hence stabilizing the inactive state. However, the presence of kinases in many essential cellular functions and their homologous nature makes specificity and efficacy particularly hard to achieve. Often, we are interested in drugs that bind preferentially to specific kinases *without* affecting other kinases. Given this, the characterization of diverse inactive states is of particular importance, both in terms of a structural understanding of these states, as well as knowledge of the thermodynamic stabilities relative to active state. For instance, given a target kinase, finding a uniquely stable inactive state with a novel binding site could lead to a more specific, less promiscuous (and hence toxic) drug. However, the most robust way to do this computationally, obtaining a MD trajectory that traverses the space of active and inactive states multiple times, is impossible. Additionally, the transition from DFG-in to DFG-out is highly non-local and a concerted combination of several long-range and large-scale motions, so that characterizing it and studying it even with the aid of enhanced sampling is difficult[120, 121].

Another important covariate to study kinase conformational ensembles is that of point mutations, which are often the cause of incorrect signalling leading to pathology. One key reason for

this is that the balance of probability between active and inactive conformations is delicate and often flipped on changing single residues, as shown by the system we have chosen in this work, DDR1.

The AF2-RAVE based approach for solving this problem involves combining structural ensembles obtained from AF2 with a machine learning algorithm to learn order parameters for biasing. To demonstrate this, we use the discoidin domain receptor tyrosine kinase 1 (DDR1), which in wild-type is more stable in an inactive DFG-out conformation. However, in several single site mutations, specifically D671N, Y755A and Y759A, the relative DFG conformational stabilities are flipped. One of the reasons we choose this set of systems is that a recent paper[4] provides extremely detailed and valuable work on their DFG stabilities, with atypically long unbiased MD trajectories. We show that our results agree with theirs qualitatively, in that we predict the flipping of DFG conformation preference on mutation. While we sacrifice some accuracy qualitatively, our method is faster by roughly 2-3 orders of magnitude (exact simulation lengths described in Results), and has the potential of being reused without relearning a lot of the information.

4.1 Results and Discussion

Similar to the protocol mentioned in the chapter 3, I start by using reduced-MSA AlphaFold2 to generate a diverse set of initial conformations and cluster them using regular-space clustering. I choose this clustering method because the reduced-MSA AF2 outputs tend to be sparse in regions of interest, and I want to prioritize the tails of the distribution over highly sampled regions. The centers of these clusters are then used to seed unbiased MD trajectories, which

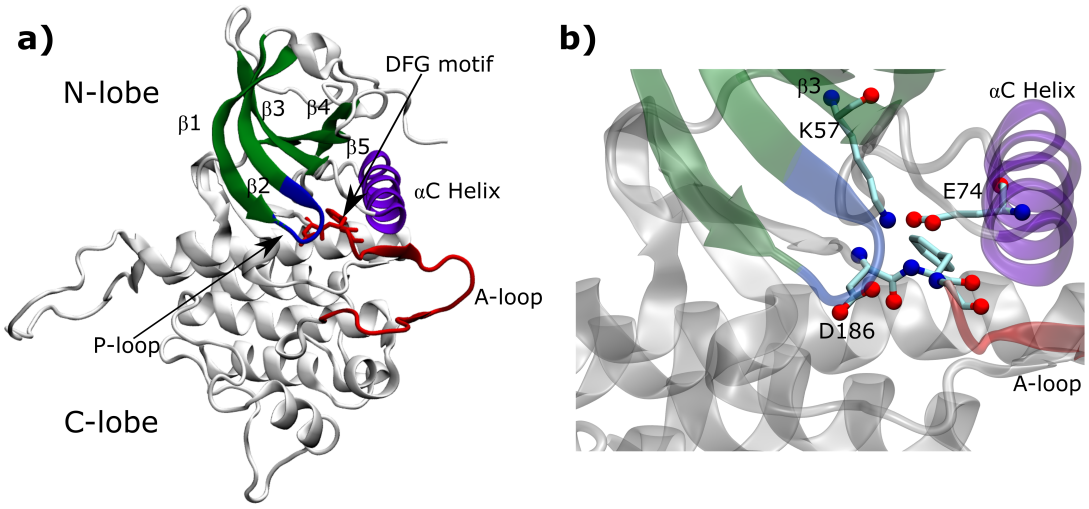


Figure 4.1: a) Structural anatomy of the DDR1 kinase molecule showing the activation loop (A-loop) (red), Gly-rich P-loop (blue), αC-helix (purple), and the characteristic N-lobe β sheets (green). These motifs are relevant to the DFG-in to DFG-out transition. **b)** The conserved salt bridge between K57 in β_3 and E74 in the αC-helix that is crucial in ligand dissociation and basic kinase function [3].

are the input trajectories for RAVE. This learns a two-dimensional reaction coordinate expressed as an information bottleneck. I then run well-tempered metadynamics biased along this two-dimensional reaction coordinate. Finally, I use the metadynamics-learned bias for the wild-type protein as a static bias to sample distributions for the wild-type and mutant sequences.

I discuss the results in two sections. First, I discuss the modified AF2 outputs for all four sequences and the process of learning a biasing potential. Next, I discuss results from biased dynamics for DDR1 WT and mutants.

When I refer to kinase DFG-in, DFG-out, and DFG-inter structures, I use the Dunbrack method [92] for classification using distance cutoffs. Representative structures for these states are shown in Fig. 4.2. When referring to dimensions or RCs that are learned through RAVE, I label them as information bottlenecks (IBs) and assign variables σ.

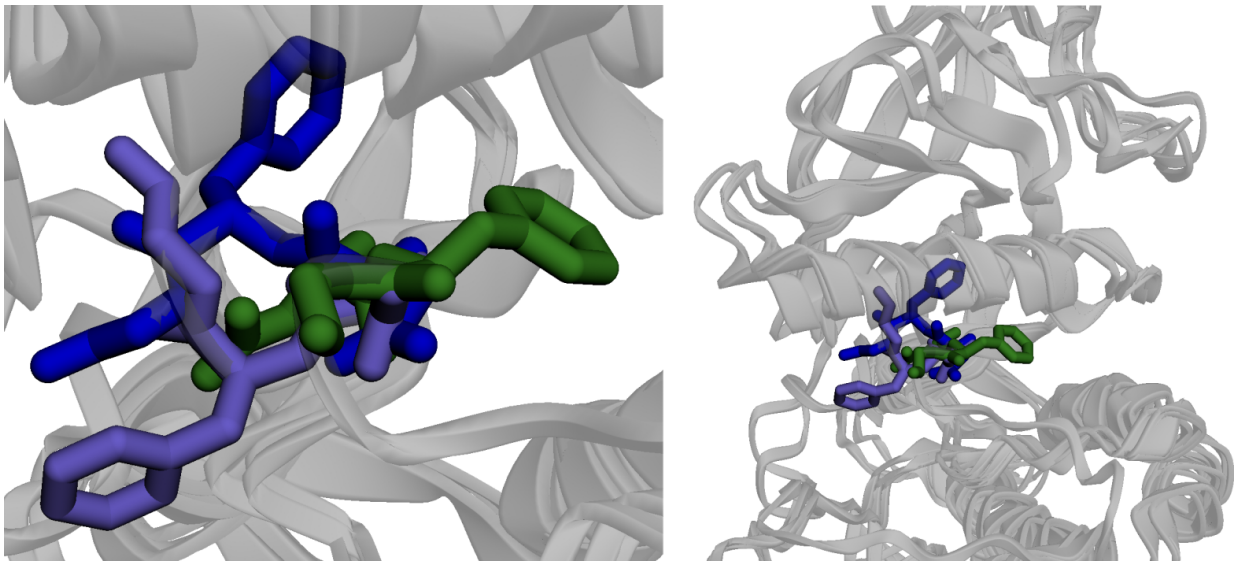


Figure 4.2: Representative structures from reduced-MSA AF2 for the DFG-in (purple), DFG-inter (blue), and DFG-out (green) states shown in two views and superimposed on the same structure.

4.1.1 Learning a Bias Potential from AF2-RAVE on WT DDR1

I first generate 1280 structures for each kinase: 640 for MSAs of depth 8 and 16, with 128 random seeds generating 5 structures each. AlphaFold2, even with the reduced MSA approach, is unable to distinguish between the expected conformational diversity of these four sequences, giving effectively identical results for all of them. This can be seen from Fig. 4.3a, where I show the populations of active, inactive, and the known intermediate or transition state, referred to as the "DFG-up" or "DFG-inter" state, after filtering out obviously unphysical structures, such as structures with broken bonds. These structures are classified using the Dunbrack classification described below. Note that although a significant DFG-out population is observed, even in the wild type, which is known to have higher inactive-state stability, AF2 predicts the more commonly observed DFG-in structure. These results disagree with the population densities implied by previous long MD simulations of the same kinase, despite my thorough search of MSA hyperparameters to force increased structural diversity [4].

The transition state is currently commonly called "inter" [92], as it has been consistently found to be necessary for observing the DFG-in to DFG-out trajectory. Previously, it was referred to as "DFG-up" because of the upward-pointing position of the Asp residue side chain in the traditional structural view, with the N-lobe above the C-lobe. The previously named "DFG-down" position is referred to as unassigned, as it is a high-energy and physically unlikely structure. The fact that the transition state is observed from reduced MSA AlphaFold2 is extremely significant and useful, as previous attempts to use RAVE only with crystal structures of DFG-in and DFG-out states had failed, because simulations tend to move toward and then become trapped in the DFG-down configuration. I show further analysis of the AF2 reduced-MSA outputs in the SI.

Next, SPIB is used on these structures to propose possible RCs for biasing, with a lag time of 20 ns. The active-inactive conformational transition is highly delocalized and requires an understanding of several intramolecular interactions. In particular, the αC-helix forms a conserved salt bridge with the $\beta3$ strand that plays a crucial role in the molecule's transition [3, 8]. Further, globally, the opening and closing of the two lobes, N-lobe and C-lobe, define the active-inactive transition [16]. From a dynamics perspective, the main motifs involved in this transition include the activation loop (A-loop), P-loop, and αC-helix, which mainly belong to the N-lobe (Fig. C.1). To this end, I include several other CVs, described below, including the distances used for the DFG classification above. These CVs roughly correspond to those used in several previous studies focusing on different parts of the kinase molecular structure [7, 92, 122, 123]. In Table C.1, I list all the residues involved in distances that I consider, indicating my acronym, the conserved residue if applicable, the residue index for DDR1 numbered from 1, and a description of the motif it belongs to. In Table C.2, I list all the distances used as initial inputs for SPIB and indicate them visually in Fig. C.1.

In this nomenclature, the Dunbrack distances, which I use to project my final potentials of mean force (PMFs), are ChelX CA–DFGPhe CZ and sbridgeK CA–DFGPhe CZ. Referring to these as d_1 and d_2, the structures are classified as follows: DFG-in if $d_1 < 11$ and $d_2 > 11$, DFG-out if $d_1 > 11$ and $d_2 < 14$, DFG-inter if $d_1 < 11$ and $d_2 < 11$, and unassigned if $d_1 > 11$ and $d_2 > 14$.

In Fig. C.2, I show the distributions of AF2 outputs for all four DDR1 sequences over the full set of input CVs. I find that they all look quite similar, despite the known differences among these mutants. This is interesting, considering that AF2 is usually expected to perform more confidently on evolutionarily faithful sequences. However, in this case, I suspect that the unusual DFG-out stability of DDR1 contributes to the easy access it has to conformational diversity. Nevertheless, my AF2 outputs still predict higher stability for the DFG-in structures across all DDR1 sequences.

To learn a biasing potential, I run multiple metadynamics trajectories in parallel that share bias potentials. This is done only for the wild-type sequence, and the same bias is then used for further calculations on all sequences. These initial structures are chosen as in the original AF2-RAVE paper [94], as follows. First, I run AF2 using ColabFold with manually set MSA depths of 8 and 16. Next, I run regular-space clustering with a minimum distance parameter of 9 on standardized CV values, using the set of CVs described above.

I use the following stopping conditions for the initial metadynamics runs: (1) if the walker starts in a DFG-in or DFG-out structure, it must reach the opposite structure; and (2) if the walker is not initially in one of the two main metastable states, it must reach one of them. I stop only if the transition remains stable for 1 ns during the biased simulation. These simulations are between 5 ns and 20 ns long.

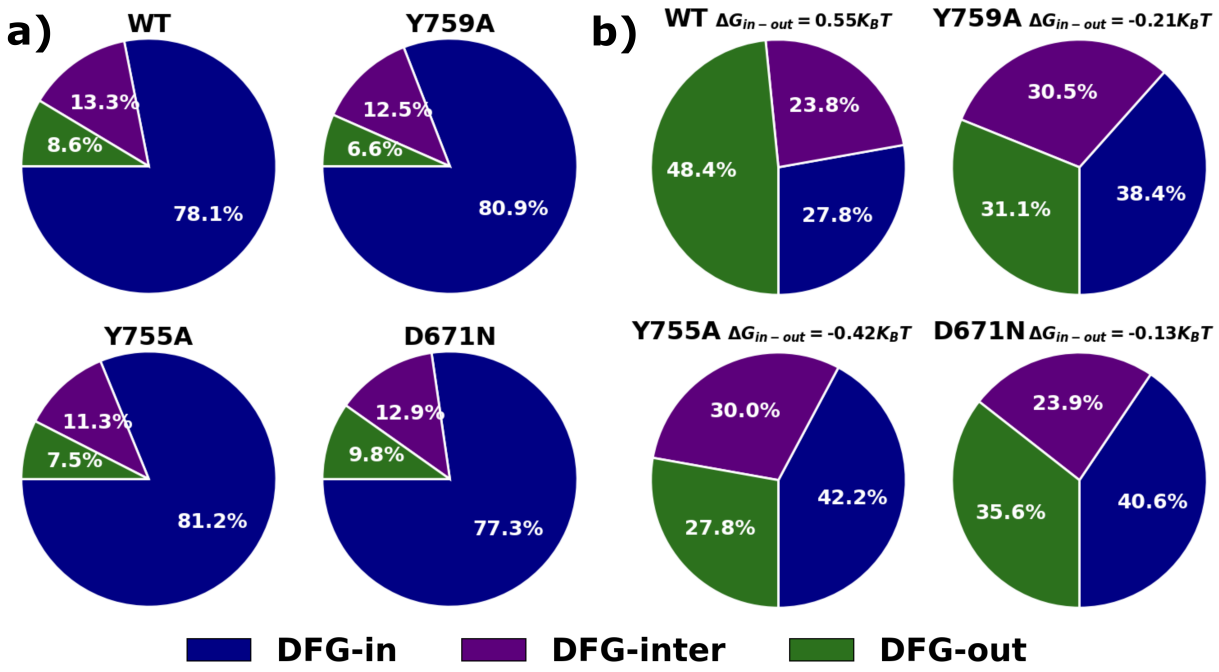


Figure 4.3: Populations of active, inactive, and the known transition state, referred to as the "DFG-up" or "DFG-inter" state, for wild type and mutants D671N, Y755A, and Y759A. **a)** Populations obtained through reduced-MSA AF2, combining MSA depths 8 and 16. **b)** Populations obtained using my AF2-RAVE protocol. I clearly observe that AF2 alone is unable to distinguish between wild type and mutants, and in particular gives the wrong order of stability for the mutants. In contrast, AF2-RAVE is able to identify the reversal in stability upon point mutation and gives more thermodynamically representative populations for these states, in excellent agreement with benchmark calculations performed by Hanson et al. [4] using unbiased MD simulations.

In Fig. C.4, I show the bias learned for the process. To demonstrate the transferability and efficiency of my protocol, I learn an IB and a bias potential using only the wild type. This also provides a basis for learning universal biases for systems with transferable CVs, potentially allowing more efficient sampling for homologous families of proteins.

4.1.2 Biased Dynamics on DDR1 Mutants

In this study, my main result is the prediction of the relative stabilities of the DFG-in and DFG-out structures for wild-type DDR1 and its mutants. In Fig. SC.5, I show the predicted PMFs for these structures. I compute DFG-in versus DFG-out relative thermodynamic stability

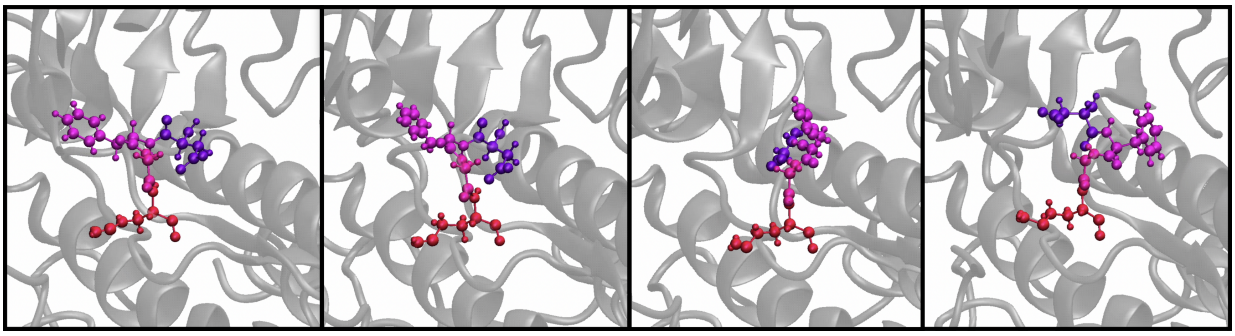


Figure 4.4: An example trajectory for the DFG-out to DFG-in transition obtained using AF2-RAVE, with residues D (in red), F (in purple), and G (in blue). See Code Availability for details of an example video showing the transition.

by integrating probabilities over Dunbrack definitions for kinase structural states and calculating the ΔG between these states: (i) WT: 0.5 K_BT, (ii) D671N: -0.2 K_BT, (iii) Y755A: -0.42 K_BT, and (iv) Y759A: -0.13 K_BT. Each of these values was computed with 5 trajectories in each direction and had standard deviations of (i) WT: 0.23 K_BT, (ii) D671N: 0.12 K_BT, (iii) Y755A: 0.29 K_BT, and (iv) Y759A: 0.11 K_BT. This flipping in the Boltzmann ranking of active and inactive states is consistent with previous findings [4], which were obtained using long unbiased MD simulations. I also integrate over my reweighted data to obtain the Dunbrack populations for comparison with those from AF2, shown in Fig. 4.3. Each run is approximately 20 ns long, and I use 5 runs in both directions for each structure to compute my PMFs and free-energy differences.

In Fig. 4.4, I show snapshots of one example sampled trajectory. Some salient details of the transition were also observed. I find that, during the transition, the breakage and formation of the salt bridge is clear, and the large-scale upward motion of the αC-helix is essential for sampling dynamics, as noted previously.

4.2 Conclusion

In this work, I extend my previous AF2-RAVE protocol to obtain Boltzmann-ranked conformational diversity in protein kinases, specifically from the perspective of the pharmaceutically relevant and evolutionarily conserved DFG loop. The key differences in the protocol are as follows: (i) I demonstrate that a static bias learned using metadynamics can be an effective method with advantages over a dynamic bias; (ii) I learn not only a transferable SPIB, but also a bias that is transferable across homologous systems; (iii) I use the SPIB step only once, as reduced-MSA AF2 is able to provide sufficient information for the level of detail I am seeking; and (iv) here collective variables as input were handpicked as I aim to characterize and quantify a specific known transition rather than explore unknown conformational spaces.

I have previously shown that the protocol can capture a versatile range of transitions, including rotameric metastability, large-scale helical motions, and partial disordering [94]. Here, I study the DFG-loop conformational change, which is of strong therapeutic importance. Moreover, I choose DDR1 because there exists an unusually thorough sampling study with unbiased trajectories by Hanson et al. [4], which provides a more robust comparison than enhanced-sampling-based work on large biomolecules is usually afforded. With AF2-RAVE, I obtain the same conformational ranking and similar thermodynamic stabilities as found by Hanson et al. [4] for the DFG-in versus DFG-out conformations of DDR1. However, the total MD simulation time in my work is around two to three orders of magnitude shorter.

It is important to note that the trajectories rely on collective variables that allow the sampling of the DFG-inter transition state. Without AF2 structures to seed the initial unbiased trajectories, the collective variable learned usually leads to a DFG-down structure, resulting in an

unsuccessful transition trajectory. This emphasizes the importance of AF2-related methods as reliable generators of initial structural hypotheses. These generated conformations can act as 'breadcrumbs' that guide SPIB toward learning a more informative reaction coordinate.

This approach can be repeated for other kinases as well. Most importantly this approach can be used to sample novel states that can then used as relatively high-throughput inputs for cryptic pocket prediction algorithms [124](see SI). That means given the protocols ability to sample, rank states, we are one step closer to use conformational selection hypothesis for structure based drug discovery. In the next chapter I will discuss about the structure picking and docking protocol that is required in the downstream of AF2-RAVE.

Chapter 5: Empowering AlphaFold2 for protein conformation selective drug discovery with AlphaFold2-RAVE

*[This chapter is based on: Xinyu Gu, **Akashnathan Aranganathan** and Pratyush Tiwary. (2024). Empowering AlphaFold2 for protein conformation selective drug discovery with AlphaFold2-RAVE. eLife. Vol. 13, pp RP99702. DOI: 10.7554/eLife.99702.1[125]. I contributed to the methodology development and reaction-coordinate design through SPIB. I also performed the parameter search needed to develop a template-only version of AF2. Finally, I performed MSA analysis to interpret the special behavior of DDR1 from AF2's perspective. I provided the codes for AF2-RAVE]*

In the previous chapter, I showed that AF2-RAVE can be used to obtain Boltzmann-ranked conformational ensembles for kinase DFG-loop motions. In particular, I demonstrated that AF2 generated structures can act as reliable initial structural hypotheses, while physics-based sampling and reweighting are required to assign thermodynamic importance to the resulting metastable states. In this chapter, I extend that framework to the next question: can AF2-RAVE ranked kinase conformations be used for docking and structure-based drug discovery?

As discussed in Chapter 1, the conformational selection hypothesis suggests that ligands may preferentially bind and stabilize pre-existing metastable conformations in the protein ensemble. For structure-based drug discovery, this means that identifying and ranking metastable

protein conformations can be useful only if these conformations also provide ligand compatible binding pockets. Therefore, after using AF2-RAVE to enrich and rank metastable structures, a docking protocol is required to test whether the selected conformations are suitable for ligand binding.

In this chapter, AF2-RAVE is combined with docking methods from the Schrödinger suite to create an AF2-RAVE Glide workflow(Fig. 5.1). The role of AF2-RAVE is to generate and rank metastable protein conformations, while the role of docking is to evaluate whether these conformations can accommodate known ligands. I primarily use Glide XP [126, 127, 128] and Induced-Fit Docking (IFD) [129, 130] to dock ligands into selected kinase structures. Recent generative docking methods such as DiffDock [131] and DynamicBind [132] provide alternative approaches, but here the focus is mainly on Glide and IFD because they provide a standard and widely used framework for retrospective docking comparisons.

Once again protein kinases are used to demonstrate this workflow - Abl1, DDR1, and Src. These systems are particularly useful because kinase inhibitors are often classified by the conformational state they bind. The human genome contains about 500 protein kinases which share highly conserved structures in their catalytic ATP-binding pocket, due to the selection pressure towards functional catalysis. This poses significant challenges in developing selective small-molecule ATP-competitive kinase inhibitors, as they must effectively target the intended kinase while avoiding off-target interactions and associated side effects. Research efforts aimed at achieving selectivity have led to the development of 4 distinct types of kinase inhibitors,[133] including two types of ATP-competitive kinase inhibitors: type I inhibitors bind to the ATP-binding site adopting the active catalytic state, while type II inhibitors target the binding pockets adjacent to the ATP-binding site adopting the inactive state. These two states primarily differ in

the configurations of their activation loop (A-loop), which is a flexible loop of approximately 20 residues. In the active state, the A-loop is "extended" to create a cleft for substrate peptides to bind, while in the inactive state, it is collapsed or "folded" onto the protein surface, blocking substrate binding. Additionally, in the active state, the three residues 'Asp-Phe-Gly' (DFG motif) at the N-terminus of the A-loop bind to an ATP-binding Mg^{2+} ion, with the Asp side-chain pointing inward to the ATP-binding pocket, while in the inactive state, the Asp side-chain is flipped outward, and the DFG-motif adopts the DFG-out conformation.[8] Henceforth, we will refer to the type I inhibitor binding state as the active DFG-in state and designate the type II inhibitor binding state as the classical DFG-out state, as previously proposed in the literature[134, 135, 136] (demonstrated in Fig. 5.2 A and B using Abl1 kinase as an example).

The chapter begins by showing that default AF2 structures of Abl1 and DDR1 are insufficient for docking type II kinase inhibitors, even when standard docking protocols are used. Then the chapter examines whether reduced MSA AF2 can generate decoy structures that resemble type-II inhibitor compatible DFG-out states. For DDR1, reduced MSA AF2 produces a small number of DFG-out-like structures, but only a subset of these structures are suitable for docking. I then show that AF2-RAVE can enrich these holo-like structures by ranking them using their relative free energies. Finally, DDR1 derived conformational information is used to guide structure generation and docking for homologous kinases such as Abl1 and Src, demonstrating how AF2-RAVE ranked metastable states can be transferred across related kinase systems.

A key point here is, that apart from the DFG loop flip seen in the previous chapter I also explore the conformational change of the entire A-loop. Additionally, I show that DDR1 is the only kinase out of the three that gives all biologically relevant states. Thus, I study DDR1 in great detail to transfer the learned information to the other homologs. I also have identified the reason

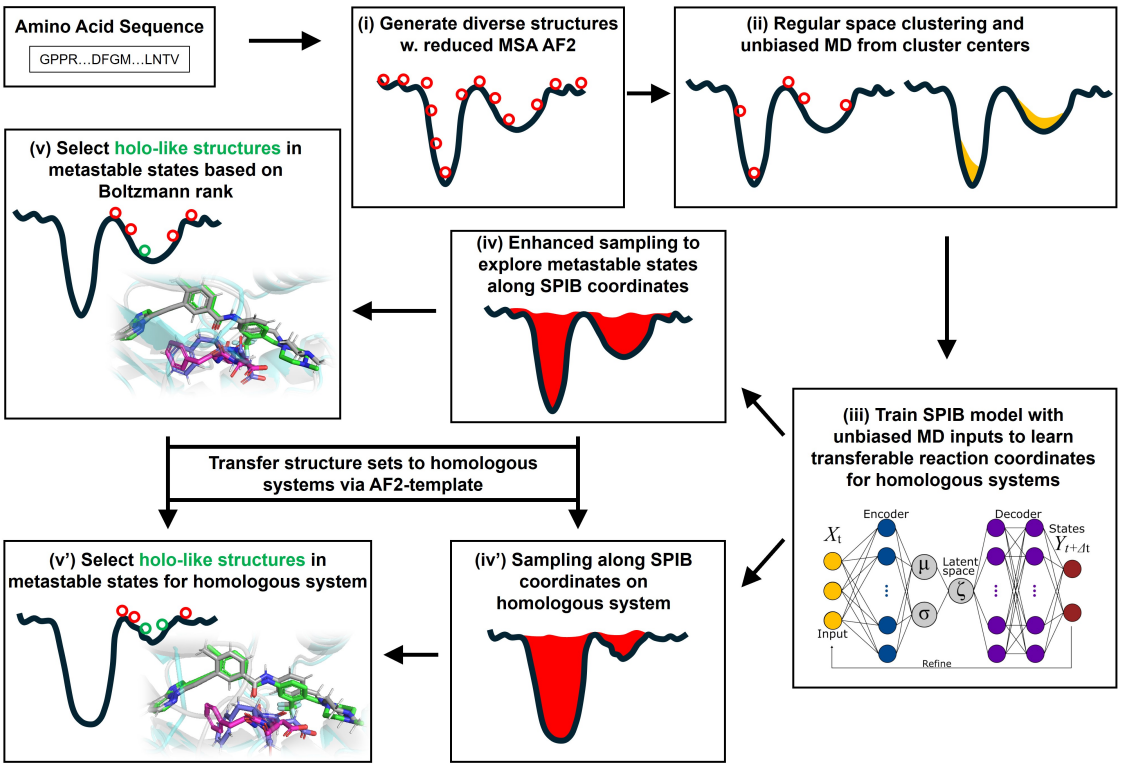


Figure 5.1: A schematic of the AF2-RAVE Glide workflow: (i) decoy structures generated by reduced MSA AF2, (ii) regular space clustering and unbiased MD simulations starting from cluster centers, (iii) State Predictive Information Bottleneck model (SPIB, a RAVE variant) to learn reaction coordinates from unbiased MD, (iv) enhanced sampling runs to calculate free energy landscape, (v) distinguish holo-like structures from decoys in metastable states based on Boltzmann rank and conduct Glide or Induced-fit Docking (IFD) on holo-like structures for ligands targeting metastable states. (iv') and (v'): the decoy structure set and the learnt SPIB coordinates are transferable to homologous systems.

why reduced MSA approach for AF2 generates more states for DDR1. Once again the key is in the MSA signals.

Overall, this chapter presents AF2-RAVE Glide as a sequence to docking protocol for conformational selection based drug discovery. The central idea is that AF2 based methods can generate structural diversity, AF2-RAVE can assign thermodynamic relevance to the generated conformations, and docking can test whether the top-ranked metastable states provide ligand compatible pockets. This connects the sequence-ensemble-function perspective developed in earlier chapters to a practical structure based drug discovery application.

5.1 Results and discussion

5.1.1 AF2 structures fail to dock with type II kinase inhibitors

The utility of AF2-generated structures for structure-based drug discovery and virtual screening campaigns has been a subject of controversy and skeptical enthusiasm.[137, 138, 139, 140] For proteins lacking crystal and cryo-EM structures but with homologous structures available in PDB, such as CDK20 kinase, AF2 demonstrates effectiveness as a homology modeling method for generating initial structures suitable for subsequent virtual screening. [141]. Other AF2-based homology modeling method can bias AF2-generated structures towards user-selected template structures with specific druggable conformations.[142] However, a significant drop in the hits enrichment factor during virtual screening has been reported when employing AF2 structures as rigid receptors for docking, compared to using holo PDB structures. This occurs even in cases where the binding pockets of AF2 structures differ only slightly at 2 to 3 residues, from those of holo PDBs.[139, 140] Given that ligands can induce slight relocation and side-chain rotation of pocket residues upon binding, it is important to note that AF2 does not account for this induced-fit effect, as it does not encode co-factors like ligands. Therefore, it appears necessary to perform ligand induced-fit modeling or relaxation on AF2 structures before engaging in any further structure-based drug design. Molecular dynamics (MD) simulations biased towards adjacent holo-template structure have proved effective in refining apo structures and improving their early enrichment performance.[143] It has also been demonstrated that AF2 structures can achieve comparable accuracy to crystal holo structures in Free Energy Perturbation (FEP) calculations, by superposing AF2 structures with crystal structures, grafting the co-crystallized ligands

onto the AF2 structure, and optimizing the AF2 structure/ligand complex to account for subtle induced-fit effects.[144] AF2 structures, decorated with ligands from template ligand grafting method or known-hits docking method, and further refined by Schrödinger IFD-MD protocol, exhibit promising performance in early enrichment[145] and the prediction of novel ligand/protein complex structures.[146] Therefore, if large backbone motion is not required, it appears feasible to refine the apo pockets in AF2 structures into holo-like pockets for structure based drug design. However, when AF2 structures exhibit significant steric clashes with holo ligands, especially in cases ligands targeting metastable states, direct application of out-of-the-box AF2 structures in docking methods for virtual screening and early enrichment may pose more challenges. The AF2 predicted kinase structures predominantly exhibit the DFG-in state, with over 95% of human kinases predicted in this conformation.[92, 147] Significant A-loop motion and backbone flipping of the DFG motif are necessary to transition from the AF2-predicted DFG-in state to holo-like states, for type II ligands targeting the classical DFG-out state. As a result, AF2 structures of Abl1/DDR1 kinases exhibit superior performance when docking with type I inhibitors (achieving a minimum ligand RMSD of 2.14 Å) compared to docking with type II inhibitors using Glide XP (Fig. 5.2 C&D). Even with Glide's Induced-fit Docking (Fig. E.5) and the highly side-chain clashes forgiving docking method DiffDock (Fig. E.7), AF2 structures struggle to dock with these metastable state-targeting ligands (type II inhibitors), with ligand RMSDs above 8 Å across all docking poses from all three docking methods.

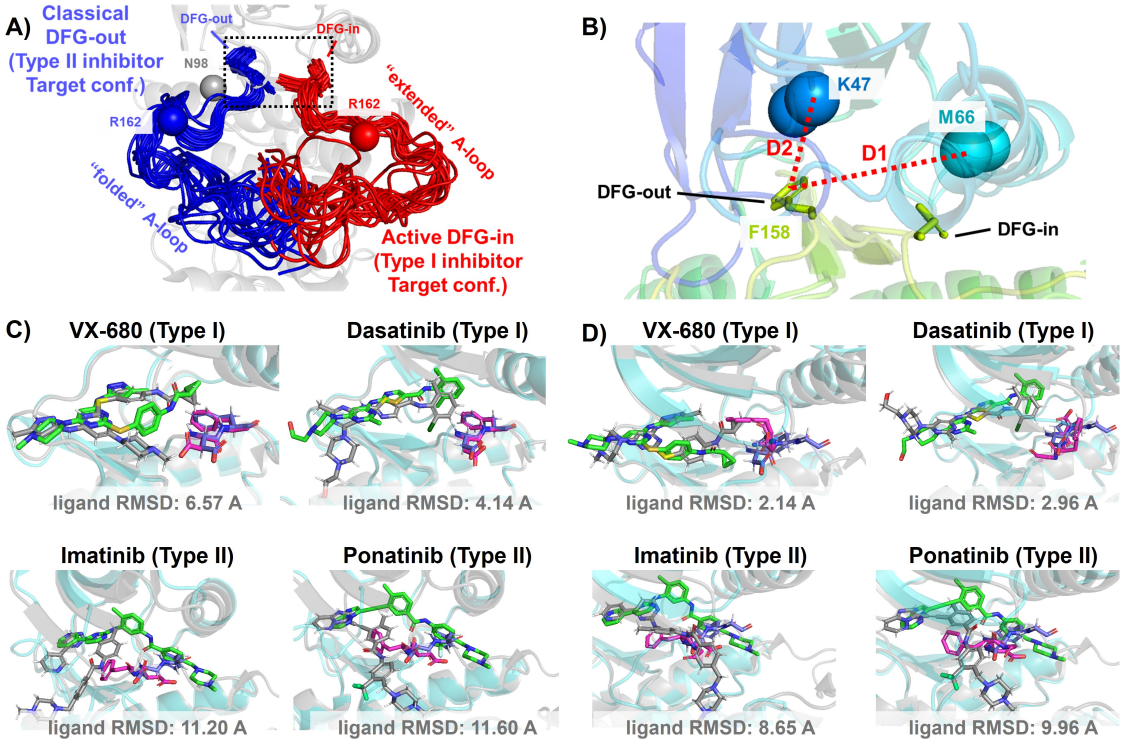


Figure 5.2: A)The NMR (Nuclear Magnetic Resonance) structures of Abl1 kinase overlay, comparing the activation loop (A-loop) in the active DFG-in state (red, PDB: 6XR6) with the classical DFG-out state (blue, PDB: 6XRG). Type I inhibitors target the active DFG-in state, where the DFG motif adopting the DFG-in configuration and the A-loop adopting the "extended" configuration, while type II inhibitors target the classical DFG-out state, where the DFG motif adopting the DFG-out configuration and the A-loop is "folded". The distance between CB atoms of residue N98 (grey bead) and residue R162 (red/blue bead) in Abl1 kinase serves as an order parameter here to illustrate the location of the A-loop. The dashed black block emphasizes the different configurations of the DFG motif in this two states. B) The Dunbrack definition for DFG motif configuration is employed here. The Dunbrack space is delineated by two order parameters: D1=dist(F158-CZ, M66-CA), D2=dist(F158-CZ, K47-CA). C) or D) The docking poses with the lowest ligand RMSD for 4 known kinase inhibitors targeting the Abl1 or DDR1 kinase AF2 structure, generated by Glide XP. Co-crystallized structures are shown as light-cyan cartoons (proteins), green sticks (ligand) and magenta sticks (DFG motif). Docking poses are shown as light-gray cartoons (proteins), gray sticks (ligand) and blue sticks (DFG motif). Comparing with type I inhibitors, AF2 structures of protein kinases fail to dock with type II inhibitors.

5.1.2 Holo-like metastable structures may be present among decoys generated from rMSA AF2

AF2-based methods can achieve structural diversity by introducing dropouts in MSA inputs stochastically (rMSA AF2) or in a clustering manner (AF2-cluster). Additionally, models modified from the AF2 framework, such as the flow-match generative model AlphaFlow[148], have been developed to explore the diversity of conformational space. Similar protocols to rMSA AF2 have demonstrated the potential to address the induced-fit effect by generating diverse structures at the binding pocket, ranging from closed apo pockets to opened holo-like pockets.[149] Other investigations have also indicated that larger backbone motions, such as DFG-motif backbone flipping and A-loop movement in at least some protein kinases, can be captured by the rMSA AF2 ensemble, although the distributions of conformations deviate significantly from the correct Boltzmann distributions.[114, 150]

In this section, I utilized rMSA AF2 to generate 1280 diverse structures for Abl1 kinase or DDR1 kinase: 640 for MSAs of depth 16 and 32. See Supplementary Material for information regarding Src kinase. Following a filtering step based on RMSD from corresponding AF2 structures, 1198 and 1147 structures remain for Abl1 and DDR1, respectively (Fig. E.3). As shown in Fig. 5.3 A&B, for Abl1 kinase, only 4 out of 1198 structures have a folded A-loop, when using a distance cutoff of 15 Å between CB atoms of N98 and R162 in Abl1. However, for DDR1 kinase, 124 out of 1147 rMSA AF2 structures exhibit a folded A-loop, when using a salt bridge distance cutoff of 10 Å between the aligned residue pairs in DDR1 (E110 and R191). I then clustered the A-loop folded structures in the Dunbrack space. For Abl1, only two clusters were identified: one adopts the DFG-in state, and the other adopts the DFG-inter state, referring to an intermediate

conformation during the backbone flipping from DFG-in to DFG-out, according to the Dunbrack definition. For DDR1, structures were divided into 5 clusters, with one cluster of size 15 being the closest to the classical DFG-out state.

Subsequently type II inhibitors, ponatinib and imatinib, were docked to the most "classical DFG-out"-like cluster from rMSA AF2 ensembles (indicated as red circles in Fig. 5.3 A&B), using IFD. Despite the A-loop being folded, the DFG-motif in structures from the Abl1 cluster is not strictly DFG-out. As expected, type II ligands fail to dock with those structures, with all IFD poses exhibiting ligand RMSDs above 9 Å (Fig. E.5 A, Fig. 5.5 B). While AF2-based methods can indeed generate decoy structures that deviate from the native state, it remains uncertain whether these decoys correspond to metastable basins. Additionally, it's unclear whether these decoys includes structures that can represent the specific metastable states required for the intended types of drug design.

In contrast to Abl1, rMSA AF2 ensemble for DDR1 contains holo-like structure for type II inhibitors. One structure (which I label the "holo-model", shown as a red circle filled with green in Fig. 5.3 B), out of the 15 in the DDR1 classical DFG-out cluster docked ponatinib with a remarkably low RMSD of 0.89 Å using IFD (Fig. 5.3 C). However, IFD poses from all the other structures exhibit large ligand RMSDs above 6 Å (Fig. E.5 B). Hence, an enrichment process to select holo-like structures from decoys becomes essential to ensure a practical number of pocket structures for ensemble docking and virtual screening.

The "holo-model" structure was also docked with another type II ligand, imatinib. In this case, the steric clashes are more significant, rendering the refinement for the induced-fit effect more challenging compared to the ponatinib scenario. I thus introduced an extra trimming step (see Methods section for details) for the DFG-Phe residue which exhibit significant clashes with

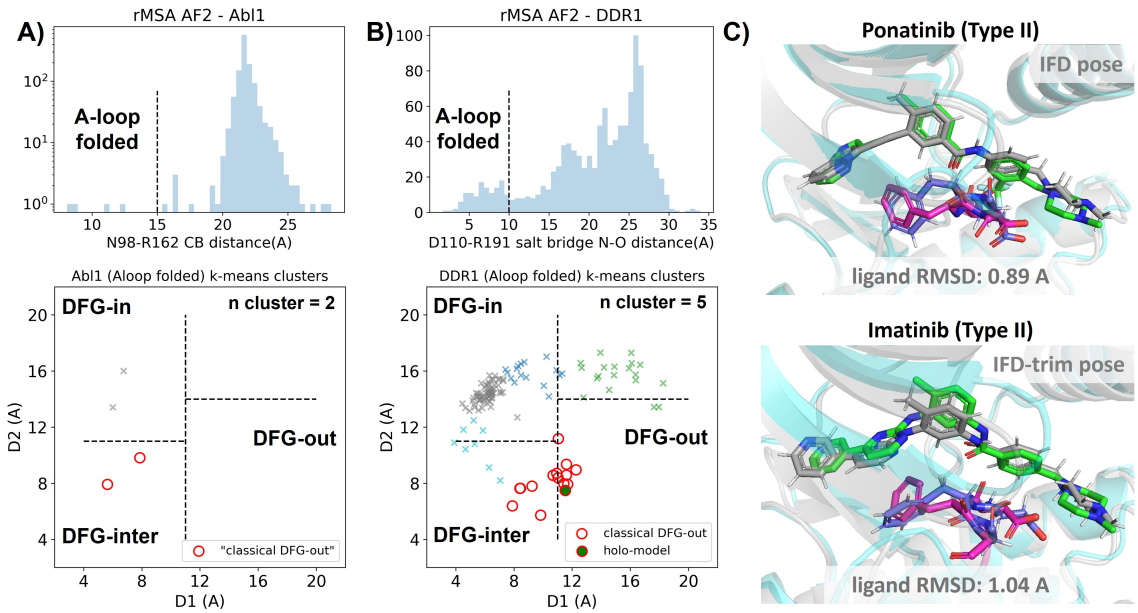


Figure 5.3: A) Upper panel: distribution of A-loop location for the reduced MSA AF2 structures of Abl1 kinase. 4 out of 1198 structures are A-loop folded. Lower panel: k-means clustering of 4 A-loop folded Abl1 structures in Dunbrack space, with the number of clusters (n cluster) set to 2. The structures in the cluster closest to the classical DFG-out state (red circles) fail to dock with type II inhibitors using induced-fit docking (IFD). B) Upper panel: distribution of A-loop location for the reduced MSA AF2 structures of DDR1 kinase. 124 out of 1147 structures are A-loop folded. Lower panel: k-means clustering of 124 A-loop folded DDR1 structures in Dunbrack space, with n cluster set to 5. Among the 15 structures in the cluster closest to the classical DFG-out state (red circles), one structure ("holo-model", highlighted by a red circle filled with green) demonstrates successful docking with type II inhibitors, showcasing a ligand RMSD < 2 Å, utilizing IFD or an extended-sampling version of IFD, IFD-trim. C) The docking poses with the lowest ligand RMSD for 2 type II kinase inhibitors targeting the DDR1 kinase "holo-model" structure, generated by IFD or IFD-trim. The color code is the same as Fig. 5.2.

the holo-imatinib in the "holo-model" structure (Fig. E.6 B). The IFD-trim protocol successfully achieves a minimum ligand RMSD of 1.04 Å for docking poses of the "holo-model" structure with imatinib (Fig. 5.3 C, Fig. E.6 C&D).

5.1.3 AF2-RAVE on DDR1 enriches holo-like classical DFG-out structures in rMSA AF2 decoys

To assess the utility of physics-based methods in selecting holo-like structures from AF2-generated ensembles, I utilized a physics-based protocol, AF2-RAVE[94] to explore the energy landscape of DDR1 kinase. I employed the identical set of collective variables (CVs) as in my previous study[114] to perform regular space clustering for the DDR1 rMSA AF2 ensemble. Two structures were selected from the clustering centers for each combination of DFG type (in, inter, or out) and A-loop position (folded or extended), resulting in a total of 12 initial structures (Fig. E.8). Subsequently, 50 ns unbiased MD simulations were conducted starting from each initial structure. The CVs extracted from all MD trajectories were input into the SPIB model with a linear encoder, to learn the reaction coordinates indicative of the slow motions linked to transitions between metastable states. The learnt SPIB reaction coordinates possess clear physical interpretations, with the first coordinate signifying the A-loop position and the second indicating the DFG-type (Fig. 5.4 A&B). This alignment with physical features is expected as I deliberately selected diverse and representative initial structures for unbiased MD simulations. Additionally, it's noteworthy that these reaction coordinates are transferable across various kinases and can effectively discern different metastable states.

Interestingly, the 15 classical DFG-out structures within the DDR1 rMSA AF2 ensemble are situated in a region of the latent space that were not thoroughly explored by the 12 unbiased MD trajectories. To address this gap, I employed enhanced sampling to sample along the SPIB-approximated reaction coordinates and compute the free energy profile inside the classical DFG-out basin. Considering both the flipping of the DFG motif and the overall motion of the

large flexible A-loop, it might be impractical to sample direct back-and-forth transitions between various states using metadynamics. Therefore, I opted for umbrella sampling for its simplicity. The reliability of umbrella sampling hinges on two issues, first whether the latent space adequately represents the conformational space and second, whether there is sufficient overlap between different windows for efficient reweighting. Addressing the first challenge remains an ongoing endeavor in the dimensionality reduction research field, and I anticipate that my SPIB latent space is sufficient enough for my current purpose. The second challenge can be managed through careful setup of umbrella sampling windows and bias strength. Given my current setup, sampling the extensive motion of A-loop relocation remains challenging, showing insufficient overlap between the A-loop folded and extended regions (Fig. E.9). Consequently, the quantitative reliability of the absolute ΔG values between different states is limited, allowing me only to qualitatively assess the relative thermodynamic stability of the DFG-in versus the DFG-out basins. Nevertheless, the qualitative relative stability observed from umbrella sampling aligns with previous studies[4, 114] (Fig. E.10). Furthermore, the local potential of mean force (PMF) surrounding each basin should provide quantitatively reliable insights, given their thorough sampling through umbrella sampling. Reassuringly, when I ranked the classical DFG-out structures within the DDR1 rMSA AF2 ensemble, using the local PMF values in the latent space, the "holo-model" structure emerged among the top 2 structures with free energy relative to the minimum smaller than 1 kJ/mol, as illustrated in Fig. 5.4 D.

Additionally, I used DiffDock to conduct docking experiments with type II ligands on the 15 classical DFG-out structures within the DDR1 rMSA AF2 ensemble. Despite poses from most structures surprisingly demonstrating very low ligand RMSD, typically below 2 Å (Fig. E.7), they exhibited significant steric clashes and low DiffDock confidence score due to Diffdock's

disregard for side-chain configurations and clashes. Notably, upon ranking structures based on their AF2-RAVE PMF and examining the corresponding poses with the lowest ligand RMSD, I observed that only the poses derived from the top 2 structures selected by AF2-RAVE had Diffdock confidence scores higher than -1.5, surpassing the default threshold of the DiffDock model (Supplementary File 1). This implies the general advantage of structures selected by physics-based methods across various docking methods.

In summary, Boltzmann ranking from AF2-RAVE effectively distinguishes holo-like structures from other decoys. Beginning with the DDR1 rMSA AF2 ensemble, AF2-RAVE substantially enhances the likelihood of identifying the holo-like structure from 1 out of 15 to at least 1 out of 2 when using a PMF cutoff of 1 kJ/mol. I noted that although the PMF values and Boltzmann ranks may fluctuate with the setup of umbrella sampling, the enrichment effect of holo-like structures remains (Fig. E.11). Besides, I applied an alternative protocol to compute the PMF profile within the DDR1 classical DFG-out basin, by running unbiased MD simulations starting from the 15 classical DFG-out decoys in the DDR1 rMSA AF2 ensemble. However, this unbiased protocol poses a risk of failing to sample rare events if barriers exit within the region of interest. In this case, the mini-barriers inside the classical DFG-out basin are low enough to enable efficient sampling using 50 ns unbiased MD simulations. Consequently, the Boltzmann ranks derived from unbiased MD simulations also successfully enriched the single holo-like structure in the DDR1 rMSA AF2 ensemble to the top 5 of the structure set (Fig. E.12).

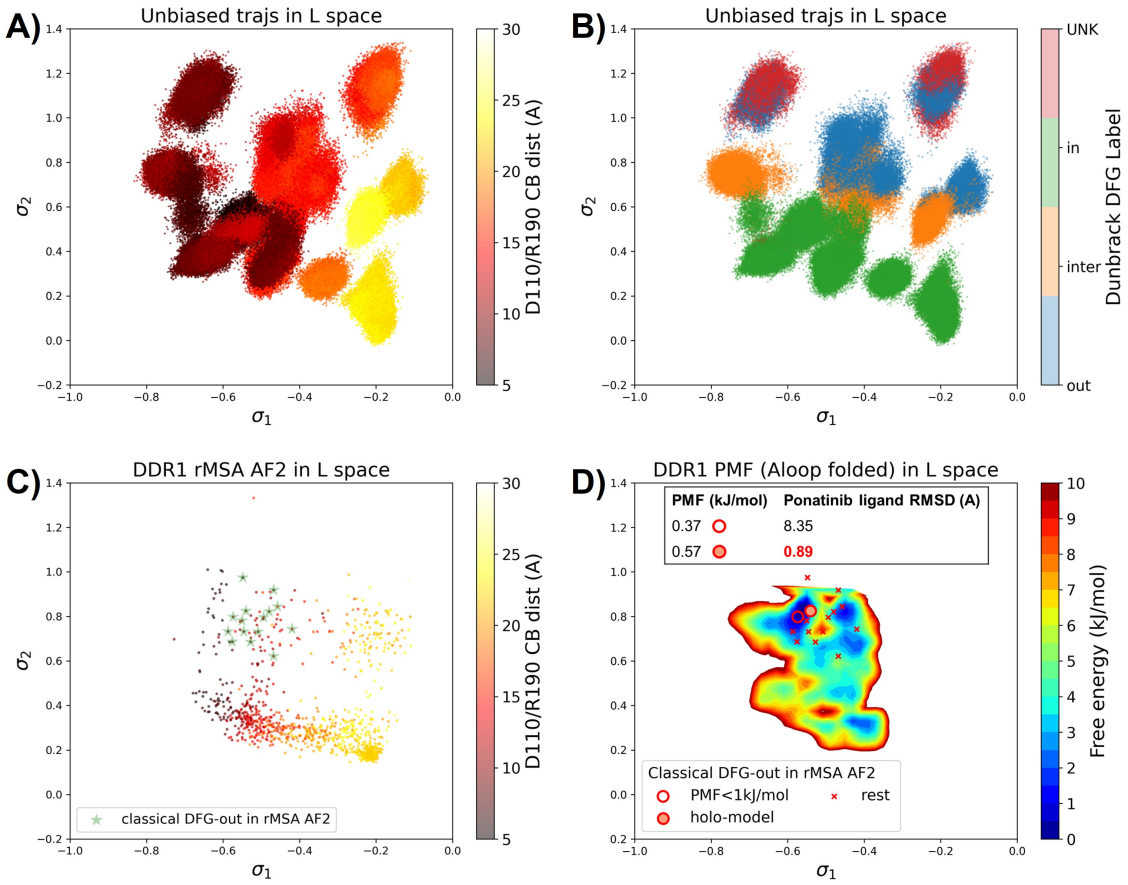


Figure 5.4: A or B) The unbiased MD trajectories of DDR1 are projected onto the learnt SPIB latent space. In plot A), the colors of sample points represent the A-loop location, while in plot B), they depict the Dunbrack DFG state. The first SPIB coordinate, σ_1, correlates with the A-loop location, and the second SPIB coordinate, σ_2, correlates with configuration of the DFG motif. C) The reduced MSA AF2 structures of DDR1 are projected onto the latent space. Sample points are color-coded based on the A-loop location. Light green stars highlight the 15 classical DFG-out structures selected based on prior information in Fig. 5.3. D) Free energy profile in the A-loop folded region of the latent space, calculated from umbrella sampling simulations. The 15 classical DFG-out structures from reduced MSA AF2 are shown as red cross and circles (structures with free energy less than 1 kJ/mol). The "holo-model" structure is emphasized using a red circle filled with red. The embedding table shows the lowest ligand RMSD in IFD poses of the rMSA AF2 structure with ponatinib. The "holo-model" is among the 2 structures selected by AF2-RAVE (PMF $<$ 1 kJ/mol).

5.1.4 Transferable learning of holo-like structure for Abl1 and Src kinases with AF2-RAVE from DDR1 templates

As mentioned earlier, my current Abl1 rMSA AF2 ensemble lacks any decoy structure in the classical DFG-out state. This is further illustrated in Fig. 5.5 A, where the Abl1 rMSA AF2 ensemble is projected onto the same latent space learnt during the DDR1 AF2-RAVE protocol. Hence, it's necessary to prepare the Abl1 decoy set adopting the classical DFG-out state before I can rank them using Boltzmann weights derived from physics-based methods and conduct further docking for selected holo-like structures.

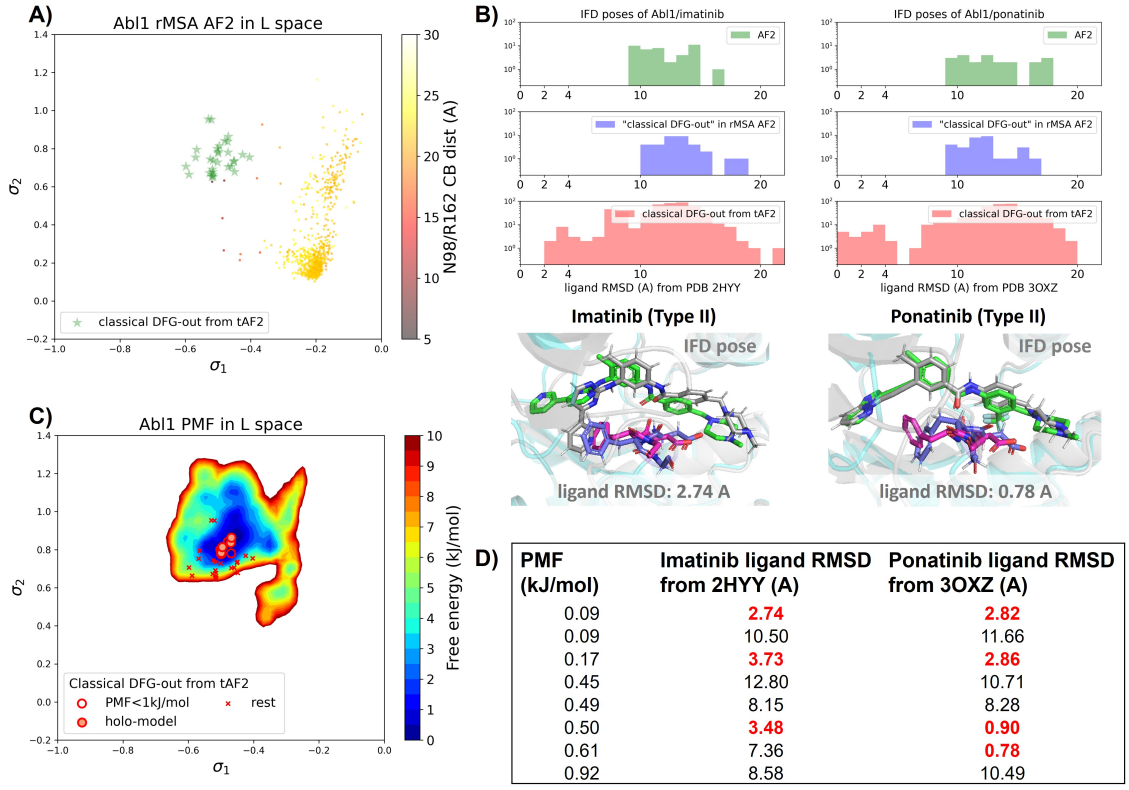
There are several approaches to generate the Abl1 decoy sets. One can conduct enhanced sampling on the latent space starting from Abl1 rMSA AF2 structures to reach the classical DFG-out basin. Subsequently, MD structures from this basin can serve as templates for asking AF2 to generate crystal-like structures in classical DFG-out state. However, for simplicity, I opted to use the 15 classical DFG-out structures from the DDR1 rMSA AF2 ensemble directly as templates and employed an AF2-based homology modeling protocol, referred to as AF2-template (tAF2 for short, detailed protocol can be found in the Methods section), to generate a decoy set comprising 30 Abl1 structures, as illustrated by the green stars in Fig. 5.5 A.

Compared to the AF2 and rMSA AF2 structures, the performance of IFD on type II inhibitors shows a significant improvement when using the 30 tAF2 decoy structures of Abl1. The lowest ligand RMSD achieved is 2.74 Å for imatinib and 0.78 Å for ponatinib (Fig. 5.5 B). However, only 4 structures (labeled as "holo-model" structures hereafter) out of the 30 decoys are capable of docking with type II inhibitors with ligand RMSD $<$ 3 Å, and all other structures produce IFD poses with ligand RMSD$>$ 6 Å (Fig. E.14).

I then investigated whether physics-based methods could enrich the 4 "holo-model" structures from the 30 tAF2 decoys to a practical number of holo-like candidates. To explore the local classical DFG-out basin, I conducted 50 ns unbiased MD simulations starting from each tAF2 decoy structure and combined the trajectories to obtain the final Boltzmann distribution. Remarkably, the PMF for the Abl1 classical DFG-out basin calculated from the unbiased protocol (Fig. 5.5 C) showed the enrichment of all 4 "holo-model" structures among the top 8 tAF2 structures with PMF value $<$ 1 kJ/mol. The PMF values of the top 8 tAF2 structures and the lowest ligand RMSD from the corresponding structures are presented in Fig. 5.5 D. I also employed umbrella sampling for the Abl1 kinase by transferring setups and initial structures using the AF2-template from DDR1 umbrella sampling. However, I noted a significant occurrence of αC-helix breakage in the Abl1 umbrella sampling trajectories compared to DDR1 (Fig. E.19). After excluding windows with broken αC-helix, the Abl1 PMF derived from umbrella sampling ultimately enriched all 4 "holo-model" structures to the top 6 among the 30 tAF2 structures (Fig. E.20). In Supplementary Information, I also provide results for Src kinase which are in the same quality as for Abl1 kinase.

5.1.5 Discussion

Through my retrospective analysis, I have thus demonstrated that the default AlphaFold2 models are ineffective for docking ligands targeting metastable protein kinase conformations. While AF2-based methods can be coaxed into generating diverse structures, they still struggle to produce reliable accuracy decoys for metastable conformations since the AF2 ensembles do not follow Boltzmann distribution. This failure is evident in the inability to generate an Abl1 AF2-



D)

PMF (kJ/mol)	Imatinib ligand RMSD from 2HYY (A)	Ponatinib ligand RMSD from 3OXZ (A)
0.09	**2.74**	**2.82**
0.09	10.50	11.66
0.17	**3.73**	**2.86**
0.45	12.80	10.71
0.49	8.15	8.28
0.50	**3.48**	**0.90**
0.61	7.36	**0.78**
0.92	8.58	10.49

Figure 5.5: A) The reduced MSA AF2 structures of Abl1 are projected onto the latent space. Sample points are color-coded based on the A-loop location. Light green stars highlight the 30 AF2-template Abl1 structures modelled from the 15 DDR1 classical DFG-out structures. B) Upper panel: the distribution of ligand RMSD for the IFD poses of Abl1 structures and two type II ligands. Lower panel: IFD poses with the lowest ligand RMSD for Abl1 AF2-template structures and two type II ligands. The color code is the same as Fig. 5.2. C) Free energy profile in the latent space, calculated from unbiased MD simulations. The 30 Abl1 classical DFG-out structures from AF2-template are shown as red cross and circles (structures with free energy less than 1 kJ/mol). The "holo-model" structures are emphasized using red circles filled with red. D) The table shows the lowest ligand RMSD in IFD poses of the AF2-template structures with two type II inhibitors. All the four "holo-models" are among the 8 structures selected by AF2-RAVE (PMF $<$ 1 kJ/mol).

ensemble containing holo-like structures for type II inhibitors. To further investigate whether this limitation is common among AF2-based methods, including the rMSA AF2 method I employed earlier, I tested another AF2-based approach, AF2-cluster.[95] The AF2-cluster ensemble of Abl1 comprises more A-loop folded structures (21 out of 197) compared to my Abl1 rMSA AF2 ensemble (4 out of 1198). However, similar to rMSA AF2, there are still no decoys in the classical DFG-out state, and all A-loop folded structures are located far from the PMF basin of the classical DFG-out state in the latent space (Fig. E.15 C). This indicates that AF2-cluster, like rMSA AF2, also fails to generate Abl1 metastable states effectively. Interestingly, I observed comparable structural diversity in sampling the A-loop folded configurations within the AF2-cluster ensembles for DDR1 and Abl1 (Fig. E.15 C&D). While for the rMSA AF2 method, the promiscuous kinase DDR1 ensemble exhibit superior structural diversity compared to the Abl1 kinase. This enhanced diversity leads to the identification of one dockable structure for type II inhibitors among decoys in DDR1 rMSA AF2 ensemble. Through the application of a homology modeling method, AF2-template, I demonstrated that the classical DFG-out decoys in the DDR1 rMSA AF2 ensemble can be transferred to Abl1 kinase. Furthermore, I tested an additional kinase, Src, for which non-native state decoys are reported to be even more challenging to produce using AF2 subsampling methods than Abl1 kinase[150]. I have also verified that rMSA AF2 and AF2-cluster struggle in producing distinct decoys from native structure of Src kinase (Fig. E.3 & Fig. E.4). Besides, neither the rMSA AF2 nor the AF2-cluster ensembles for Src kinase adequately sample the classical DFG-out basin in the latent space (Fig. E.17). Remarkably, AF2-template, as a homolgy modelling method, can easily produce the classical DFG-out structure of Src kinase, using the top 2 AF2-RAVE-picked DDR1 classical DFG-out structures as templates (Fig. E.16). The IFD poses from the tAF2 structure of Src kinase shown a minimal ligand RMSD

of 2.82 Å with imatinib (Fig. E.18).

Table 5.1: Comparing the IFD performance of various structure generation methods for docking type II kinase inhibitors

Source-protein (# of structures)	Lowest imatinib ligand RMSD (Å)	Lowest ponatinib ligand RMSD (Å)	Ratio of structs. w. ligand RMSD < 3 Å
AF2-Abl1 (1)	9.22	9.4	0/1
AF2-DDR1 (1)	9.24	9.33	0/1
rMSA AF2-Abl1 (2)	10.14	9.11	0/2
rMSA AF2-DDR1 (15)	**1.04***	**0.89**	1/15
tAF2-Abl1 (30)	**2.74**	**0.78**	4/30
AF2-RAVE-DDR1 (2)	**1.04***	**0.89**	**1/2**
AF2-RAVE-Abl1 (8)	**2.74**	**0.78**	**4/8**

* result from IFD-trim

With the rapid expansion of chemical space, virtual screening on libraries containing billions of diverse molecules becomes enticing for novel drug discovery [151]. Therefore, the enrichment of candidate holo-like structures emerges as a necessary step, offering significant benefits in terms of computational efficiency and feasibility. As summarized in Table 5.1, unlike the non-dockable AF2 structures for type II inhibitors, the diverse rMSA AF2 ensemble (referred to as rAF2 in Table 5.1 for brevity) shows potential in generating holo-like structures within a large set of decoys. However, it's only upon AF2-RAVE ranking and selection that the ratio of holo-like structures in selected structure set increases to a plausible value of 50%, facilitating further virtual screening on computational models of protein pockets.

5.2 Conclusion

AlphaFold2 has arguably revolutionized protein structure prediction, but its direct use in drug discovery remains challenging, especially for ligands targeting non-native or metastable protein conformations. In this chapter, I showed through retrospective kinase inhibitor studies that

combining AlphaFold2, statistical-mechanics-based enhanced sampling, and induced-fit docking can address this limitation. Specifically, I used AF2-RAVE to rank metastable kinase conformations and Glide Induced-Fit Docking (IFD) to evaluate type II inhibitor binding pockets. This AF2-RAVE Glide workflow yielded holo-like structure candidates with a 50% successful docking rate for known type II inhibitors.

A key result is that AF2 can be further empowered through a templating procedure. Holo-like DFG-out structures and the DDR1 latent space obtained from AF2-RAVE were transferable to homologous kinases, including the more challenging Abl1 and Src systems [150]. Using DDR1-derived metastable structures as templates allowed AF2 to generate classical DFG-out conformations for these kinases, reducing the need to relearn the latent space and sampling protocol from scratch. This transferability suggests that AF2-RAVE can be used not only to rank conformations within one protein, but also to propagate useful conformational information across related protein families.

Overall, these retrospective docking results support the practical use of the conformational selection hypothesis for structure-based drug discovery. AF2-based methods can generate diverse structural hypotheses, AF2-RAVE can assign thermodynamic weights to these structures, and docking can test whether the selected metastable states provide ligand-compatible pockets. The next step is to move beyond retrospective validation and perform prospective docking studies, where AF2-RAVE-selected conformations are used to screen compound libraries and experimentally discover new binders for a target protein.

Chapter 6: Hierarchical AF2-RAVE for Timescale-Resolved Holo-like Structure Generation

[This chapter is based on: Xinyu Gu, Venkata S. S. Adury*, **Akashnathan Aranganathan**, Xinhao Zhuang, Kristen M. Varney, David J. Weber & Pratyush Tiwary. Hierarchical AF2RAVE for Multiconformation Virtual Screening Targeting S100 Ca^{2+}-Binding Proteins. (2025) Journal of Chemical Theory and Computation.(*- equal contribution)[152]. I contributed to the methodology development, which branched from the previous chapter and was designed to avoid the computational cost of induced-fit docking. I also contributed to the initial ensemble generation for a homo-dimer system and provided the code for AF2-RAVE.]*

In the previous chapters, I showed that combining AI-based structure prediction with physics-based sampling improves the identification and Boltzmann ranking of metastable protein conformations. AF2-derived structural hypotheses can serve as reliable starting points for molecular dynamics, while AF2-RAVE refines these hypotheses and assigns thermodynamic weights to metastable states. I also showed that this ensemble-based perspective is useful for structure-based drug design because it accounts for protein flexibility beyond a single dominant static structure.

However, the previous chapters also revealed two important points. First, compared with MD-generated structures, AF2-generated structures are often more crystal-like and can perform better in static-structure docking protocols. Second, classification of metastable states at longer

timescales is useful for refining the structural search because it primarily captures large-scale conformational changes. Local rearrangements, especially side-chain rotamer changes and pocket-level packing, remain important for ligand docking. In the previous chapter, this local flexibility was addressed using induced-fit docking, but such approaches are computationally expensive and difficult to scale to large virtual screening campaigns [129, 130, 151, 153].

This motivates a hierarchical extension of AF2-RAVE. As discussed in Chapter 1, conformational selection provides a useful framework for treating ligand binding as the stabilization of pre-existing metastable states. For practical docking, however, it is not sufficient to sample only slow backbone-level conformational changes; one must also capture local side-chain and pocket rearrangements that affect ligand compatibility. In this chapter, I use a hierarchical AF2-RAVE strategy in which metastable states are classified at multiple timescales using different input features and time lags. The goal is to generate a diverse but tractable set of holo-like structures that capture both global conformational changes and local pocket-level rearrangements, making them suitable for rigid docking and high-throughput virtual screening.

I apply this hierarchical AF2-RAVE pipeline to the calcium-binding S100 protein family [154, 155]. S100 proteins typically function as homodimers and expose a hydrophobic pocket upon calcium binding, enabling interactions with diverse protein and nucleic-acid targets. These interactions regulate cellular processes such as proliferation, differentiation, apoptosis, calcium homeostasis, inflammation, and cell migration [156]. Several S100 isoforms are disease-associated biomarkers; for example, S100B is associated with malignant melanoma and other cancers, while S100A1 has been linked to Alzheimer's disease [157, 158]. Despite their high sequence and structural similarity, S100B and S100A1 mediate distinct biological interactions, making them useful systems for testing whether ensemble-based structure generation can

aid selective inhibitor design.

In this chapter, I use hierarchical AF2-RAVE to explore the free-energy landscapes and metastability of Ca^{2+}-S100B and Ca^{2+}-S100A1, starting from sequence information. I then evaluate the resulting holo-like structures using retrospective docking with Glide XP [126, 127, 128] and enrichment testing with UCSF DOCK 6.11, referred to here as DOCK [159, 160]. These tests use a previously unreported Ca^{2+}-S100B inhibitor dataset [161] and compare AF2-RAVE-generated structures with standard AF2 predictions and X-ray crystallographic Ca^{2+}-S100B/inhibitor complexes. Retrospective docking evaluates whether Glide XP can reproduce known binding poses, while enrichment tests whether DOCK can distinguish true inhibitors from property-matched decoys.

The key result is that the AF2-RAVE structure set outperforms standard AF2 structures in both retrospective docking and enrichment. Notably, AF2-RAVE also outperforms X-ray structures in enrichment, possibly because crystallization artifacts or assignment ambiguities in some X-ray structures can allow decoys to artificially improve docking scores [162]. Overall, the hierarchical AF2-RAVE structures show improved diversity and holo-like character, making them strong candidates for rigid docking and high-throughput virtual screening.

6.1 Methods

6.1.1 Hierarchical AF2-RAVE funnel for multiconfiguration holo-like receptor generation

In high-throughput virtual screening, protein receptor structures are usually kept rigid for computational efficiency. [163] This requires that both the global backbone and the pocket side-

chain packing of the protein receptor structures are already in holo-like states, ready for ligand docking. However, experimentally determined holo structures are not always available, and apo-like structures are often the only starting point for computational predictions[153]. Modeling large-scale backbone transitions from apo to holo can be challenging and cannot be addressed even by flexible docking [145]. Numerous computational simulation and enhanced sampling techniques have been developed to identify cryptic pockets for drug discovery.[149, 164, 165, 166] My previous work demonstrated that the standard AF2-RAVE protocol can sample and enrich holo-like structures with significant backbone rearrangements from the apo state, effectively addressing the conformational selection effect in protein kinases and their type II inhibitors[125].

When a ligand selectively stabilizes a particular backbone state, it shifts the local environment and alters the population distribution of metastable pocket states. Here, I simplify the induced-fit effect by assuming that while configurational populations change after ligand binding, the possible side-chain configurations of metastable pocket states remain largely unchanged. Thus, I assume that the final holo structure still resides within a metastable basin of the apo case. Under these assumptions, holo-like models that incorporate both conformational selection and induced-fit effects can be generated by sampling metastable states in the absence of ligands.

However, large-scale backbone changes and pocket side-chain rotamer changes are typically linked to metastability through different descriptors and across distinct timescales[17]. To decouple metastability at different levels and explore metastable states systematically, I used a hierarchical approach based on the standard AF2-RAVE protocol[94, 114, 167], to conduct dimensionality reduction and structure generation as shown in Fig. 6.1. Specifically, I first identified backbone motions on a slower timescale, followed by an analysis of pocket side-chain rotamer dynamics within each respective backbone mode, on a faster timescale.

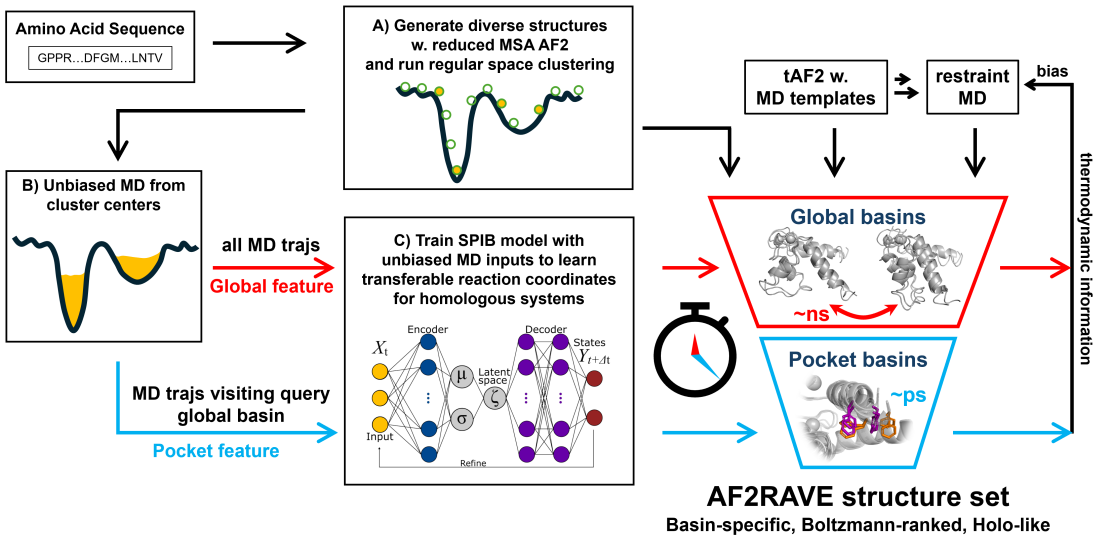


Figure 6.1: The hierarchical AF2-RAVE funnel for generating holo-like structures in metastable states. A) A diverse protein structural ensemble is generated from the sequence using reduced MSA AlphaFold2 (rAF2). B) Short unbiased MD simulations are performed, initiated from cluster centers identified in the rAF2 ensemble. C) Dimensionality reduction is applied to the MD trajectories (denoted trajs above) using the SPIB model[5] with different descriptors and timescales to decouple backbone motions from pocket side-chain dynamics. Guided by free-energy surfaces in both the global SPIB latent space and the pocket SPIB latent space, structures from AF2-based methods are further generated, refined, and selected. The final AF2-RAVE structure set consists of diverse, metastable, holo-like conformations.

6.1.1.1 Generating diverse structural ensembles from AF2-based methods

Vanilla AlphaFold2 (AF2) predictions, using the full multiple sequence alignment (MSA) of available evolutionary information, typically result in little structural diversity across multiple runs. To explore a broader range of conformations, several AF2-based pipelines have introduced stochasticity. This is achieved by substituting residues with alanines along randomly selected columns in MSA, as in SPEACH_AF [98], or by reducing MSA depth through random subsampling (referred to as reduced MSA AF2 or "rAF2") [91, 93, 96, 150] or clustering-based approaches [95, 99] and several others[168]. These methods effectively mask or reduce certain residue-residue contact signals from the full MSA, enabling predictions to potentially sample alternative structural states, but with no thermodynamic ranking of their propensities.

In this study, I use rAF2 to generate diverse structural ensembles, producing 640 structures per MSA depth. While this approach can produce highly varied structures including complete hallucinations[91, 169], I use the AF2-RAVE protocol to harness this diversity. By integrating dynamic information, AF2-RAVE identifies the truly metastable regions of the protein conformational landscape.

Unfortunately for both S100 dimer systems (S100A1 dimer or S100B dimer), the rAF2 structural ensembles generated with a MSA depth of 64 do not exhibit conformational diversity, as shown in Fig. F.2A. I thus used the structure with the highest pLDDT (predicted Local Distance Difference Test[170]) score at an MSA depth of 64 as my reference AF2 structure for each system. At the same time, ensembles with MSA depths of 8, 16, and 32 often produce hallucinated orientations and interfaces between the two chains in the dimer structure, leading to backbone C_α RMSD values of approximately 20 Å when compared to the AF2 reference structure, as shown in Fig. F.2A. To address this, the AF2 reference structure was used as a template for generating new template rMSA AF2 ensembles with MSA depths of 8, 16, and 32. This approach resulted in diverse structural ensembles with typical dimer orientations, as shown in Fig. F.2B. To eliminate highly unphysical structures, an RMSD filter was applied to exclude all structures with RMSD values greater than 7 Å, when compared to the AF2 reference structure. Structures containing steric clashes were removed if any non-bonded heavy atom pairs had a distance of less than 0.7 Å, as shown in Fig. F.2C. Additionally, the mean value and standard deviation of peptide bond lengths (reference mean value: 1.32 Å) were calculated for each structure. An extra filter was applied to remove structures with mean peptide bond lengths exceeding 1.4 Å and standard deviations greater than 0.2 Å. Following these filtering steps, 258 and 168 structures remained in the final template rMSA AF2 ensembles for S100A1 and S100B dimers, respectively. For simplicity,

I refer to these refined structural ensembles as the rAF2 ensembles of S100A1 and S100B in the following sections.

6.1.1.2 Unbiased MD simulations

Regular-space clustering was performed for the rAF2 ensemble of each S100 system (S100A1 or S100B), using 28 collective variables (CVs) representing the global/backbone conformation of monomers. These CVs consist of pairwise distances between C_α atoms of 8 anchor residues, which are the two terminal residues of the four rigid alpha helices in one monomer, as shown in Fig. F.4A. The anchor residues were identified based on fluctuation analysis from 1 μs-long MD simulations of six Ca^{2+}-S100 systems, as shown in Fig. F.3 (see SI for setup of this MD). The regular-space clustering process produced 10 cluster centers for S100A1 and 15 cluster centers for S100B, which were treated as initial protein structures for subsequent MD simulations. Calcium ions were added to the center of mass of all the oxygen atoms from Ca^{2+}-binding residues in each EF-hand[171].

For each initial structure, a 100 ns production run of unbiased MD simulation was performed, following the setup described in the SI. To prevent Ca^{2+} ions from dissociating from EF-hands while preserving their native coordination mode[172], a truncated harmonic constraint was applied during equilibration runs and production runs, to pull calcium ions back if their distance from coordinating oxygen atoms exceeded 5 Å, as shown in Fig. F.5C and further detailed in SI.

6.1.1.3 Global SPIB and local/pocket SPIB

I applied hierarchical dimensionality reduction using the SPIB model to molecular dynamics (MD) trajectory data. All SPIB models learned two-dimensional latent spaces and employed linear encoders for better interpretability of the latent variables. To distinguish metastable states across different timescales, I trained two layers of SPIB models with distinct time lags and two sets of input features: global features and local/pocket features, corresponding to the global and local SPIB models respectively. Free-energy surfaces projected onto the 2D SPIB latent space were computed via WHAM[173] and used to identify metastable basins. Detailed training procedures are provided in the SI.

For the global SPIB model, the input CVs were derived from the 28 global features previously used in regular-space clustering. The global SPIB model trained with a time lag of 25 ns on the Ca^{2+}-S100B system (Fig. F.6) was used to define the global SPIB latent space throughout this study. Since the global features are transferable between Ca^{2+}-S100A1 and Ca^{2+}-S100B due to their high sequence similarity, the resulting global SPIB latent space is also applicable to Ca^{2+}-S100A1. When viewed along the global SPIB latent space, islets with free-energy values below 4 kJ/mol relative to the minimum were defined as different metastable basins. For each of the two deepest basins, I selected the three unbiased MD trajectories for which the mean IB values were closest to the basin minimum on the latent space. I then filtered out trajectories with high variances in IB values. As shown in Fig. F.7B, this ensured that the selected trajectories for each global basin shared the same backbone mode and could be used as input for the local SPIB model in subsequent analysis.

To derive local features for describing side-chain packing modes, I first applied the binding

site detection method Graph Attention Site Prediction (GrASP) model[124] to the S100B AF2 structure and identified pocket residues as those with heavy atoms having a GrASP score above 0.2. Additionally, three C-terminal residues, C85, F88, and F89, were included to account for critical C-terminal conformational changes relevant to Ca^{2+}-S100B inhibitor design, as reported in previous studies[174]. The illustration of pocket residues can be found in Fig. F.4B. I then derived 53 local features, which are pairwise distances between distal atoms of pocket residue pairs with a residue index difference less than 10. Assuming that the backbone mode remains the same in the training set of each local SPIB model, relative side-chain positions between residue pairs close in sequence should be sufficient for representing side-chain rotamer packing states.

Given trajectories within a certain backbone mode (here I focus on the native backbone basins of both Ca^{2+}-S100B and Ca^{2+}-S100A1), I trained the SPIB model based on the 53 local features. Unlike backbone motions, side-chain rotamers usually fluctuate on much faster timescales. A hyperparameter sweep showed that the latent space remained stable across time lags from 50 ps to 800 ps, therefore, I selected the 300 ps model as representative for the local SPIB analysis. The resulting latent variables and latent space are presented in Fig. F.8 and Fig. 6.3.

6.1.1.4 AF2-template and restrained MD for AF2-RAVE structure sets generation

In the local SPIB latent space, adjacent bins with free-energy values below 2 kJ/mol were grouped into different metastable basins. MD frames from training trajectories, output at 10 ps intervals, were projected onto the SPIB latent space. Frames within each local metastable basin

were concatenated in the order of simulation time, and 20 representative frames were uniformly selected at equal intervals, as shown in Fig. F.9A. For each local basin, these 20 frames served as template structures for AF2-template (tAF2) modeling. For a given query sequence, tAF2 structures are generated by ColabFold[11] by uploading desired template structures and using the "single-sequence" mode for the Evoformer module. For each local basin, 20 tAF2 models are generated, and the top two structures with the highest pLDDT (Fig. F.9B) were used as initial protein structures for subsequent restrained MD simulation. The restrained MD simulations are short, low temperature runs. The conformations of key residues with slow dynamics were biased toward free-energy basins identified in the local SPIB latent space, whereas rotamers with fast dynamics retained "crystal-like" conformations similar to those in the tAF2 models. Final frames from the restrained MD production runs were compiled into the AF2-RAVE structure set (Fig. F.9C). This structure set was later used for docking validation, as described in Appendix F.5. Detailed setup procedures are provided in the SI.

6.1.2 Docking validation as a downstream test

The main focus of this chapter is the hierarchical AF2-RAVE procedure used to generate timescale-resolved, holo-like receptor structures. I use retrospective docking and enrichment only as downstream validation to test whether these structures are useful for rigid virtual screening. The detailed ligand preparation, docking, enrichment, and BEDROC protocols are provided in Appendix F.5. In the main text, I retain only the results needed to show that the hierarchical AF2-RAVE structure set improves retrospective docking and enrichment relative to standard AF2 and X-ray-derived receptor choices.

6.2 Results

6.2.1 Global SPIB captured the backbone breathing of the hydrophobic pocket in Ca^{2+}-S100

S100 proteins typically form homodimers, with each monomer comprising an N-terminal pseudo EF-hand motif (residues 3–39) and a C-terminal canonical EF-hand motif (residues 51–89), connected by a hinge loop. Each EF-hand consists of two α-helices linked by a calcium-binding loop. Upon calcium ions binding, significant conformational changes occur at the C-terminal EF-hand, resulting in the exposure of a hydrophobic pocket formed by helix 3, helix 4, and the hinge region.[171] This pocket then recruits target proteins primarily through hydrophobic protein–protein interactions.

Since the input features used for global SPIB are transferable across the S100 protein family, the learned global SPIB coordinates (Fig. F.6) and corresponding latent space are also transferable. By projecting the free-energy surfaces for both Ca^{2+}-S100A1 and Ca^{2+}-S100B onto this shared latent space, I identified two distinct global basins that represent the backbone breathing of the pocket, as shown in Fig. F.1. While the basins occupy similar regions in the latent space for both isoforms, the precise positions of their minima differ. These differences can be utilized for isoform-specific selective drug design.

Projections of available holo structures (complexed with peptides or small-molecule inhibitors, PDB IDs provided in SI) onto the latent space show that most resolved complexes fall within the native global basin, with a widely open hydrophobic pocket, shown as stars in Fig. F.1. For prospective drug discovery, I anticipate that the alternate basin with a more closed pocket is

likely to be relevant. However, since all known inhibitors primarily target the native global basin, my subsequent pocket side-chain-level local SPIB analysis and docking validation in this study were focused within the native basin.

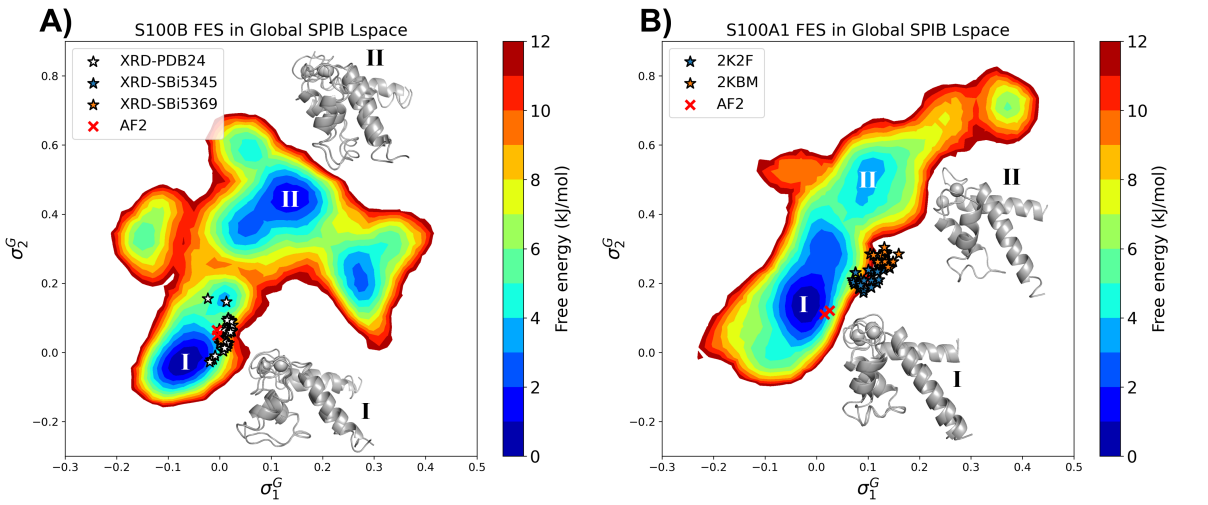


Figure 6.2: Free-energy surfaces for Ca^{2+}-S100B (A) and Ca^{2+}-S100A1 (B) in a transferable global SPIB latent space. Two major global basins are identified, reflecting the breathing of the hydrophobic pocket of calcium-bound S100 proteins. Representative structures from unbiased MD for both basins and both Ca^{2+}-S100 systems are depicted in gray, with only one monomer shown for clarity. The projections of AF2 structures (red crosses) and experimentally determined holo structures (stars) indicate that these structures all reside within basin I, the native global basin. For Ca^{2+}-S100B, experimentally determined holo structures includes 24 published XRD structures in PDB (denoted as PDB24, projected as white stars) and 2 unpublished XRD structures with inhibitor SBi5345 or SBi5369 (projected as blue and orange stars). For Ca^{2+}-S100A1, two peptide-bound holo NMR structural ensembles were projected as blue and orange stars.

6.2.2 Local SPIB captured metastable side-chain packing states, covering both the AF2-predicted apo and known holo conformations for the Ca^{2+}-S100 proteins

For the native backbone basins of both Ca^{2+}-S100A1 and Ca^{2+}-S100B, local (pocket-level) SPIB analysis was performed to explore all possible long-lived side-chain packing configurations. Interpretation of the learned SPIB reaction coordinates revealed the most relevant

features associated with the slow transitions between metastable side-chain states (Fig. 6.3 left panel). The dominant slow side-chain motions involve residues primarily in the C-terminal and hinge regions. Specifically, key residues identified by SPIB include D46, V47, Q48, A53, I57, E60, L61, L81, C85, F88, and F89 in Ca^{2+}-S100A1, while for Ca^{2+}-S100B they include F44, L45, V53, M80, I81, C85, F88, and F89. Notably among the S100 family, the C-terminal and hinge regions are the least conserved segments [175], suggesting that they play critical roles in interacting with isoform-specific binding partners [176]. Consistently, SPIB naturally highlights them as strategic target residues for selective drug design. In particular, for Ca^{2+}-S100B, the design of its covalent inhibitors (e.g., SC124, SC1475, SC1982, SBi4172, and SBi4434), which attach to Cys85, shows conserved interactions with residues F44, F88, and F89 [177]. Additionally, X-ray structures of the complexes between Ca^{2+}-S100B with pentamidine analogs have shown that the C-terminus adopts two distinct conformations, defined by the positions of F88 and F89, termed the "FF-Gate". Symmetric pentamidine analogs are associated with the closed FF-Gate conformation, while asymmetric analogs induce the open state[174]. In agreement with these experimental observations, the local SPIB coordinate, σ_1^L captured the FF-Gate conformational transitions, with the free-energy surface in the SPIB latent space revealing distinct basins for both open and closed FF-Gate states, as shown in Fig. 6.3B.

The projections of both the AF2-predicted apo and the known holo side-chain rotamer configurations fall within distinct metastable basins on the free-energy surface (FES) in the pocket SPIB latent space, as indicated by crosses and stars in Fig. 6.3. When mapping all reduced-MSA AF2 structures with native global backbones (Fig. F.7A), these structures exhibit greater side-chain packing diversity compared to vanilla AF2 predictions (Fig. 6.3, middle panel). However, they still fail to cover all potential metastable basins of side-chain configurations, most notably,

they miss those corresponding to experimentally resolved holo structures (Fig. 6.3, right panel).

To generate a Ca^{2+}-S100B structural set with coverage of all local SPIB basins, I performed AF2-template modeling using MD-derived structures from each basin as templates. This approach biases the conformations of critical pocket residues toward specific basins, while preserving crystal-like configurations in the remaining residues, given AlphaFold2's inherent preference for crystal-like structures. The retention of this crystal-like trait is expected to benefit future virtual screening efforts, as most docking algorithms are trained and optimized on experimentally resolved crystal structures.

However, since the template module of AlphaFold2 utilizes only the C_α and C_β atom positions from the template structure, I find that the AF2-template alone is insufficient to fully capture the detailed side-chain configurations in templates (Fig. F.9B). To address this, short, restrained MD simulations were performed to bias key side-chains toward the desired local SPIB basin. This process generated the final AF2-RAVE structure set within the native global basin of Ca^{2+}-S100B, which was then used for subsequent retrospective docking and enrichment validation (Fig. 6.4A, Fig. F.9C, Fig. F.12).

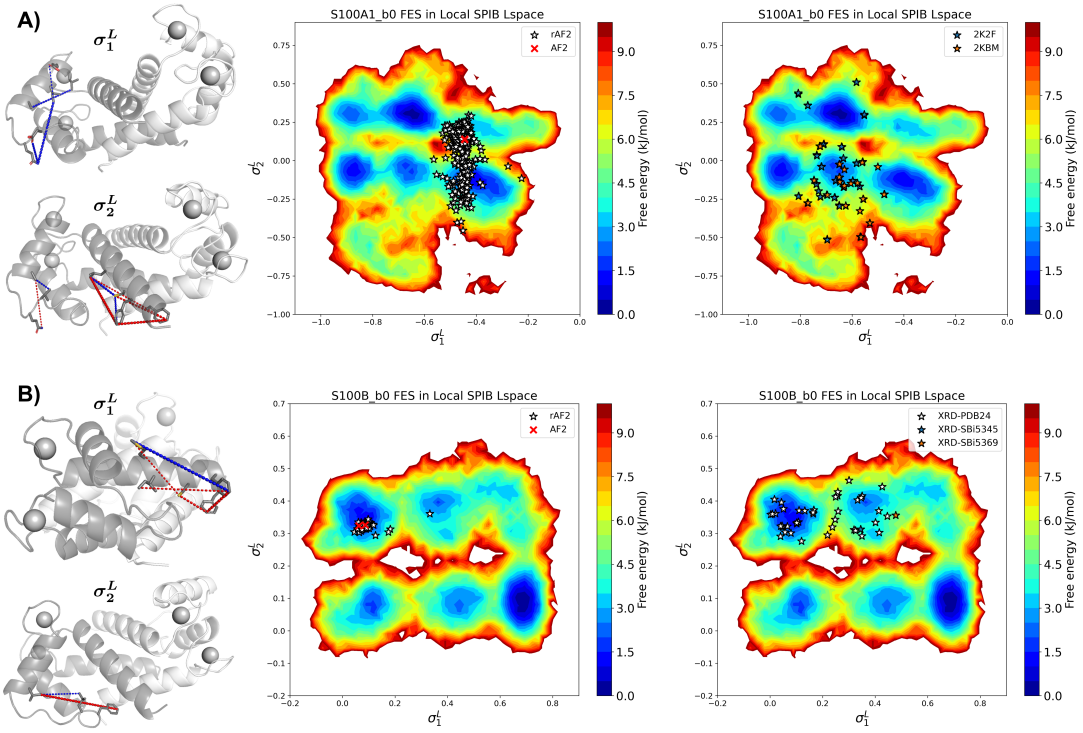


Figure 6.3: Free-energy surfaces for Ca^{2+}-S100B (A) and Ca^{2+}-S100A1 (B) within their native global basins, represented in the local/pocket SPIB latent spaces. The left panel illustrates the most dominant features (whose weights are larger than 0.5 maximum weights) contributing to the learned SPIB coordinates, σ_1^L,σ_2^L. The color and width of the dashed lines indicate the sign and absolute magnitude of the coefficients in the learned linear combination. In both systems, slow conformational changes are primarily associated with side-chain rotamer transitions around the hinge loop region and the C-terminal, especially the FF-gate region. The projection of the rAF2 ensemble (middle panel, white stars) and AF2 structures (middle panel, red crosses) onto the pocket latent space reveals poor coverage of possible metastable basins for side-chain configurations. Notably, the experimentally determined holo structures (right panel, all stars) fall into metastable basins that are not well sampled by AF2-generated models.

6.2.3 Retrospective docking validation on AF2-RAVE structure set of Ca^{2+}-S100B

To assess the suitability of the AF2-RAVE structure set for rigid docking, I performed retrospective docking using Glide XP on two newly designed Ca^{2+}-S100B inhibitors, SBi5345 and SBi5369, both with X-ray holo structures resolved. Using the X-ray holo structure as a reference, the lowest ligand RMSD docking poses from AF2-RAVE consistently outperformed

those from AF2 and were comparable to poses obtained using the holo XRD protein structures (Fig. 6.4B&C), in terms of ligand RMSD. To further evaluate the interacting surface in docking poses, I analyzed protein–ligand interaction fingerprints (PLIFs), which confirmed the superior performance of AF2-RAVE structures (Fig. F.10, Fig. F.11). Notably, both inhibitors bind to Ca^{2+}-S100B in the FF-gate closed conformation, (shown as red stars in Fig. 6.4A), and the lowest ligand RMSD poses from AF2-RAVE structure set were given by the structure with similar FF-gate configurations ($\sigma_1^L \approx 0.4$, as shown in Fig. 6.4C), closely matching the XRD reference.

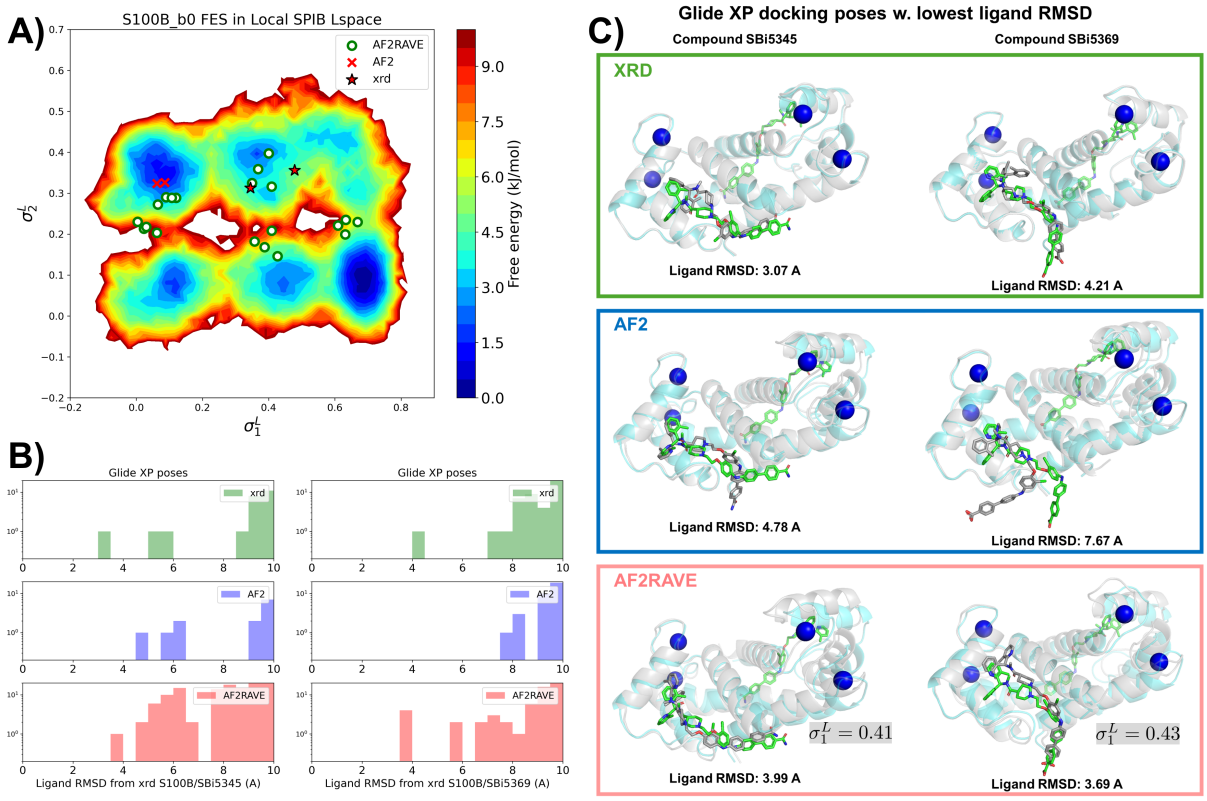


Figure 6.4: A) Projection of XRD (Ca^{2+}-S100B with SBi5345 or SBi5369, shown as red stars), AF2 (red crosses), AF2-RAVE structure set (green circles) onto the local/pocket free-energy surfaces for Ca^{2+}-S100B with native global basin. B) Ligand RMSD distribution of the Glide XP docking poses generated from the three structure sets. The left panel corresponds to docking poses with inhibitor SBi5345, while the right panel corresponds to inhibitor SBi5369. C) Representative docking poses with the lowest ligand RMSD for each structure set. Docking poses (proteins in grey, ligands as grey sticks) are superimposed onto the corresponding X-ray holo structures (proteins in cyan, ligands as green sticks) for comparison.

6.2.4 Retrospective enrichment validation on AF2-RAVE structure set of Ca^{2+}-S100B

I observe that the crystal structure (XRD) and the AlphaFold2 model (AF2) yield reasonable enrichment performance and are quite similar. However, relying on a single crystal or AF2 structure can limit enrichment, as the ligand set is structurally diverse and may not engage all the interactions that a specific structure was optimized to accommodate. The variability in enrichment observed when using different ligand-bound crystal structures is illustrated in Fig. F.19 in the SI.

One of the crystal structures exhibited side-chain variation that suggested the presence of two possible orientations for a residue, as highlighted in Fig. 6.5A. To emulate a fully blind scenario—where such prior knowledge would not be available—I allowed DOCK Prep (Chimera 1.18[178]) to automatically select one of the orientations. Interestingly, the alternate orientation resulted in improved enrichment, as indicated by the BEDROC profile in Fig. 6.5B.

Fig. 6.5B compares enrichment results obtained by averaging docking scores across all AF2-RAVE structures, with those from XRD and AF2 structures. The AF2-RAVE protocol enables rational sampling of side-chain conformations, generating structurally diverse binding pocket environments. I find that restricting to only the top-ranked structure by pLDDT, led to a marked improvement in enrichment performance. While more structures would improve the diversity, it comes at the cost of accuracy, which is crucial in blind enrichment studies I intend to pursue in the future. Fig. F.16 in the SI shows the drop in enrichment with structures ranked first and second by pLDDT for each pocket.

AF2-RAVE outperforms not only AlphaFold2 but interestingly also all available crystal

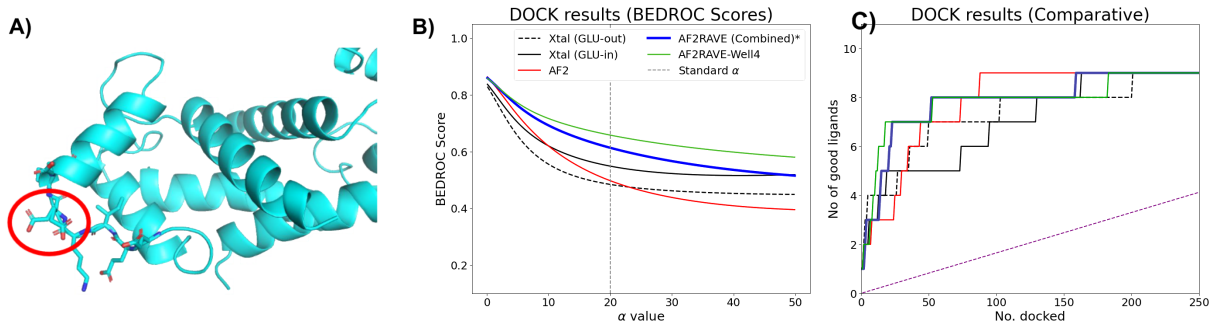


Figure 6.5: A) The side-chain assignment of GLU50 is ambiguous in the crystal structure of Ca^{2+}-S100B/SBi5345. B) BEDROC profiles and C) enrichment profiles for different structural inputs. In both B) and C), the dotted black line represents the outward-flipped orientation of GLU50, while the solid black line denotes the inward-flipped orientation. DockPrep favors the outward-facing conformation, which yields lower enrichment performance compared to the inward-facing conformation. The light green line highlights the best-performing structure (from Basin 4) in AF2-RAVE structure set.

structures by providing structural ensembles that are more effective for enrichment compared to standard AlphaFold2 predictions and ligand-specific crystal structures. This may be attributed to the importance of terminal residues in shaping the binding site—regions that are often unresolved in crystallographic experiments[179, 180]. In the case of Ca^{2+}-S100B, many binding pocket residues are also solvent-exposed, making them especially susceptible to dynamic rearrangements that AF2-RAVE is able to capture. I even find some evidence of this in the ambiguity of certain residues in the crystal structure as shown in Fig. 6.5A where two orientations are observed in the crystal structure. AF2-RAVE's ability to handle dynamic information and capture solvent interaction were crucial in obtaining this enrichment for Ca^{2+}-S100B.

6.3 Discussion

Starting from the amino acid sequence, the hierarchical AF2-RAVE pipeline enables a comprehensive exploration of the free-energy landscape and metastable states of the Ca^{2+}-S100 protein systems. It captures both large-scale backbone breathing motions and diverse holo-like

side-chain configurations missing in AF2-based methods. I find that these physically meaningful and diverse conformations generated by AF2-RAVE clearly enhanced the enrichment of true Ca^{2+}-S100B inhibitors over decoys in virtual screening, outperforming both AlphaFold2 and XRD structures.

The global SPIB model successfully learned a transferable latent space for the Ca^{2+}-S100 protein family. Combined with free-energy calculations, it revealed a dominant slow backbone motion—a breathing-like transition of the C-terminal EF-hand between a native open state and a more compact closed state—occurring on nanosecond timescales or longer. To date, all resolved holo structures of Ca^{2+}-S100 proteins bound to peptides or inhibitors adopt the native open conformation. These inhibitors, which target the open state, are typically large and highly hydrophobic to mimic and block the hydrophobic interactions of S100-binding proteins. As a result, they are not drug-like. Notably, no known ligands target the metastable closed state, presenting an opportunity to discover novel scaffolds and more drug-like compounds.

Additionally, the pocket SPIB model captured slow dynamics of residues in the C-terminal and hinge regions, associated with pocket side-chain packing transitions on picosecond timescales. It also identified free-energy basins for holo-like side-chain conformations that are not covered by AF2-based methods, for both Ca^{2+}-S100A1 and Ca^{2+}-S100B. Importantly, uncovering this side-chain-level metastability proved crucial for rigid docking applications.

To evaluate the plausibility of using my AF2-RAVE-derived holo-like models for rigid docking and virtual screening, I performed retrospective docking and enrichment tests on an unpublished Ca^{2+}-S100B inhibitor dataset. In retrospective docking tests, the AF2-RAVE structure set performs comparably to experimentally resolved holo X-ray structures and outperforms AF2 models, despite all three structure sets sharing the same backbone states. In enrichment testing,

AF2-RAVE outperforms both AF2 and X-ray structures. Surprisingly, X-ray structures show similar enrichment performance to AF2, likely due to crystallographic artifacts or ambiguities that allow certain decoys to artificially achieve high docking scores. These findings underscore the advantage of AF2-RAVE in generating physically relevant, diverse conformations optimized for rigid docking and high-throughput virtual screening.

For virtual screening that relies on rigid docking, accurately modeling holo-like protein structures requires capturing both large-scale backbone motions and subtle side-chain rearrangements. The hierarchical AF2-RAVE pipeline addresses both aspects, offering structural ensembles that reflect the dynamic nature of target proteins across different timescales. Even in flexible docking scenarios, where specific residues must be assigned to be flexible/perturbed, SPIB-learned weights can offer dynamics-informed guidance to identify the most relevant flexible residues. This holds promise for maintaining docking accuracy while significantly enhancing flexible docking efficiency compared to treating the entire binding pocket as flexible.

AF2-RAVE structure sets offer strong potential for prospective drug discovery and novel scaffold identification, particularly for challenging targets like the Ca^{2+}-S100 protein family, for which no drug-like inhibitors currently exist. Moreover, no isoform-specific inhibitors exist for the Ca^{2+}-S100A1 isoform to date. While in this work I focus on retrospective studies of Ca^{2+}-S100B specific binders, I believe this same protocol will help me in prospective discovery of Ca^{2+}-S100A1 specific binders, work which is currently being pursued. This hierarchical AF2-RAVE funnel applied to Ca^{2+}-S100 isoforms revealed that while backbone-level metastability is similar between Ca^{2+}-S100A1 and Ca^{2+}-S100B, the local SPIB analysis uncovers more significant differences in side-chain-level metastability. Using these distinct metastable states through multi-configuration virtual screening provides a promising strategy for selective inhibitor de-

sign targeting a specific isoform. Looking ahead, AF2-RAVE may enable the development of isoform-specific therapeutics targeting Ca^{2+}-S100A1 for Alzheimer's disease treatment.

Together, these chapters summarize my efforts to develop computational methods for advancing structure-based drug discovery in Prof. Pratyush Tiwary's lab. Through close collaboration with postdoctoral researchers and graduate students in the group, I worked on integrating AI-based structure prediction with physics-based sampling and thermodynamic reweighting to identify, rank, and use metastable protein conformations for drug discovery. In the upcoming chapters, I discuss how AF2-RAVE-based approaches can be extended to immunotherapy applications.

Chapter 7: Extending AF2-RAVE for improved evaluation of T Cell Receptor - Peptide Major Histocompatibility Complexes from Multimer Models

The previous chapters focused on AF2-RAVE as a way to generate and Boltzmann-rank metastable conformations of single protein chains or domains. However, many biological processes are mediated not by isolated proteins, but by protein complexes whose function depends on molecular recognition, association, and dissociation. This is especially important in immunology, where T-cell receptor (TCR) recognition of peptide-major histocompatibility complex (pMHC) depends not only on the structures of the individual partners, but also on the conformational ensemble and stability of the bound complex [181, 182, 183].

Recent cofolding models, including AlphaFold-Multimer, RoseTTAFold, Boltz-1, Chai-1, and AlphaFold3, extend deep-learning-based structure prediction from single-chain proteins to multi-chain protein complexes [10, 15, 184, 185, 186]. These methods provide a powerful route to generating structural hypotheses for protein-protein interfaces directly from sequence. Nevertheless, as with single-chain AF2 predictions, cofolding outputs are not Boltzmann-weighted ensembles and do not directly provide the thermodynamic or kinetic information needed to assess complex stability, or dissociation propensity. Therefore, a possible next question is whether AF2-RAVE like workflows can be extended from single-protein conformational ensembles to

protein-complex ensembles.

In this chapter, I explore this idea using TCR-peptide complex recognition as a model system. I use cofolding predictions as initial structural hypotheses for protein complexes and then apply physics-based sampling and reweighting to assess the stability and conformational variability of the predicted complexes. In addition to other filtering algorithms used to remove unphysical structures, an enhanced-sampling-based score is used to filter out complexes corresponding to weak binding modes. The goal is to move beyond a single predicted bound structure and toward an ensemble-based description of protein-complex recognition, where interface conformations, metastable states, and relative thermodynamic stability can be compared across related complexes. To answer this I present a short chapter on evaluation of multimer complexes from cofolding models.

7.1 TCR peptide MHC recognition

In adaptive immunology, engagement of the T-cell receptor (TCR) with peptide-major histocompatibility complex (pMHC) is a central event in antigen recognition. Productive TCR–pMHC binding can initiate T-cell activation, cytokine release, cytotoxicity, memory formation, or tolerance. Because of this central role, TCR-pMHC recognition is directly relevant to infection, autoimmunity, vaccine design, cancer neoantigen discovery, and engineered T-cell therapies.

The TCR is a heterodimer composed of α and β chains, each contributing three complementarity determining regions: CDR1, CDR2, and CDR3. In a common binding mode for TCR recognition of peptide–MHC class I, the TCR docks diagonally over the peptide-binding groove, with the TCRα chain positioned over the MHC α_2 helix and the TCRβ chain positioned over the

MHC α_1 helix. In this canonical picture, CDR1 and CDR2 primarily contact the MHC helices, whereas CDR3 loops often make more direct contacts with the antigenic peptide.

However, this canonical view should not be interpreted as a universal rule. Different TCRs can vary in docking angle, tilt, footprint, and interaction pattern depending on the TCR sequence, peptide identity, MHC sequence, and local environment. This suggests that a TCR-pMHC complex may not be fully described by a single rigid binding pose. Instead, multiple interface conformations or binding poses may exist as metastable states. Such differences in binding geometry and interface dynamics may influence downstream TCR signaling, providing a motivation to study TCR-pMHC recognition from an ensemble perspective. Thus, this problem is very suitable for the goal of this chapter.

7.2 Related Work

General complex predictors such as AlphaFold-Multimer [184] can generate candidate interfaces for many protein complexes, but TCR-pMHC modeling is a special case. The peptide is short and flexible, the TCR CDR loops are highly variable, and the biologically relevant interface is a composite surface formed by both the peptide and the MHC. As a result, general confidence metrics may penalize or obscure alternative binding modes, and the predicted complex can be strongly affected by template information. Template information is therefore both a strength and a limitation. When a related TCR-pMHC structure exists, templates can steer the model toward a plausible docking topology and prevent unrealistic placement of the TCR over the pMHC surface. This is valuable because completely template free predictions can fail by placing the TCR away from the peptide-binding groove or by generating a binding pose in which the peptide is

displaced from the groove. At the same time, templates can bias the ensemble toward known geometries and suppress alternative basins that may be biologically relevant.

Specialized TCR-pMHC methods address some of these problems by incorporating prior information about the geometry of the complex. TCRdock [187] is one such method, developed soon after AlphaFold2, that uses TCR-pMHC specific hand-made template information to bias AlphaFold2 predictions toward more reliable docking geometries. Similarly, TCRmodel2 [188] leverages the AlphaFold2 framework together with TCR-pMHC-specific sequence, template, and filtering strategies to improve complex prediction. Despite these advances, TCR-pMHC structure prediction remains challenging, largely because the number of experimentally resolved TCR-pMHC complexes is still limited compared with the diversity of possible TCRs, peptides, and MHC alleles [188]. Moreover, a general structural confidence score or energy-based score is not sufficient to fully characterize a binding pose, because environmental effects are need to be considered to estimate biophysical quantities such as K_D and k_{off}. Thus, TCR-pMHC recognition provides an ideal test case for AF2-RAVE.

7.3 Antigen Recognition and the HHAT Wild-Type/Neoantigen System

The HHAT, or hedgehog acyltransferase, system is a useful example because it links neoantigen recognition to a specific structural mechanism. For the studied HLA-A*02:06 system, the wild-type peptide is KQWLVWLLL, whereas the L75F mutant peptide, often denoted HHATp8F, is KQWLVWLFL [189, 190]. The 302TIL TCR recognizes the mutant pMHC more strongly than the wild-type pMHC, and this difference has been attributed to mutation-induced preorganization of the peptide. In particular, the Leu-to-Phe mutation alters the conformation of an upstream

tryptophan residue, positioning it more favorably for TCR engagement [189, 190].

Although PDB entries are available for the HHAT wild-type and mutant TCR-pMHC systems, they do not contain resolved coordinates for the TCR chains. However, the peptide conformations in the presence and absence of TCR binding provide useful structural information, especially for evaluating whether the predicted peptide rotamers and recognition geometry are reasonable. This makes the HHAT system a valuable test case: the full complex is challenging enough that AF2 has limited direct structural guidance, but the available peptide conformational information still provides an experimental reference for evaluating AF2-RAVE predictions.

Experimentally resolved HHAT pMHC and TCR-pMHC structures provide useful references for evaluating the predicted complexes. I used the available peptide conformations to assess the predicted HHAT peptide rotamers, and used the TCR3d-curated 6UK2 biological assembly as the full-complex reference for comparing TCR docking, CDR-loop geometry, and interface contact preservation [191]. Thus, the HHAT system is a useful test case because both peptide-level and full TCR-pMHC recognition features can be compared against experimental structural information.

7.4 Results

7.4.1 Metadynamics ranks AlphaFold-Multimer binding poses

To begin, following the AF2-RAVE workflow, I generated approximately 1900 initial structures for the TCR–pMHC complex using AlphaFold-Multimer (AF-M) [192]. Because this is a multichain protein complex, I used larger reduced-MSA depths, typically 64 or 128, rather than the smaller depths such as 8 or 16 used for single-chain proteins. In the single-chain case, the

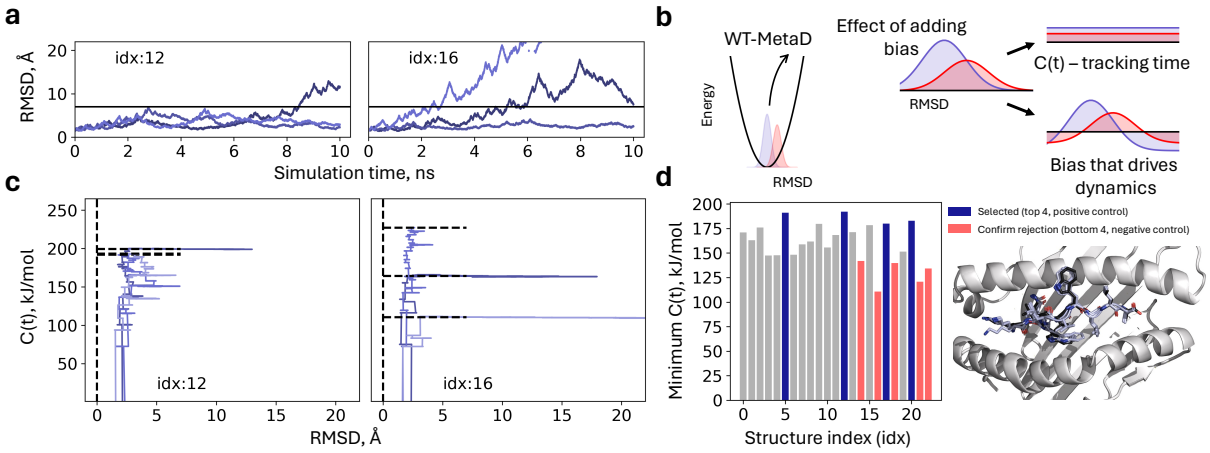


Figure 7.1: Binding-pose metadynamics for complex filtering: a) Interface RMSD, used as the biasing coordinate, plotted against simulation time for two selected initial binding poses out of 23. These two poses were selected as positive-control (left) and negative-control (right) examples. Each time series corresponds to one replicate. b) Conceptual representation of $c(t)$ between two consecutive bias-deposition steps. c) $c(t)$ versus RMSD plot depicting the average bias deposited during unbinding. d) Ranking based on the minimum value of $c(t)$. A structural comparison between the predicted peptide conformation (light violet) and the experimental conformation (black) is shown on the right.

goal is usually to perturb tertiary contacts within one folded protein. In the complex case, the goal is instead to perturb the interchain contacts that define the binding pose while preserving the folds of the individual chains. In this sense, the problem is closer to a cofolding, or fold-and-dock, setting, where both the structures of the partners and their relative interface geometry must be assessed [187].

I observed that sequence information alone did not generate structures corresponding to the experimentally observed peptide conformations and TCR-pMHC binding geometries(Fig. G.1). Therefore, similar to the S100 systems discussed in the previous chapter, I used template information to guide the initial complex generation. First, I ran a default AF-M prediction using full MSA and template information to obtain one reference complex. I then used this complex as a template and performed a reduced-MSA AF-M protocol to generate a diverse set of initial TCR-pMHC structures. Because multichain structure predictors can produce unphysical complex

models, I applied geometric and chemical filters to remove structures with broken peptide bonds, severe steric clashes, or other violations of basic molecular geometry. After filtering and applying the regular-space clustering procedure prescribed by AF2-RAVE, I obtained 23 distinct binding modes.

To select the most stable binding modes from these candidates, I performed binding-pose metadynamics. In this step, a short metadynamics was run using the interface RMSD as the biased coordinate [193]. The metadynamics parameters were intentionally chosen to strongly perturb the interface, because the goal was not to obtain a fully converged free-energy surface but rather to test the robustness of each binding pose under enhanced fluctuations. In other words, poses that remain stable even after frequent bias deposition are interpreted as stronger binding modes, whereas weakly bound or poorly packed poses are expected to dissociate or rearrange more readily. The reversible work associated with the applied bias can therefore be used as an enhanced-sampling-based score to help rank and filter the predicted TCR-pMHC binding poses. This scoring term corresponds to the $c(t)$ [6] quantity discussed in Chapter 1. I therefore performed three 10 ns replicates of binding-pose metadynamics and used the minimum $c(t)$ value as the score for selecting more stable binding poses.

As a proof of concept, I selected a positive and negative control based on the simulation and visual inspection. Essentially, I picked the one where even the interaction between the TCR chains was weak as the negative control and the one with the least change in RMSD as the positive control. The RMSD time series for the controls are shown in Fig. 7.1a. Now, the score for this test is given by the minimum of $c(t)$ (Fig. 7.1b). The behavior of $c(t)$ can also be seen in Fig. 7.1c and in Fig. G.2. Thus, based on the score, the ranks are shown in Fig. 7.1d. Clearly, our negative control was dropped, and the stable positive control was picked by this scoring

mechanism. Further, the MHC domains of the final ranked conformations were aligned with the experimentally resolved MHC domain under TCR influence. Then, the overlay of the peptide is shown in Fig. 7.1d. The peptides overlap remarkably well, thus qualitatively showing that such a filtering procedure can be useful for handling complexes.

7.4.2 AF2-RAVE for TCR-pMHC Complexes

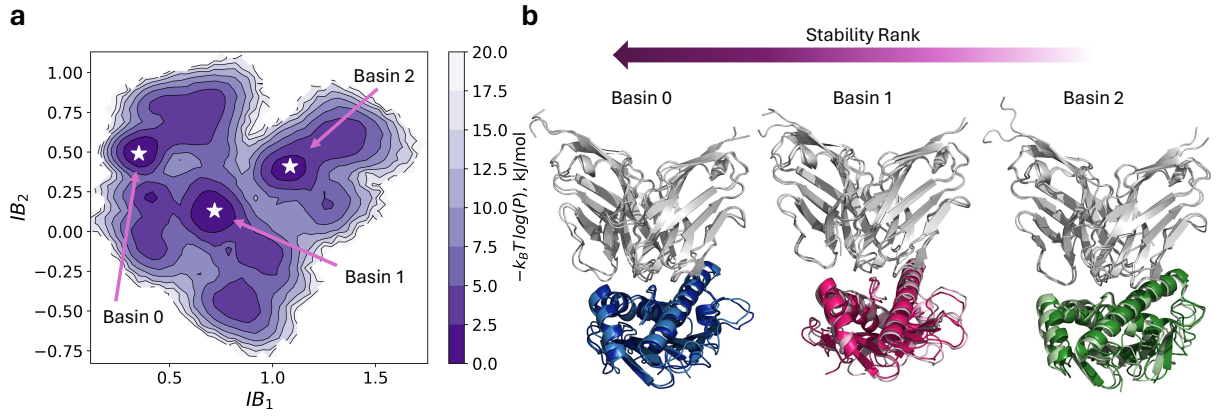


Figure 7.2: AF2-RAVE-selected conformations: a) SPIB latent space for the TCR-pMHC complex. b) AF2-RAVE picked structures (2 per basin) with TCRs (silver) aligned and fixed.

To complete the AF2-RAVE protocol, 100 ns of unbiased molecular dynamics simulation was performed for each complex, giving a total of 400 ns of unbiased simulation. Interface features were selected from all pairwise contacts at the TCR-pMHC interface using the coefficient of variation and the AMINO scheme [194] prescribed in the updated AF2-RAVE protocol [167]. This procedure produced a set of ∼20 interface features.

Using these features, SPIB-derived reaction coordinates were learned with a time lag of 2 ns (Fig. 7.2a). The learned latent space identified three metastable regions. Representative structures from these regions are shown in Fig. 7.2b. The first two basins correspond to alternative modes of peptide engagement by the TCR CDR loops, consistent with canonical TCR-pMHC recognition.

The third basin represents a distinct recognition state in which the TCRα chain forms the largest number of contacts with the peptide. Although this state is less stable than the dominant basins, it may represent a metastable recognition mode or an intermediate along the unbinding pathway.

7.4.3 Structural validation of AF2-RAVE basins

To evaluate whether the AF2-RAVE-selected basins recover the experimentally observed TCR-pMHC recognition geometry, I compared representative structures from basin 0 and basin 1 against the TCR3d-corrected 6UK2 biological assembly. For each comparison, the MHC domain was aligned first, and the resulting coordinates were used to evaluate the peptide, CDR loops, and TCR placement relative to the reference (Fig. 7.3). This alignment isolates differences in the TCR-peptide recognition geometry from small differences in the MHC scaffold.

After MHC alignment, basin 0 and basin 1 both placed the peptide close to the reference peptide conformation, with peptide backbone RMSDs around $\sim 1\mathring{A}$ (Other metrics are in Table 7.1). The two basins differed more strongly in the placement of the TCR and specifically CDR loops. Basin 1 showed closer agreement for both CDR3 loops, with TCRα CDR3 and TCRβ CDR3 backbone RMSDs around $\sim 1\mathring{A}$. In contrast, basin 0 showed larger CDR3 deviations, consistent with an alternative TCR recognition pose. To compare the interface interaction pattern, I also calculated the Q value defined in Eq. 1.19, restricting the native contact set to heavy-atom contacts between the TCR CDR loops and either the peptide or the MHC in the corrected reference structure. Both basins preserve the peptide-facing CDR contact network well, while basin 1 better preserves the broader CDR-MHC interaction network.

Together, these results show that AF2-RAVE can move beyond static structure prediction

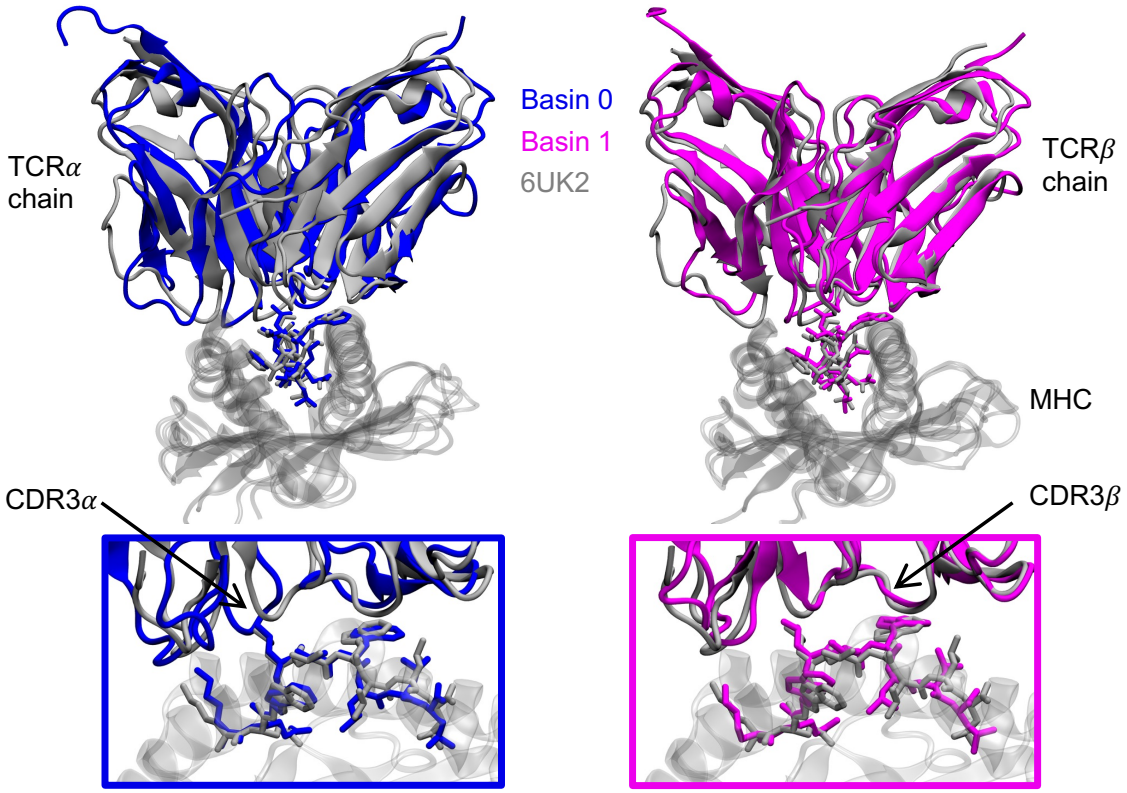


Figure 7.3: Validation of AF2-RAVE-selected TCR-pMHC structures against the crystal structure: Representative structures from basin 0 (left) and 1 (right) are aligned to the reference MHC (PDB: 6UK2). The top MHC-aligned structures compares the global TCR-pMHC docking/positioning geometry, while the bottom highlights the interface interaction between the peptide, CDR loops and MHC.

by separating predicted complexes into dynamically meaningful basins and revealing alternative modes of antigen recognition.

7.5 Conclusion and Future Work

Overall, this chapter demonstrates how to handle multiple-chain systems using reduced-MSA and AF2-RAVE protocols. I investigated the role of templates and the effect of reduced MSAs in protein-complex prediction, with a particular focus on TCR-pMHC recognition. By following the AF2-RAVE protocol, I also demonstrated the importance of geometry- and chemistry-based filters and included the binding-pose metadynamics-based filter.

Table 7.1: Quantitative validation of AF2-RAVE-selected structures after MHC alignment. RMSDs were calculated after aligning the MHC backbone to the corrected 6UK2 biological assembly. Q values were calculated using native heavy-atom contacts between the CDR loops and either the peptide or the MHC.

Metric	Basin 0	Basin 1
Peptide all-heavy atom RMSD (Å)	1.33	1.53
TCR backbone RMSD (Å)	5.38	2.14
TCRα CDR3 RMSD (Å)	3.31	1.02
TCRβ CDR3 RMSD (Å)	3.01	0.84
All CDR backbone RMSD (Å)	4.01	2.33
Q value, CDR-peptide	0.94	0.94
Q value, CDR-MHC	0.63	0.76

Finally, I completed the AF2-RAVE workflow and identified three distinct metastable states for the TCR-pMHC complex. These results were evaluated using the experimentally resolved pMHC structure in the presence of the TCR as a reference. For the selected peptide rotamer, the TCR appears to sample three stable or metastable binding modes. Future analysis of the neoantigen mutant, together with additional unbinding simulations, will help clarify the molecular basis of TCR recognition for this pMHC ligand. This direction is particularly important from an immunology perspective and from a methods perspective, given the current limitations of cofolding models for predicting flexible and weakly stable biomolecular complexes.

Chapter 8: CAR-BOOSTing T-Cell Responsiveness through Synaptic Cleft Engineering

[This chapter is based on: Justin M. Mirazze, **Akashnathan Aranganathan***, Jaehyun Suh*, Xiang Chen*, et al. CAR-BOOSTing T-Cell Responsiveness through Synaptic Cleft Engineering. Nature Chemical Biology. (**under revision**)(*- equal contribution)[195]. I contributed to the method development and analysis of cryo-ET data for synaptic cleft calculation.]*

8.1 Introduction

The engineering of T-cells with chimeric antigen receptors (CARs)—synthetic proteins that enable antigen recognition independently of MHC complexes—has undergone major advances in recent years[196, 197, 198, 199]. CAR T-cells have demonstrated success in treating relapsed/refractory B-lineage leukemia and lymphoma. However, many patients fail to stay in remission, often due to antigen loss on leukemic cells[200, 201, 202, 203, 204]. Poor CAR T-cell cytotoxicity against tumors expressing $< 10^3$ ligands[205, 206] is strongly associated with relapse. To address such immune escape, clinicians have developed strategies, including: 1) targeting multiple antigens[207, 208]: 2) incorporating CAR elements into the TCR complex through T-cell receptor fusion constructs (TruCs)[209], synthetic TCR and antigen receptors (STARs)[210], in which the VH and VL domains are fused to TCRα and β chains, re-

spectively, and HLA-independent TCRs (HITs), which similarly couple antigen recognition to the endogenous TCR complex while incorporating 4-1BB costimulation and, in some designs, co-expressed CD80/4-1BBL signaling domains[211]); 3) harnessing TCR/CAR crosstalk to enhance anti-tumor cytotoxicity while sparing healthy tissues[212]; and 4) replacing the CD8α hinge/transmembrane with CD28[213], amongst others. While these approaches have all increased CAR responsiveness, the determinants of CAR sensitivity remain incompletely understood.

Here, we investigated how CAR design parameters influence sensitivity using CD19-targeting constructs with varied hinge, transmembrane, and co-stimulatory domains, and found that—within specific hinge/co-stimulatory combinations—the hinge region exerted the strongest impact. To enable rational design, we developed a computational pipeline—Chimeric Antigen Receptor Biophysically Oriented Optimization of Synaptic Topology (CAR-BOOST)—that tunes hinge domains to the biophysical constraints of the native immune synapse. We validated CAR-BOOST across CD19-, CD20-, CD22-, and CD33-targeting constructs, in settings spanning both liquid and solid tumors.

8.2 Results

8.2.1 The hinge region of the CD19 CAR construct regulates CAR T-cell responsiveness to CD19^{low} leukemia

Tisagenlecleucel (tisa-cel) and axicabtagene ciloleucel (axi-cel) constructs incorporate a CD19-targeting FMC63 scFv and a CD3-ζ signaling domain (Fig. 8.1a), but differ in other modular components. Tisa-cel includes a 45-amino acid hinge (H) and transmembrane domain

(TMD), both derived from CD8α, together with an intracellular 4-1BB costimulatory (CS) domain. In contrast, axi-cel includes a triple alanine linker and a 39-amino acid hinge derived from CD28, together with a CD28 TMD and CS domain[199, 214] (Fig. 8.1a). To directly compare these CARs, we cloned them into lentiviral vectors[214] with modified spacers for clonability and comparability. Both CD28H-TMD-CS (axi-cel-like) and CD8αH-TMD-4-1BB CS domain (tisa-cel-like) constructs efficiently killed NALM6 leukemia with high CD19 expression ($\sim 7 \times 10^4$ molecules/cell), but only the axi-cel-like CAR (CD28-based) was cytotoxic in vitro against CD19$^{\text{low}}$ cells ($\sim 1 \times$ 10^3 CD19 molecules/cell) (Fig. 8.1b).

We then systematically explored how permutations of H, TMD, and CS domains influence surface expression and cytotoxic potential of FMC63-bearing CARs (Fig. 8.1c and Supplementary Fig. H.1a,b). Notably, constructs containing a 28H-8T combination, irrespective of the CS domain, failed to express at the cell surface and therefore could not be included in our functional comparisons (Supplementary Fig. H.1b). Among the remaining constructs, all mediated efficient killing of CD19$^{\text{high}}$ leukemia; however, only CARs incorporating a CD28 hinge, regardless of whether they contained a CD28 or 4-1BB CS domain, retained high cytotoxicity against CD19$^{\text{low}}$ leukemia (Fig. 8.1d). Indeed, linear regression analysis of the killing index revealed the H design as the dominant factor influencing cytotoxicity, with the CS domain contributing more modestly (Fig. 8.1e,f), in particular when targeting CD19$^{\text{low}}$ cells (Fig. 8.1g).

8.2.2 Optimizing the CD8α-derived hinge for increased responsiveness of the tisa-cel-like CAR to CD19^{low} leukemia

Our previous studies[215, 216] along with additional analyses of the intrinsic disorder in the CD8α-derived and the CD28-derived hinges (Supplementary Fig. H.1c,d) highlighted the structural distinctions of these hinges. We therefore systematically tested whether truncation of the CD8α hinge in the tisa-cel-like CAR would alter its efficacy against CD19^{low} tumors. We generated a series of CD19 (FMC63) CAR designs using the CD8α hinge as a framework, either truncating the hinge at the N-terminus or extending the hinge with a flexible linker composed of glycines (G) and serines (S) (Fig. 8.2a). We then assessed expression of all CAR constructs on transduced T-cells; due to differences in cell surface detection, with lower detection of CARs harboring shorter hinges, transduction levels were evaluated via a truncated form of EGFR (tEGFR) containing extracellular domain III and the EGFR transmembrane domain[217] (Supplementary Fig. H.1e,f).

We first evaluated the different CAR-T designs for their ability to bind target cells. While this parameter was found to be predictive of in vivo tumor eradication in previous studies[218, 219], in vivo potential is influenced by a convolution of multiple parameters including cytotoxicity, expansion, resistance to exhaustion, and others[220, 221, 222, 223]. We found that CAR T-cell binding to antigen, assessed using the Lumicks zMovi® platform, did correlate with surface staining (Supplementary Fig. H.1f-h), but it did not correlate with tumor cytotoxicity (Supplementary Fig. H.1h and Fig. 8.2b). Of note, CAR constructs with shorter CD8α hinge lengths exhibited lower target binding but higher cytotoxicity.

We then assessed the responsiveness of CAR T-cells with different CD8α hinge lengths

as a function of their cytokine secretion profiles. Using a quantitative IMMUNOtron robotic platform[224], assessing kinetic responses as a function of CAR ligand abundance (Supplementary Fig. H.2a-d), we found that IL-2 was the most suitable readout for the comparative evaluation of our modular CAR designs. IL-2 secretion profiles against target cells expressing varying levels of CD19 could be fitted and parameterized with the EC_{50} (i.e. the abundance of CD19 required to achieve 50% of the maximal response for each CAR construct). We found that the CD19 CARs with 8H15 and 8H27 exhibited 9- and 30-fold improvements in IL-2 EC_{50} relative to the canonical tisa-cel-like 8H47CAR (Fig. 8.2c), respectively. Evaluation of IFNγ production by the different CARs also identified the 8H27 CD19 CAR as the most responsive to antigen^{low} leukemia (Supplementary Fig. H.2e). For IFNγ, EC_{50}s, across 4 donors, 8H27 showed significantly lower EC_{50} than 8H15 (paired t-test on $\log_{10}$ EC_{50}, one-sided p = 0.006), corresponding to an ˜4-fold improvement in potency (Supplementary Fig. H.2e). We leveraged these measurements to generate a comprehensive map of the responsiveness of all CAR constructs across all CAR ligand densities (Fig. 8.2d), revealing a strong correlation between cytotoxicity and IL-2 EC_{50} (Fig. 8.2e). Truncating the CD8α hinges in CD19 CARs—from the canonical 8H47 to shorter variants such as 8H27, 8H15, or 8H8—significantly enhanced CAR-T responsiveness. Among these, 8H15 and 8H27 demonstrated superior performance compared to all others, suggesting an optimal hinge length for maximizing CAR activity (Fig. 8.2b,c and Supplementary Fig. H.2e,f). We then evaluated three CAR-T designs incorporating the canonical 8H47 hinge or the shortened 8H8 or 8H15 hinges in vivo, in NOD.Cg-Prkdc^{scid}Il2rg^{tm1Wjl}/SzJ (NSG) mice engrafted with NALM6 leukemia. All constructs efficiently eradicated CD19^{high} NALM6 cells in NSG mice; however, 8H15 CD19 CAR T-cells mediated the most durable control of CD19^{low} NALM6 tumors (Fig. 8.2f). Consistent with this finding, the median time to tumor relapse was

approximately 10, 17, and 27 days for mice treated with 8H47, 8H8, and 8H15 CD19 CAR T-cells, respectively (Fig. 8.2g). Notably, time to relapse inversely correlated with in vitro IL-2 secretion EC_{50} measured for each CAR design (Fig. 8.2g), supporting antigen-dependent IL-2 production as a robust ex vivo surrogate for in vivo anti-tumor performance.

Given the superior clinical efficacy of Axi-cel (28H-28T-28CS) relative to Tisa-cel (8H-8T-41BBCS) in B cell lymphoma[225, 226, 227] and prior evidence that replacing the CD8α hinge with a CD28 hinge enhances antigen sensitivity[213], we next assessed whether our optimized shortened CD8α hinge (8H15 and 8H27) could outperform an Axi-cel-like benchmark (28H42-28T-28CS (28H42-28)). Indeed, in mice engrafted with $CD19^{low}$ leukemia, treatment with 8H15- and 8H27-based CARs resulted in significantly improved tumor control and survival compared with either the conventional 8H47-BB (tisa-cel-like) or 28H42-28 (axi-cel like) constructs ($p<0.001$ at days 18 and 25 for both donors; mean survivals of 46 and 43 days vs. 29 and 34 days, respectively; Fig. 8.2h and Supplementary Fig. H.2g,h). Moreover, incorporation of the optimized 8H27 hinge into the Axi-cel–like backbone (8H27-28T-28CS (8H27-28)) significantly enhanced anti-tumor efficacy relative to the standard Axi-cel–like construct ($p<0.0001$ at day 25 for both donors, Fig. 8.2h and Supplementary Fig. H.2g,h) although it did not fully recapitulate the performance of the corresponding 8H27 CAR incorporating a 4-1BB costimulatory domain. Collectively, these data demonstrate that CD19 CARs bearing shortened CD8α hinges and 4-1BB costimulation outperform conventional CD28 hinge-based CAR architectures, underscoring the critical contribution of integrated CAR structural design—rather than co-stimulatory domain choice alone—to effective tumor clearance in low-antigen settings.

These optimized CARs exhibited markedly enhanced responsiveness, with up to 100-fold increases in antigen sensitivity across liquid tumors, monitored as a function of IL-2 secretion

(Supplementary Fig. H.2i). The CD8α hinge length dependency similarly extended to the eradication of a CD19-expressing melanoma solid-tumor model (Supplementary Fig. H.2j). Thus, the responsiveness of CD19 CAR T-cells against CD19^{low} tumors—measured by cytokine secretion as well as in vitro and in vivo killing—proved to be highly sensitive to the CAR's hinge design, with dramatic changes resulting from the deletion of as few as 7 amino acids in the CD8α hinge and an optimal sequence length of 15-27 amino acids.

8.2.3 Modeling CAR-generated synaptic clefts using an AlphaFold2RAVE (AF2RAVE)-based model.

Building on these findings, we hypothesized that the CD8α hinge sequence length modulates the depth of the CAR-induced synaptic cleft, enabling spatial conformation with the canonical TCR immune synapse (IS) geometry and thereby optimizing CAR T-cell responsiveness[228, 229, 230, 231]. Consistent with this model, we recently demonstrated that incorporation of TCR signaling components, such as ZAP-70, into a CD19-directed CAR alters both its recruitment and dwell time at the immunological synapse compared with CARs containing only a co-stimulatory domain[232]. Guided by these observations, we therefore first characterized the geometric features of the endogenous TCR IS[232, 233, 234, 235, 236, 237, 238]. The endogenous TCR-driven synapse is composed of a central supramolecular activation cluster (cSMAC) accumulating TCR-MHC, CD2-CD58, and CD28-CD80/86 complexes, a peripheral SMAC (pSMAC) with ICAM-1/LFA-1 complexes, and a distal SMAC (dSMAC) composed of CD45. Recently, it has been shown that some co-stimulatory protein receptors, such as CD2, move toward the dSMAC and are presented in the corolla region of the canonical TCR synapse[239, 240]. We explored

how this spatial organization imposes intermembrane distance constraints essential for T-cell activation (Fig. 8.3a) by modeling these complexes using AlphaFold-Multimer with reduced MSA input. We found that a functional synapse (cSMAC) must have a very tightly constrained depth between 12.5 and 14.5 nm (Fig. 8.3b, Supplementary Fig. H.3). These bounds defined three intermembrane regions: one central functional synapse region (S) and two peripheral regions, O_1 and O_2. We hypothesized that CAR/CAR ligand complexes must accumulate specifically within the S region—rather than in O_1 or O_2—to contribute to cSMAC formation and robust downstream signaling.

We then developed a protocol to computationally estimate this partitioning, assuming independent end-to-end (z_{e2e}) distance distributions for contacting receptors, enabling target-agnostic transferability and reduced computational cost (Fig. 8.3c). We validated our hypothesis assumption on the TCR IS using reduced MSA AlphaFold2 (rMSA-AF2)[91] as the distribution generator; the method qualitatively places cSMAC receptor complexes (Fig. 8.3d and Supplementary Fig. H.4) within the functional synapse region while pSMAC proteins (CD45 and ICAM-1/LFA-1 complex) were too large to fit within a cleft that was smaller than 14.5 nm (Fig. 8.3d). We then implemented an AF2RAVE-based conformational distribution generator[94] to evaluate the responsiveness of CARs with different CD8α hinge lengths, producing transferable z_{e2e} distributions (Fig. 8.3e,f, Supplementary Fig. H.5, Supplementary Fig. H.6). As opposed to other recent distribution generators[241, 242, 243, 244], we favored AF2RAVE for its ability to reweight rMSA-AF2 ensembles by accounting for environmental variables such as temperature, pressure, ion concentration and others using enhanced sampling molecular dynamics[64, 94] (Supplementary Fig. H.6). Furthermore, combining physics-based sampling with a data-driven generator enables a more efficient modelling of disordered protein regions (Appendix H, Modelling Meth-

ods). To our knowledge, this is the first study to optimize CAR design based on modeling intermembrane synapse distributions and matching the architecture of the TCR IS.

8.2.4 CAR T-cell responsiveness to a $CD19^+$ tumor is dependent on the intermembrane distance distribution of the CAR IS.

The distributions of in silico intermembrane distances—modeled for T-cells harboring CD19-targeting FMC63-based CARs with distinct CD8α hinges (Fig. 8.4a) interacting with CD19-expressing leukemia target cells—revealed that an IS formed by an 8H27 CAR is more likely to fall within the S region (13.5 $\pm$ 1 nm) as compared to other CAR constructs (Fig. 8.4b). It was therefore of interest to directly measure how different CD19 CAR constructs impacted the actual widths of the immunological synapses generated by CAR T-cell/target cell interactions.

We first considered using conventional electron microscopy (EM) to estimate the depth of the synaptic cleft[245] assembled by our different designs of CAR. However, EM relies on fixation and metal-sputtering for imaging, which can cause relatively large standard deviations for distance measurements. In contrast, standard deviations of < 1 nm have been reported[246] for microtubule measurements performed by cryogenic electron tomography (cryo-ET) (Fig. 8.4c). Indeed, using cryo-ET, we measured a standard deviation for the 25 nm microtubules of 0.9 nm (Supplementary Fig. H.7b). Most importantly, we found that the average depths for synaptic clefts associated with the 8H8, 8H15, and 8H47 CARs were 11.57$\pm$0.79, 12.22$\pm$0.83, and 14.19$\pm$0.94 nm, respectively (p=0.03 for 8H8 vs. 8H15, p=0.05 x 10^{-7} for 8H15 vs. 8H47, and p=0.04 x 10^{-12} for 8H8 vs. 8H47; Fig. 8.4c, Supplementary Fig. H.7c-e, and Supplementary Fig. H.8), quantitatively matching our theoretical predictions for the functional synapse region

(Fig. 8.4d). We then derived a simple scaling law for the depth of the synaptic cleft as a function of the hinge length and predicted that a 8H27 FMC63-based CAR would generate a 13.4 nm-wide synapse that best matches the geometry of the natural immunological synapse and results in optimal activation (Fig. 8.4e). Finally, our results revealed that the 8H47 tisa-cell-like CAR creates a suboptimal 14.2 $\pm$ 0.9 nm synaptic cleft (Fig. 8.4c)—likely accounting for its reduced potency compared to the optimal 8H27 CAR design when targeting $CD19^{low}$ tumor cells (Fig. 8.2b,c,d,e).

We then introduced a quantitative parameter ϕ, to correlate our structural predictions with our functional measurements of CAR T-cell activation. We computed ϕ as an ensemble parameter that sums the energetic and entropic costs of localizing in the O_1 or O_2 region as compared to the S region of the synapse (Fig. 8.4b,f). When the CAR/CAR ligand pair resides in the O_1 region (z_{e2e} $<$ 12.5 nm), it displaces other synaptic interactions (e.g. TCR-MHC, CD28-CD80), resulting in a suboptimal synapse. In contrast, when the pair localizes to the O_2 region ($z_{e2e} >$ 14.5 nm), it fails to physically exclude the CD45 phosphatase, thereby limiting CAR activation. Hence, ϕ = ΔG_1 + ΔG_2 is the free energy cost of accurately localizing the CAR/CAR ligand in S to the exclusion of O_1 and O_2. In other words, this parameter quantifies the relative probability of finding the CAR/CAR ligand complex within S compared to outside S in O_1 or O_2 (Appendix H, Modelling Methods). We found a quantitative agreement between the in silico model ϕ values and functional readouts such as IL-2 EC_{50} in vitro and time to tumor relapse in vivo (Fig. 8.4g). Note that the longer hinges—8H61 and 8H76—were outliers (Fig. 8.4g), likely because of the computational undersampling arising from their larger sizes. Altogether, our cryo-ET measurements matched our structural model and highlighted the critical role of the CAR design—and in particular, its hinge—for matching the natural geometry of the immunological synapse and achieving optimal CAR partitioning in S and CAR responsiveness.

8.2.5 CAR-BOOSTing model predicts maximally responsive hinges across multiple CAR targets

Functional and conformational studies of the CD19 FMC63-based CAR led to the development of our CAR-BOOST model for CAR T-cells. Our computational pipeline relies on epitope location to determine the optimal hinge length for maximizing responsiveness, based on CD19 CAR T-cell measurements and synapse modeling. Since our model is based on the functional synapse gap (S), we next experimentally validated our CAR-BOOST predictions across panels of CD8α hinge variants and scFvs targeting clinically relevant leukemia antigens—CD22 (m971)[247], CD20 (Leu-16)[248], and CD33 (SGN33)[249] —using tumor cell lines engineered to express low levels of each respective antigen (Fig. 8.5a). CAR-BOOST modeling predicted distinct epitope distances from the target cell membrane, with CD22, CD20, and CD33 epitopes spanning membrane-distal to membrane-proximal positions (Fig. 8.5a-c). Given the dynamic and flexible nature of these target proteins, we generated conformational ensembles to capture the distribution of epitope locations relative to the extracellular leaflet of the leukemia cell membrane (Fig. 8.5b and Supplementary Fig. H.9). Based on these predicted epitope distances, we designed CARs incorporating either extended CD8α hinges (up to 91 aa) for membrane-proximal epitopes (e.g., CD22) and progressively shortened hinges (as short as 3 aa) for membrane-distal epitopes (e.g., CD33), enabling direct experimental testing of CAR-BOOST-guided, target-specific hinge optimization.

Given the $\sim$14.5 nm upper bound for the immune synapse depth and the typical diameter of an scFv ($\sim$3.5 nm), the optimal hinge region should not exceed 11.0 nm. We then experimentally validated our CAR-BOOST predictions across the panels of CD8α hinges and scFVs targeting

CD20, CD22, and CD33, using cell lines engineered to express low levels of the respective target antigens. Using IL-2 secretion as a measure of T-cell activation (Fig. 8.5d), we tested hinges ranging from 3 to 91 aa as a function of the epitope location and identified optimal hinge lengths of 51, 47, 27, and 3 amino acids for CD20, CD22, CD19, and CD33 CARs, respectively—showing an inverse correlation with epitope distance from the cell membrane. We then compared the theoretical ϕ values with IL-2 EC_{50} for each CAR T-cell/leukemia target pair and found a linear relationship between CAR-BOOST-predicted ϕ values and the responsiveness of the respective CAR T-cells (Fig. 8.5e).

In the context of the CD22 CAR—which has demonstrated significant clinical success in patients with relapsed/refractory $CD22^{+}$ B-ALL[250, 251]—we found that the chosen 8H47 hinge exhibited the highest ex vivo responsiveness, as predicted by CAR-BOOST. In the context of the CD33 CAR, we observed that the shortest hinge (8H3), comprising only 3 amino acids, elicited the highest levels of IL-2 secretion in response to $CD33^{low}$ leukemia, raising the possibility that the comparatively poor clinical responses reported with a CD33 CAR incorporating a longer 42-aa CD28 hinge may, at least in part, reflect suboptimal synaptic geometry for low-antigen-density target recognition. In contrast, CD20 CAR T-cells exhibited optimal function against $CD20^{low}$ NALM6 when a longer hinge (8H51) was incorporated—consistent with the predicted proximity of the CD20 epitope to the target cell membrane. Since ϕ values are related to free energy differences, CARs that achieve more negative values (compared to positive ones) would demonstrate better activation profiles. This was confirmed experimentally for CD22, CD20 (Fig. 8.5e) and CD19 CARs (Fig. 8.4g). For CD33 CARs, all constructs, including 8H3, exhibit a positive ϕ value because the target protein alone positions the epitope in the O_2 region—outside of the optimal range for the generation of an immunological synapse < 14.5 nm

(Fig. 8.5b,e).

As expected from our modeling and experimental in vitro studies, CD22 CARs with 8H47 and 8H76 exhibited superior in vivo eradication of $CD22^{low}$ NALM6 (Fig. 8.5f,g). We also estimated the time to tumor relapse, which occurred at 1, 8, 14, and 16 days in mice treated with the 8H15, 8H91, 8H76, and 8H47CD22 CAR T-cells, respectively (Fig. 8.5h). These times were inversely correlated with the in vitro IL-2 EC_{50} of the different CD22 CARs as well as the in silico CAR-BOOST ϕ values (Fig. 8.5h). Our systematic evaluation of CARs targeting distal, median, and proximal epitopes in the CD33, CD19, CD22, and CD20 cell surface proteins validates the predictive power of the CAR-BOOST algorithm. We then extended CAR-BOOST validation to CD28- and IgG4-based hinge architectures. AF2RAVE more accurately predicted z_{e2e} distance distributions than rMSA-AF2 (Supplementary Fig. H.10a,b). These distributions aligned with optimal synaptic cleft spacing and ϕ values for CD19 CARs incorporating 8H15, 8H27, and IgG4H15 hinges, but not 28H42 (Supplementary Fig. H.10), paralleling their superior in vitro and in vivo activity (Fig. 8.2g, Supplementary Fig. H.2g,h, and Supplementary Fig. H.10c). Thus, when the CAR epitope position is identified, CAR-BOOST can leverage the endogenous TCR-pMHC functional synapse gap (S), and the thermodynamic ϕ parameter can be used to predict the optimal length for maximizing CAR sensitivity to $\mathrm{antigen}^{low}$ tumors (Fig. 8.4g and Fig. 8.5e).

8.3 Discussion

CAR-BOOST establishes a quantitative framework for precisely tuning CAR T-cell responsiveness through the rational engineering of synapse geometry. Our study identifies the hinge region as a central regulator of CAR sensitivity and introduces a systematic strategy for

optimizing CAR design. Prior work showed that CD19-directed CARs incorporating a CD28 hinge exhibit enhanced responsiveness against $CD19^{low}$ tumor cells compared with CARs bearing a CD8α hinge[213]; however, the mechanistic basis for this observation remained unresolved. Resolving this question is of direct clinical relevance, as the axi-cel construct—which incorporates a CD28-hinge and transmembrane domain—has demonstrated superior efficacy relative to tisa-cel in patients with B-cell lymphoma, but at the cost of substantially increased immune cell-associated neurotoxicity syndrome (ICANS) and cytokine release syndrome (CRS), as well as higher T-cell exhaustion profiles[225, 227, 252].

Preliminary modeling identified hydrophobic residues in the N-terminal portion of the CD28 hinge, originating from the Ig-like V-type domain, as promoting a collapsed structure[215, 228]. Similar qualitative insights have been reported[213, 228, 253, 254, 255, 256, 257, 258, 259, 260], but the hypothesis that the end-to-end distance of the CAR hinge regulates CAR-T responsiveness had never been tested. Here, we developed CAR-BOOST by integrating rMSA-guided AF2 modeling with AF2RAVE-generated conformational ensembles, and show that geometric optimization of CAR hinge regions to match immunological synapse dimensions can be leveraged to enhance CAR T-cell responsiveness against antigen-low tumors. Systematic variation of hinge length across multiple CAR targets validated the target-specific nature of this design strategy. Notably, our findings provide a quantitative mechanistic framework that explains prior qualitative observations suggesting that longer hinges are more adequate for membrane-proximal epitopes, whereas shorter hinges are advantageous for membrane-distal epitopes[213, 228, 253, 254, 255, 256, 257, 258, 259, 260]. Consistent with this model, truncation of the CD8 hinge diminished the activity of EGFR-directed CAR T-cells targeting a membrane-proximal epitope, but not a distal epitope[256]. Nonetheless, whether CAR T-cell responsiveness is directly governed

by the end-to-end distance imposed by the hinge region has not been previously quantitatively or systematically tested.

Specifically, our study yields four key insights. First, we define an optimal CAR synaptic cleft distance of 13.5 $\pm$ 1 nm, closely matching the width of a canonical TCR-mediated synaptic cleft[245]. Second, we introduce CAR-BOOST as a computational pipeline that predicts CAR responsiveness based on epitope membrane proximity and CAR structural parameters. Third, we experimentally validate this model by demonstrating a nonlinear relationship between CAR architecture (with variable hinge lengths), synaptic cleft dimensions measured by cryo-EM, and T-cell responsiveness assessed in vitro. Finally, we show that CAR constructs optimized to geometrically match the TCR-pMHC synaptic cleft confer prolonged survival in preclinical humanized tumor models.

CAR-BOOST further enables the preemptive exclusion of non-ideal epitopes. For example, the membrane-distal CD33 epitope targeted by SGN33 imposes a synaptic cleft distance exceeding 15 nm, independent of hinge length, and is therefore predicted to yield suboptimal CAR activation. Consistent with this prediction, we observed a deleterious 10-fold increase in EC_{50} compared with CD19- and CD22-directed CAR T-cells. In addition, unstructured hinge sequences longer than $\sim$51 amino acids—corresponding to an average end-to-end distance exceeding 11.3 nm—are unlikely to confer gains in CAR T-cell responsiveness. This rationale obviates the need to model increasingly long and structurally intractable long hinge variants such as 8H61, 8H76, and 8H91. In the future, it will be important to incorporate the spatial organization and lateral distributions of CARs within lipid bilayers. Nevertheless, in agreement with the present findings, prior studies have shown that CAR activation can be tuned by receptor/antigen dimensions that exclude CD45 from the immunological synapses[213, 228, 253, 254, 255, 256,

257, 258, 259, 260]. CAR-BOOST is a more interpretable and predictive framework for CAR design, supporting the incorporation of axial spacing constraints as a rational principle for next generation approach CAR engineering.

While CAR-BOOST is immediately relevant to hematological malignancies, this model can be extended to solid tumors and integrate with advanced technologies. By modifying the hinge length, CARs can potentially discriminate target cells from healthy cells that express different levels of the same antigen, adding an additional layer of specificity. Combining CAR-BOOST with dual or triple-targeting strategies could also mitigate tumor escape[261]. Furthermore, CAR-BOOST can be updated by refining S region classification and incorporating target-dependent CAR z_{e2e} distributions, moving beyond current assumptions of strict boundaries and independence conditions.

Our findings lay the groundwork for a new paradigm in CAR engineering, where computational modeling informs rational and epitope-specific CAR design. This approach, integrating in silico and in vitro validation, will support the development of more precisely targeted immunotherapies, ultimately enhancing therapeutic efficacy and improving patient outcomes.

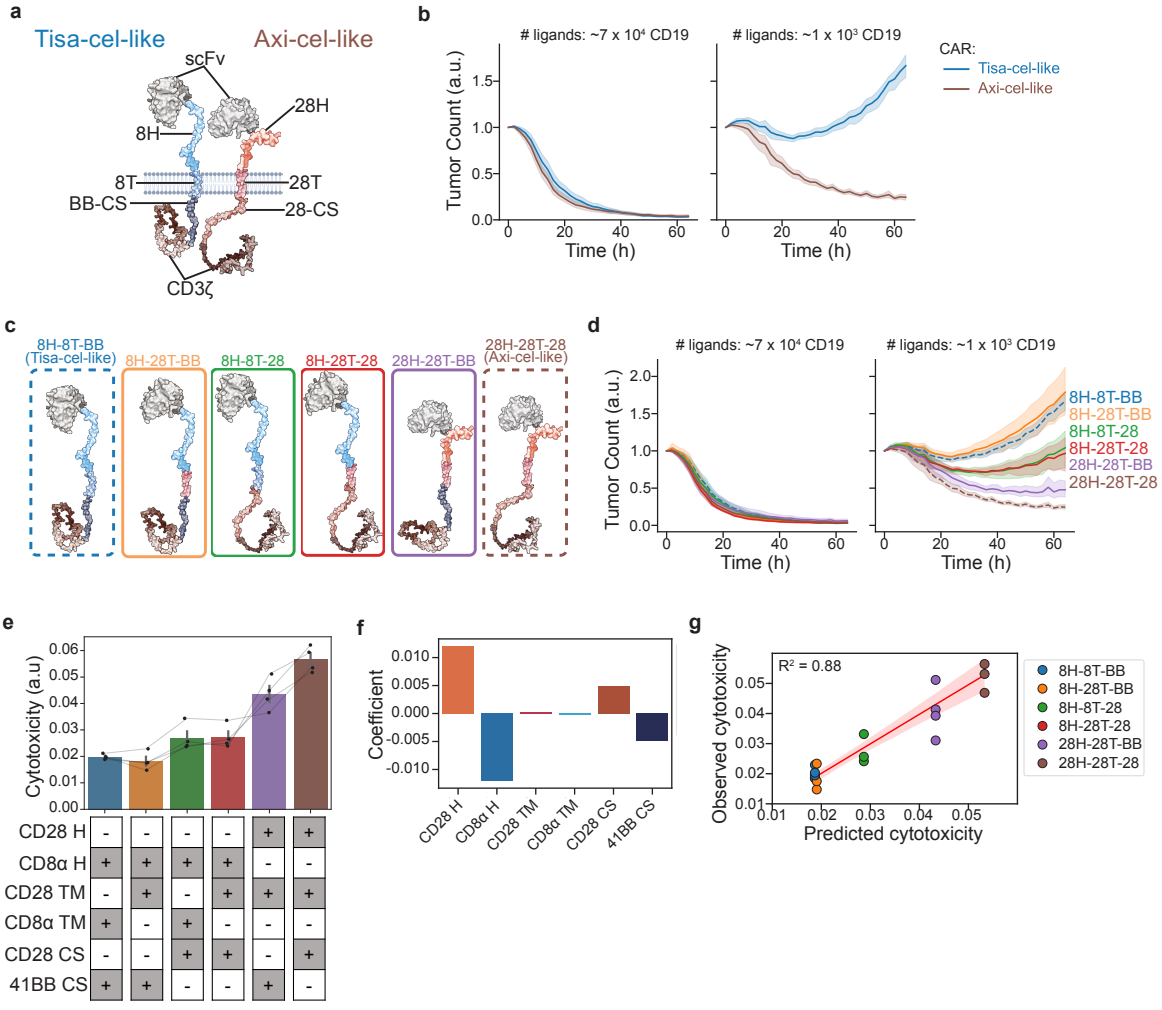

a
Tisa-cel-like
Axi-cel-like
scFv
28H
8H
8T
28T
BB-CS
28-CS
CD3ζ
b
ligands: ~7 x 10⁴ CD19
ligands: ~1 x 10³ CD19
Tumor Count (a.u.)
Time (h)
CAR:
Tisa-cel-like
Axi-cel-like
c
8H-8T-BB (Tisa-cel-like)
8H-28T-BB
8H-8T-28
8H-28T-28
28H-28T-BB
28H-28T-28 (Axi-cel-like)
d
ligands: ~7 x 10⁴ CD19
ligands: ~1 x 10³ CD19
Tumor Count (a.u.)
Time (h)
8H-8T-BB
8H-28T-BB
8H-8T-28
8H-28T-28
28H-28T-BB
28H-28T-28
e
Cytotoxicity (a.u)
CD28 H
CD8α H
CD28 TM
CD8α TM
CD28 CS
41BB CS
f
Coefficient
CD28 H
CD8α H
CD28 TM
CD8α TM
CD28 CS
41BB CS
g
R² = 0.88
Observed cytotoxicity
Predicted cytotoxicity
8H-8T-BB
8H-28T-BB
8H-8T-28
8H-28T-28
28H-28T-BB
28H-28T-28

Figure 8.1: The hinge region of the CD19 CAR construct regulates responsiveness to CD19^{low} leukemia

a, Schematic depicting the tisa-cel-like and axi-cel-like CD19 CAR constructs, highlighting their distinct hinge (H), transmembrane (TM), and costimulatory (CS) domains. b, The cytotoxic potential of axi-cel-like and tisa-cel-like CAR T-cells against clonal GFP$^+$ NALM6 lines expressing 7x10^4 (left) and 1x10^3 (right) CD19 molecules was assessed by Incucyte®. Tumor counts (a.u.) between 0 and 64 hours are presented as the ratio of GFP$^+$ NALM6 cells/mm^2 at each time point relative to the GFP$^+$ NALM6 cells/mm^2 at time 0 (representative of n=4 individual T-cell donors.. c, Schematic of novel domain-swapped CD19 CAR constructs, shown with solid borders, and their comparisons with the tisa-cel-like and axi-cel-like CARs, shown with dashed borders. d, The cytotoxic potential of the domain-swapped CD19 CARs was evaluated against NALM6 clones by Incucyte® as presented in panel b (representative of n=4 individual T-cell donors). The solid lines represent the domain-swapped CAR constructs, and the dashed lines represent the tisa-cel-like and axi-cel-like CARs. e, Bar plot and truth table depicting the normalized cytotoxicity of the different modular CAR T-cells, defined as the inverse of the area under the normalized tumor-count curve from 0 to 64 hours; higher values indicate greater tumor killing. Points represent donor-level averages, and lines connect values derived from the same donor across constructs, highlighting paired comparisons (n=4 individual T-cell donors). f, Bar plot of the model coefficients for the fitted linear regression of the H, TM, and CS domains, presenting the impact of each independent variable on cytotoxicity. g, Regression fit between the predicted and experimental normalized cytotoxicity values for the different modular CD19 CARs.

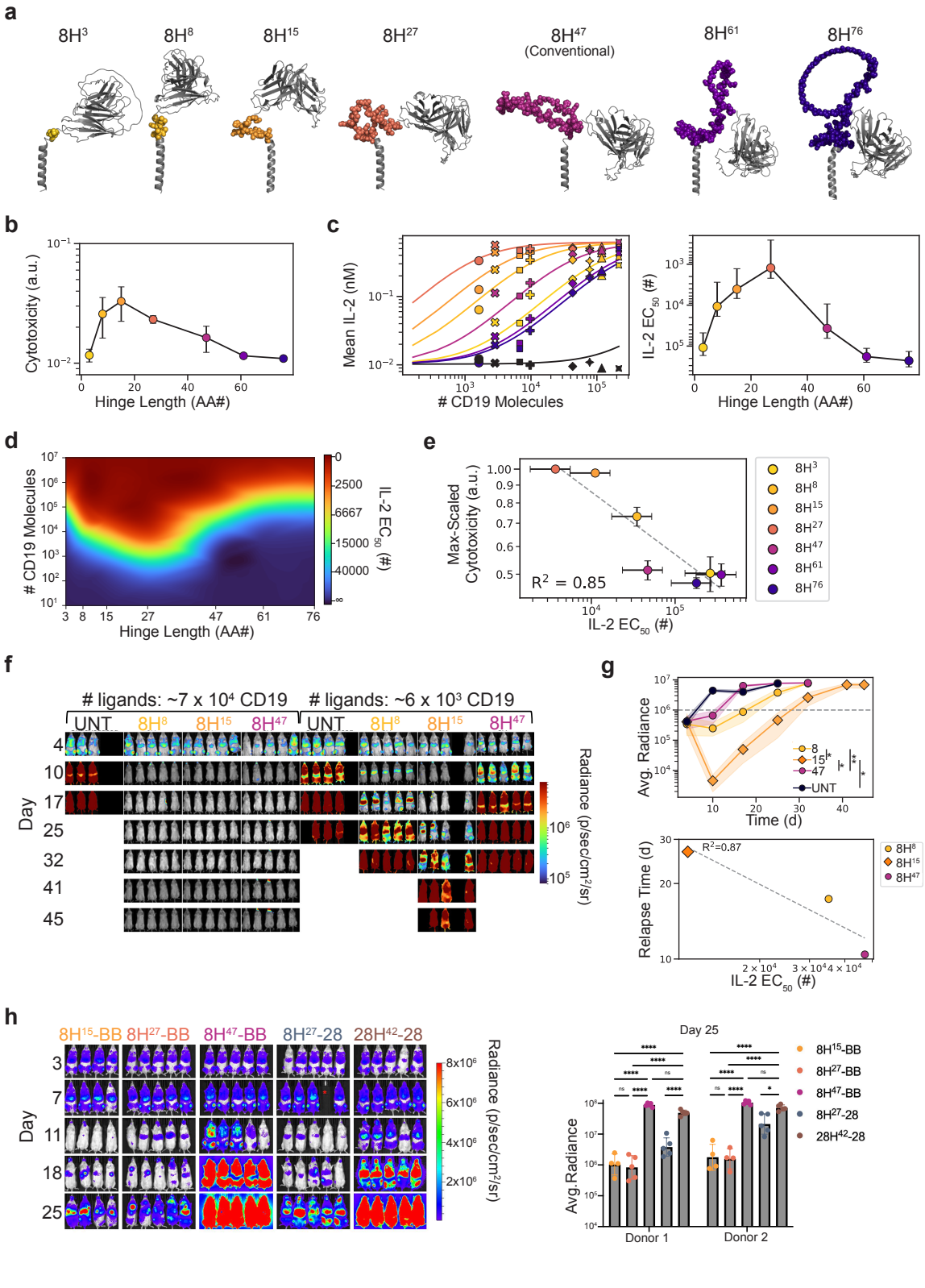

a
8H3
8H8
8H15
8H27
8H47
(Conventional)
8H61
8H76
b
Cytotoxicity (a.u.)
Hinge Length (AA#)
c
Mean IL-2 (nM)
CD19 Molecules
IL-2 EC50 (#)
Hinge Length (AA#)
d
CD19 Molecules
Hinge Length (AA#)
IL-2 EC50 (#)
e
Max-Scaled Cytotoxicity (a.u.)
R2 = 0.85
IL-2 EC50 (#)
f
ligands: ~7 x 10⁴ CD19
ligands: ~6 x 10³ CD19
UNT
8H8
8H15
8H47
Day
Radiance (p/sec/cm²/sr)
g
Avg. Radiance
Time (d)
UNT
Relapse Time (d)
R2=0.87
IL-2 EC50 (#)
h
8H15-BB
8H27-BB
8H47-BB
8H27-28
28H42-28
Day
Radiance (p/sec/cm²/sr)
Day 25
Avg.Radiance
Donor 1
Donor 2

Figure 8.2: Optimizing the CD8α hinge for increased responsiveness of tisa-cel-like and axi-cel-like CARs to CD19^{low} leukemia.

a, Alphafold2 (AF2)-predicted transmembrane and extracellular structures of CARs with different CD8α hinge lengths relative to the tisa-cel-like construct harboring a 47 amino acid (AA) CD8α hinge consisting of 45 AAs derived from CD8α and a SG linker (noted as 'conventional'). Constructs containing CD8α-derived hinges greater than 47 AA were generated by inserting a glycine-serine sequence between the SG linker and the CD8α-derived amino acids. b, Normalized cytotoxicity is plotted as a function of the number of amino acids in the CD8α CAR hinge (n=2-4 individual donors). c, IL-2 dose response curves, averaged by donor following a 24-hour co-culture of the indicated CD19 CAR T-cells, are presented as a function of CD19 antigen density (n=4 individual donors; left). Line plot of EC_{50} derived from the dose response curves for each CD19 CAR construct is presented as a function of the number of amino acids in the CD8α CAR hinge. Error bars represent the standard error of the means of EC_{50} from individual donors (right). d, Heatmap depicting the mean log-transformed level of IL-2 secretion by CD19 CAR T-cells following exposure to leukemia cells expressing different levels of CD19. The color bar on the right indicates the scale for max-normalized IL-2. The cubic method is used to extrapolate IL-2 secretion values for grid points within the observed data range (n=4 individual donors). e, Correlation between cytotoxicity and the EC_{50} for IL-2 secretion (n=2-4 individual donors). f, Bioluminescence images of NSG mice engrafted with GFP$^+$Luciferase$^+$ NALM6 (1 x 10^6) clones expressing either 7 x 10^4 (left) or 6 x 10^3 (right) CD19 molecules followed by adoptive transfer of the indicated untransduced (UNT) or CD19 CAR-transduced T-cells (day +3, n=4-5 mice per CAR T-cell group). Data are representative of two independent experiments using CAR T-cells from individual donors. g, Line plot of the average in vivo radiance of mice treated with the indicated UNT or CD19 CAR T-cells as a function of time. An average threshold radiance of 10^6 p/sec/cm^2/sr is shown as a dotted line (top). Correlation between the EC_{50} for IL-2 secretion and the time at which the average radiance intersects with the 10^6 p/sec/cm^2/sr threshold for each CAR construct (defined as 'relapse time', bottom). At day 10, differences between the indicated CAR constructs were assessed using two-sided Mann–Whitney U tests with Holm correction for multiple comparisons; asterisks denote adjusted significance. AA, amino acid. h, Bioluminescence images of NSG mice engrafted with GFP$^+$Luciferase$^+$ NALM6 clone expressing 6 x 10^3 CD19 molecules followed by adoptive transfer of the indicated CD19 CAR-transduced T-cells (day +4, n = 4-5 mice per CAR T-cell group). Data are representative of two independent T-cell donors (left). Bar plot of average in vivo tumor radiance for mice treated with the indicated CD19 CAR T-cells on day+21 following CAR T-cell infusion (one-way ANOVA with Tukey's multiple-comparisons test; *:$p < 0.05$, **:$p < 0.01$, ***:$p < 0.001$, ****:$p < 0.0001$; right).

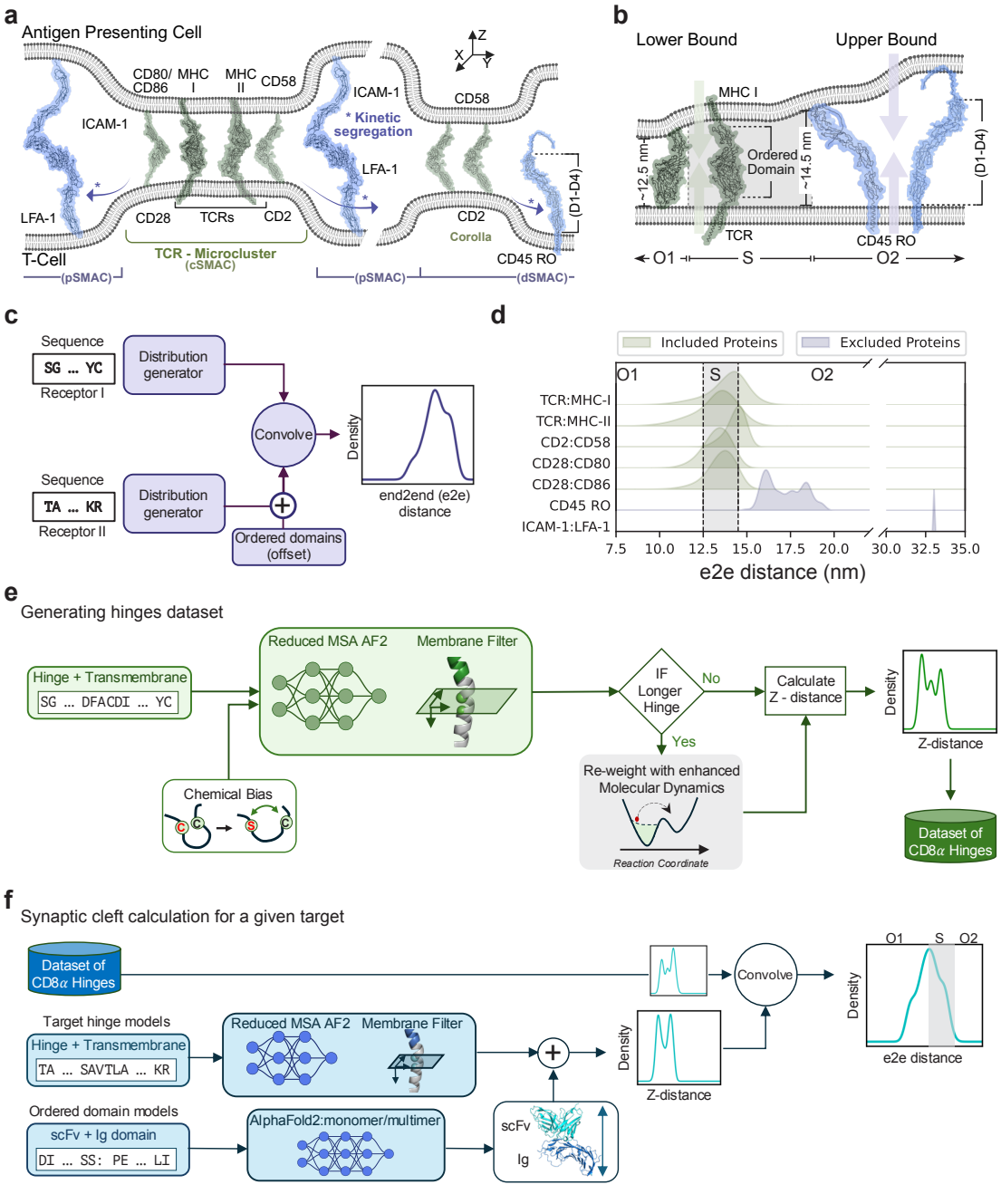
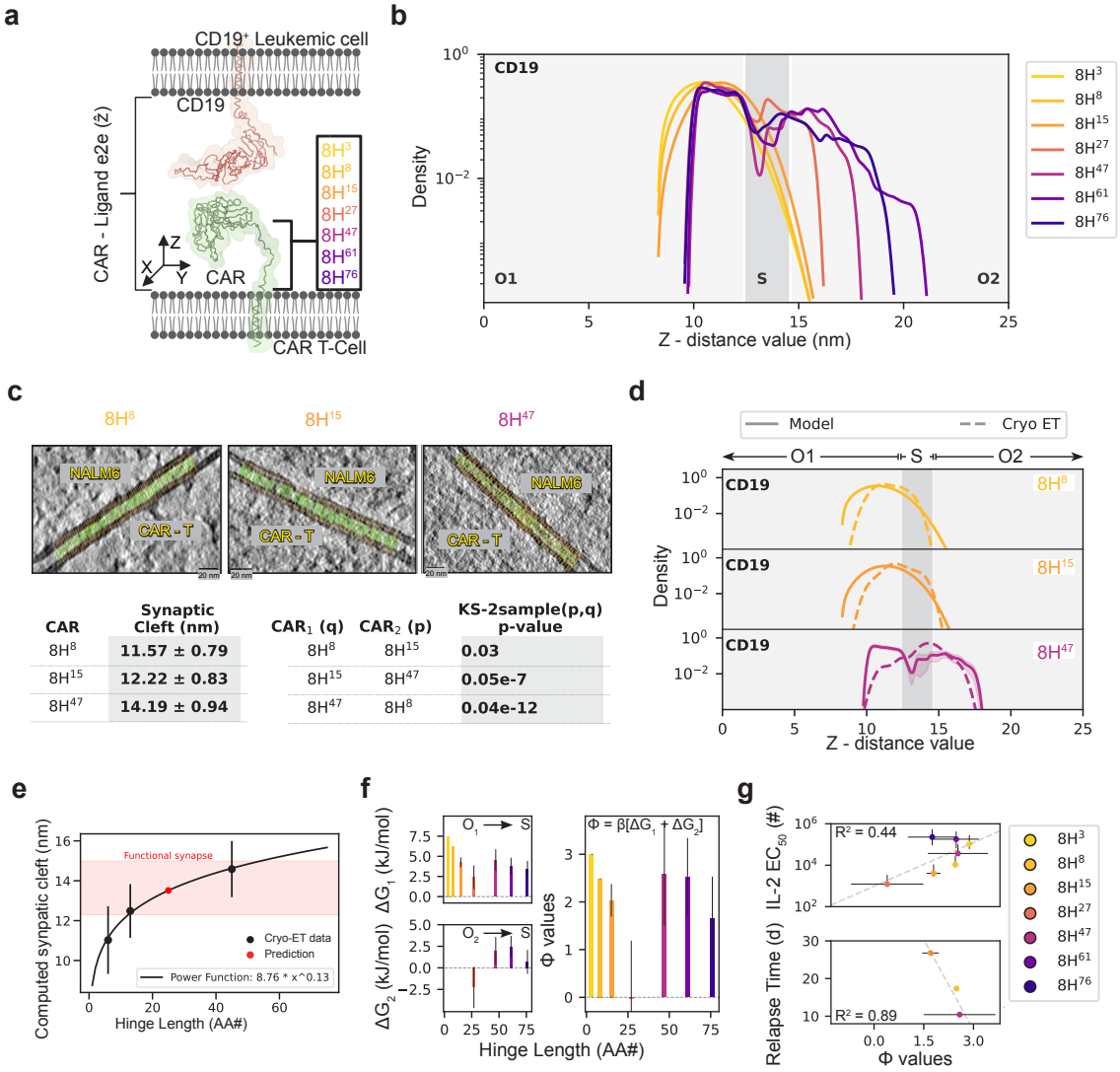

a
Antigen Presenting Cell
CD80/ CD86
MHC I
MHC II
CD58
ICAM-1
ICAM-1
* Kinetic segregation
CD58
LFA-1
LFA-1
CD28
TCRs
CD2
CD2
Corolla
(D1-D4)
T-Cell
TCR - Microcluster (cSMAC)
CD45 RO
(pSMAC)
(pSMAC)
(dSMAC)
b
Lower Bound
Upper Bound
MHC I
~12.5 nm
Ordered Domain
~14.5 nm
(D1-D4)
TCR
CD45 RO
O1
S
O2
c
Sequence
SG ... YC
Receptor I
Distribution generator
Convolve
Density
end2end (e2e) distance
Sequence
TA ... KR
Receptor II
Distribution generator
Ordered domains (offset)
d
Included Proteins
Excluded Proteins
O1
S
O2
TCR:MHC-I
TCR:MHC-II
CD2:CD58
CD28:CD80
CD28:CD86
CD45 RO
ICAM-1:LFA-1
7.5
10.0
12.5
15.0
17.5
20.0
30.0
32.5
35.0
e2e distance (nm)
e
Generating hinges dataset
Hinge + Transmembrane
SG ... DFACDI ... YC
Reduced MSA AF2
Membrane Filter
IF Longer Hinge
No
Yes
Calculate Z - distance
Density
Z-distance
Chemical Bias
Re-weight with enhanced Molecular Dynamics
Reaction Coordinate
Dataset of CD8α Hinges
f
Synaptic cleft calculation for a given target
Dataset of CD8α Hinges
Convolve
O1
S
O2
Density
e2e distance
Target hinge models
Hinge + Transmembrane
TA ... SAVTLA ... KR
Reduced MSA AF2
Membrane Filter
Density
Z-distance
Ordered domain models
scFv + Ig domain
DI ... SS: PE ... LI
AlphaFold2:monomer/multimer
scFv
Ig

Figure 8.3: Modeling synaptic clefts generated by CAR using an AlphaFold2RAVE(AF2RAVE)-based model

a, Schematic of the immunological synapse. Small protein pairs that are costimulatory, such as CD2-CD58 and CD28-CD80/86, are generally positioned in the central portion of the synapse-the central supramolecular activation cluster (cSMAC; TCR-microcluster)-potentiating signals initiated by the TCR complex or the CAR [239, 240]. ICAM-1/LFA-1 and CD45RA/CD45RO localize to the peripheral SMAC (pSMAC) and distal SMAC (dSMAC), respectively; CD2-CD58 complexes can also occupy the corolla, with cSMAC-like spacing. Structures were generated using AF-multimer with full MSA information and visualized using Biorender. b, A cartoonist hypothesis presenting an imprecise definition of the lower and upper bounds of a functional synapse (S). The upper bound was defined by taking the z_{e2e} for the ordered domains (D1-D4) of CD45RO protein (O2) obtained using rMSA AF2 (Supplementary Fig. H.3). To establish the lower bound (O1), we calculated the end-to-end distances of the ordered domains of the synapse protein complexes using an rMSA approach on AF-multimer (Supplementary Fig. H.3). c, A schematic representation of the CAR-BOOST model that assumes distributions arising from the contacting receptors in the synapse to be independent. d, CAR-BOOST modeling to T-cell synapse proteins-using rMSA AF2 as the distribution generator-are shown in the density plots. e, Schematic of the AF2RAVE-based pipeline to generate distributions of varied sequence lengths of the CD8α-derived CAR hinge and the transmembrane domain. f, Schematic depicting the rMSA AF2/ AF-multimer approach for generating multiple structures for a specific target protein to determine the location of its epitope in the synaptic cleft. The epitope's sequential information is gathered from sequence mapping. The pipeline illustrates how to estimate the optimal number of CD8α CAR hinge amino acids required for an effective CAR T-cell response against an antigen$^{\mathrm{low}}$ tumor.



CAR	Synaptic Cleft (nm)
8H8	11.57 ± 0.79
8H15	12.22 ± 0.83
8H47	14.19 ± 0.94

CAR1 (q)	CAR2 (p)	KS-2sample(p,q) p-value
8H8	8H15	0.03
8H15	8H47	0.05e-7
8H47	8H8	0.04e-12

Figure 8.4: Cryo-ET links CD8α hinge–dependent synaptic distance to CD19 CAR T-cell responsiveness

a, Cartoon representation of a synaptic cleft between CD19 on the leukemic cell and T cells harboring CAR constructs incorporating 3, 8, 15, 27, 47, 61, and 76 aa CD8α hinges. b, Density plots presenting modeled intermembrane distances between the CD19 epitope and the CAR constructs presented in Fig. 8.4a. c, Representative cryo-ET images of the intermembrane regions between CD19$^+$ NALM6 leukemia cells and T-cells expressing 8H8, 8H15, and 8H47 CD19 CAR constructs. Orange and green lines show 'Inner to Inner' and 'Outer to Outer' distances measured between the inner and outer plasma membranes of the CAR T-cell and leukemia cell, respectively (top). Calculated synaptic cleft distances $\pm$ standard deviation, as described in the Methods, are presented for each CAR and statistical differences between CAR constructs with different CD8α hinges are shown (bottom). d, Comparisons between the density plots of intermembrane distances calculated from the CAR-BOOST in silico model (solid lines) and from cryo-ET experiments (dashed lines). Note that the leftmost in silico peak formed by the 8H47 CAR is due to the membrane interactions in the simulation. e, Curve of the fitted power function depicting the mean synaptic cleft distances relative to the number of amino acids in the CD8α hinge of the different CD19 CARs. f, Bar chart displaying the ΔG values, calculated from the in silico model for each CAR construct. Note that constructs with hinge lengths of less than 27 AA are not included in the ΔG_2 due to a lack of support in the O2 region. To overcome the lack of support in different regions the sum of the two ΔG values (ϕ values) for all CAR constructs are shown in the right panel. g, ϕ values are compared against in vitro IL2-EC$_{50}$ (top) and relapse time of leukemia-bearing NSG mice following adoptive CAR T-cell transfer (bottom). The two outliers (8H61 and 8H76) in the top panel result from a lack of convergence in sampling due to the large length scale.

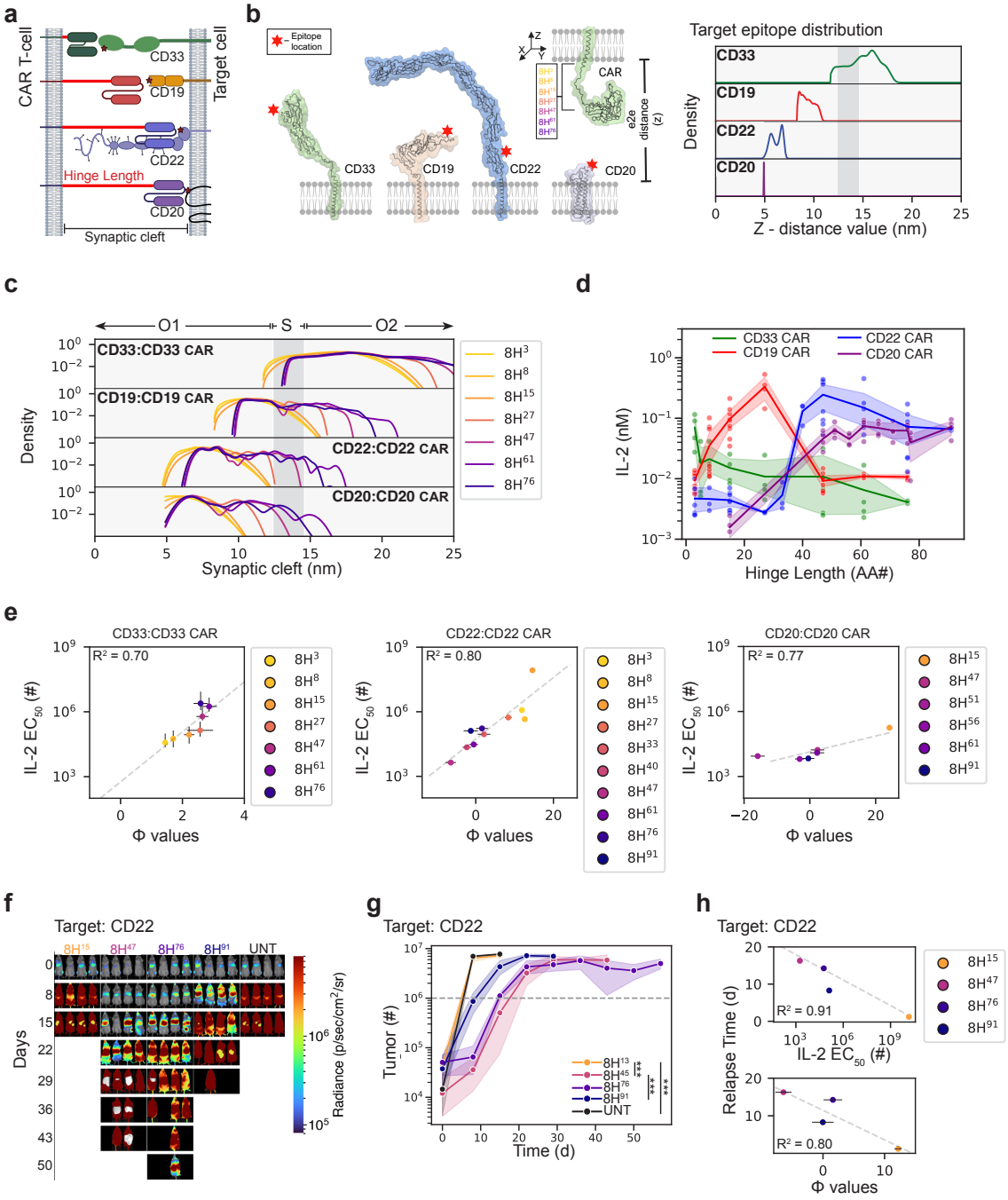

a
CAR T-cell
Target cell
CD33
CD19
CD22
Hinge Length
CD20
Synaptic cleft
b
Epitope location
CD33
CD19
CD22
CD20
CAR
e2e distance (z)
Target epitope distribution
Density
Z - distance value (nm)
c
O1
S
O2
CD33:CD33 CAR
CD19:CD19 CAR
CD22:CD22 CAR
CD20:CD20 CAR
Density
Synaptic cleft (nm)
d
CD33 CAR
CD19 CAR
CD22 CAR
CD20 CAR
IL-2 (nM)
Hinge Length (AA#)
e
CD33:CD33 CAR
R² = 0.70
CD22:CD22 CAR
R² = 0.80
CD20:CD20 CAR
R² = 0.77
IL-2 EC50 (#)
Φ values
f
Target: CD22
UNT
Days
Radiance (p/sec/cm²/sr)
g
Target: CD22
Tumor (#)
Time (d)
UNT
h
Target: CD22
Relapse Time (d)
R² = 0.91
R² = 0.80
IL-2 EC50 (#)
Φ values

Figure 8.5: The CAR-BOOST model predicts maximally responsive hinges across multiple CAR targets

a, Schema of hinge lengths for maximally responsive CAR constructs, as determined by the location of the target epitope. b, Schema illustrating the structural dimensions of the target proteins and CAR constructs with representative lengths and epitope locations (red stars) shown in the left panel. The right panel depicts the z-distance distribution of the different target epitopes from the cell surface, showcasing the variability in epitope location. c, Density plots illustrating the distribution of synaptic clefts for different CAR constructs across all target receptor epitopes. d, Line plots depicting the mean IL-2 secretion by CD33, CD19, CD22, and CD20 CAR T-cells relative to the predicted length of the CD8α hinge, as determined by the AF2 linear-regression method (n=2-4, 6-7, 2-7, and 4 donors for the different CD33, CD19, CD22, and CD20 CARs, respectively). e, Correlations between ϕ values and EC_{50}s for IL-2 secretion by CD33, CD22, and CD20 CAR T-cells following interaction with their respective target receptors. f, Bioluminescence images of NSG mice engrafted with $GFP^{+}Luciferase^{+}$ NALM6 (1×10^6) clone expressing $\sim 1 \times 10^3$ CD22 molecules/cell and adoptively transferred with CD22 CAR-transduced or untransduced (UNT) T-cells (day +3, n=4 mice per group, representative of data from 2 individual donors). g, Line plot of the average in vivo radiance of mice treated with the indicated UNT or CD22 CAR T-cells as a function of time. An average threshold radiance of 10^6 p/sec/cm^2/sr is shown as a dotted line, defined as 'relapse time.' h, Correlation between the EC_{50}s for IL-2 secretion and the time at which the average radiance intersects with the 10^6 p/sec/cm^2/sr threshold for each CAR construct (top); correlation between relapse times and ϕ values (bottom).

Chapter 9: Applied causality to infer protein dynamics and kinetics

In this chapter, I aim to develop a method that infers timescale differences present in AI-generated structural ensembles. As seen in Chapter 2, and to some extent in the application chapters, such ensembles lack the causality mechanism provided by an MD integrator, meaning there is no inherent timescale attached to the generated ensemble.

A protein's function is not determined by sequence or a single static structure alone, but by the conformational diversity encoded in its free-energy landscape [16, 17, 40, 42, 262, 263]. As discussed in Chapter 1, obtaining this conformational diversity and its associated timescales remains difficult. Experiments can probe important dynamical regimes but may lack full atomistic detail, while molecular dynamics (MD) simulations provide mechanistic trajectories but are limited by accessible simulation timescales [264, 265, 266, 267, 268, 269, 270, 271]. Enhanced sampling methods can improve conformational exploration, but they are system-dependent and require careful choices of collective variables or biasing strategies, as discussed earlier in this thesis [65].

Recent AI-based structure and ensemble-generation methods, including AlphaFold2 (AF2), RoseTTAFold, reduced-MSA AF2, and newer generative models, can rapidly produce structurally diverse protein ensembles [1, 10, 91, 95, 98, 244, 272]. These methods are closely related to the structural ensemble generation strategies discussed in Chapter 2. However, because these

models do not propagate structures through a physical integrator, their outputs do not have an intrinsic ordering in time. Therefore, even when an AI-generated ensemble contains realistic structural diversity, it remains unclear whether the differences between conformations correspond to fast local fluctuations, slower collective motions, or rare functional transitions.

To address this limitation, I use AI-generated conformational ensembles to parameterize a memory-free, overdamped Langevin equation through the Langevin Equation for Protein Dynamics (LE4PD) framework. The central idea is to construct a harmonic potential of mean force (PMF) matrix from the ensemble and diagonalize it to obtain uncoupled Langevin modes. Each mode is associated with a characteristic length scale and relaxation timescale, which can then be used to estimate residue-level autocorrelation functions (ACFs) [262, 273, 274, 275]. In this way, a static AI-generated ensemble can be converted into an approximate dynamical model with interpretable timescales.

This approach is inspired by the framework of Copperman et al., where a memory-free Langevin equation was mapped to a Gaussian network model using a PMF extracted from NMR conformations [262]. It is also conceptually related to diffusion-map and other data-driven dimensionality-reduction approaches for biomolecular dynamics [192, 276, 277, 278, 279, 280, 281, 282, 283]. The present approach differs in two ways: first, the input ensembles are generated from AI-based structural ensemble methods; second, the method uses the residue-fluctuation basis set from Ref. [274], which incorporates anisotropy and connects to principal component analysis (PCA) in appropriate limits. I test this methodology using HIV-1 protease monomers and dimers, including both wild-type (WT) and mutant (MUT) sequences.

First, I use reduced-MSA AF2 as the conformational generator to demonstrate the pipeline and to relate MSA depth to the timescales represented by the generated ensemble. I generate one

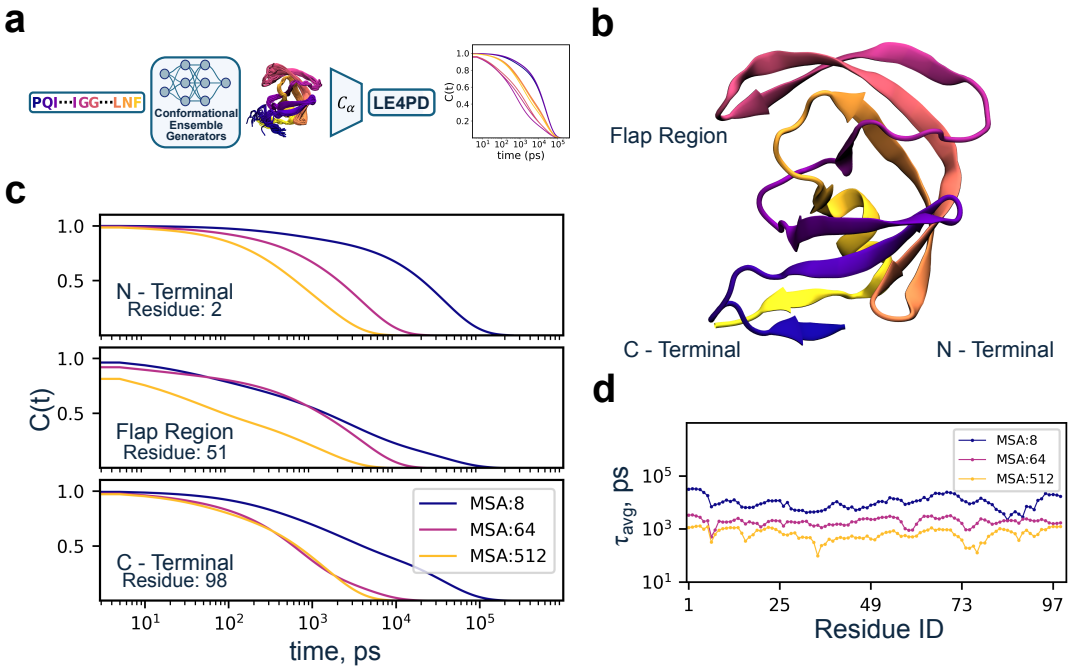


Figure 9.1: Overview of our method. a) The protein sequence is input to the machine learning architecture of choice, generating an ensemble of all-atom configurations. This ensemble is coarse-grained at the C_α level and input to the Langevin Equation for Protein Dynamics (LE4PD), which predicts the relaxation timescales (kinetics) of each residues' structural fluctuations (dynamics). b) A representative structure of the HIV-1 protease monomer is shown, with the flap (residues 49-51) and N- and C-termini labeled. c) Residue-dependent autocorrelation functions as a function of MSA depth input to AF2 for the three labeled structural regions in (b). d) Predicted relaxation time of each residue as a function of residue index for one HIV-1 protease sequence, PDB ID: 1EBW.

microsecond of atomistic MD simulation as a reference and compare the rMSA-AF2 ensembles to MD trajectories using a defined distance metric. This allows me to approximate the dynamical regime spanned by a given AI-generated ensemble. The main result is that AF2 structures generated with an MSA depth of 8 map to relaxation timescales between those observed in 100 ns and 1000 ns MD simulations, whereas AF2 ensembles generated using the default MSA cluster number of 512, or "full" MSA depth, are comparable to timescales observed in a 10 ns simulation. I further test the generality of this protocol by comparing against other contemporary ML models and by applying the workflow to both WT and MUT HIV-1 protease dimers using a cofolding model.

9.1 Results

9.1.1 Overview: A Map from Sequence to Kinetics

An overview of my method of mapping from sequence space to a coupled dynamics-kinetics space provided by the solution to the ML-parameterized Langevin Equation for Protein Dynamics (LE4PD) equation is given in Figure 9.1a. Specifically, I apply this method on a set of six unique sequences encoding the HIV-1 protease monomer (structure shown in Figure 9.1b). The solution of the LE4PD gives autocorrelation functions (kinetics) for each residue and predicted root-mean-square fluctuations (RMSF) (dynamics). Depending on the hyperparameterization of the generative ML method, e.g. MSA depth to AF2, the kinetics of each generative ML ensemble is predicted using the LE4PD, as illustrated for three HIV-1 proteases residues (Figure 9.1c). I estimate the average decorrelation time τ_{avg} by integrating the predicted autocorrelation functions (ACFs) for each residue in the protein (Figure 9.1d).

9.1.2 The Timescales of AlphaFold2-generated Ensembles are Inversely Correlated with MSA Depth

As the MSA depth in AF2 serves as a constraint on the predicted structures, as I decrease the MSA depth, I expect the ensemble's conformational variance to increase. Furthermore, since there is a general correlation between the variance of a protein's dynamics and the timescale of those dynamics,[274] I expect that lifting the constraints via MSA depth reduction will increase the predicted residue relaxation timescales.

To measure this effect, in Figure 9.2 I plot the average decorrelation time, τ_{avg}, for each

residue, which is defined simply as the integral of the residue's ACF, indexed by i:

$$\tau_{\text{avg}}(i) = \int_0^\infty \langle \Delta\vec{R}_i(0) \cdot \Delta\vec{R}_i(t) \rangle dt. \tag{9.1}$$

In Eq. 9.1, $\langle \Delta\vec{R}_i(0) \cdot \Delta\vec{R}_i(t) \rangle$ is the time correlation function for the i^{th} C_α atom. The calculation of the time correlation function is in the Materials and Methods section.

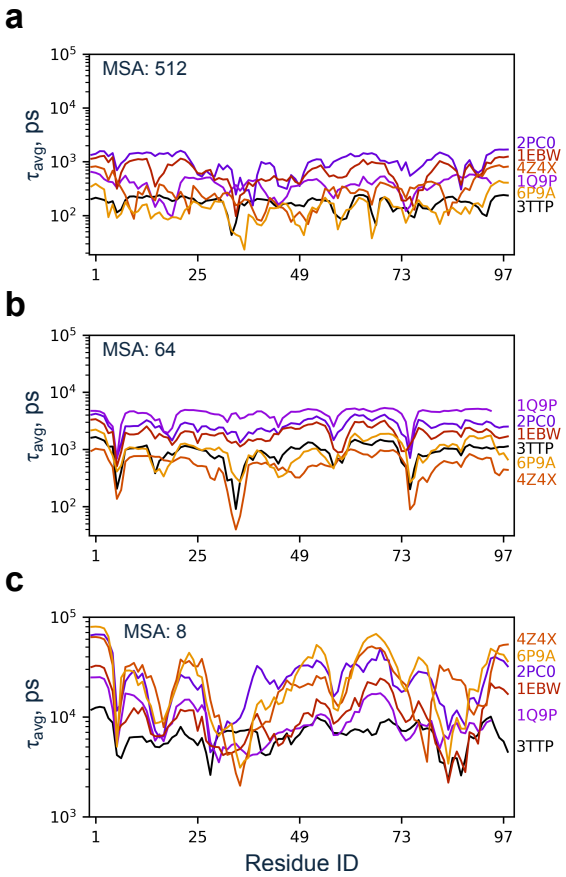


Figure 9.2: Estimated timescales for the relaxation kinetics of each residue for the six indicated sequences of HIV-1 protease at MSA depths of a) 512, b) 64, and c) 8. Each sequence variant label on the right-hand side of the plot is printed with the same color as its integrated timescale curve.

Figure 9.2 plots the timescales for each sequence at three MSA depths: 512 (Figure 9.2a), 64 (Figure 9.2b), and 8 (Figure 9.2c). When the MSA depth is at 512 and 64 (Figure 9.2a and

b), there is a clustering of either WT (1Q9P) or close-to-WT (five or fewer mutations from WT: 1EBW, 2PC0) at relaxation timescales slower than the highly mutated sequences (3TTP, 4Z4X, 6P9A), with 64 MSA depth showing a stronger separation. I observe a timescale separation between the near-WT cluster and the highly mutated cluster at certain stretches of the primary sequence, e.g., near the flap region (residues 49 through 51) at this highest MSA depth.

However, at 8 MSA depth, there is a strong dependence of sequence composition on both the average relaxation timescale and the τ_{avg} of each residue individually. However, there is no correlation between a sequence's Hamming distance from the WT and the relaxation timescales when an MSA depth of 8 is used to parameterize the LE4PD. Furthermore, there is no clustering seen in the equivalent relaxation timescale from 10, 100, and 1000 ns MD results, which are shown in the SI. The SI also contains a more detailed analysis of per-residue timescales for all six systems. From these results, the only definitive claim I can make is that, as the MSA depth is decreased, the timescales probed by AF2 show an increase, on average, in agreement with the equivalent MD results as the length of the MD simulation is increased. Unfortunately, due to the monomers' propensity to either dimerize spontaneously or form aggregates[284, 285], there is no NMR relaxation data for comparing the mutant to wild-type relaxation timescales.

However, since I see clustering at two higher MSA values studied (64 and 512) but not the lowest (8), I speculate this effect might be due to a tradeoff between the specific co-evolutionary signals or mutational information present in the MSA and the weights present in the AF2 architecture that allow it to fold proteins generally across a diverse sequence range.

9.1.3 MSA Depth is Inversely Correlated with the Span of the Slowest Collective Motions Generated by the AF2 Conformational Ensemble

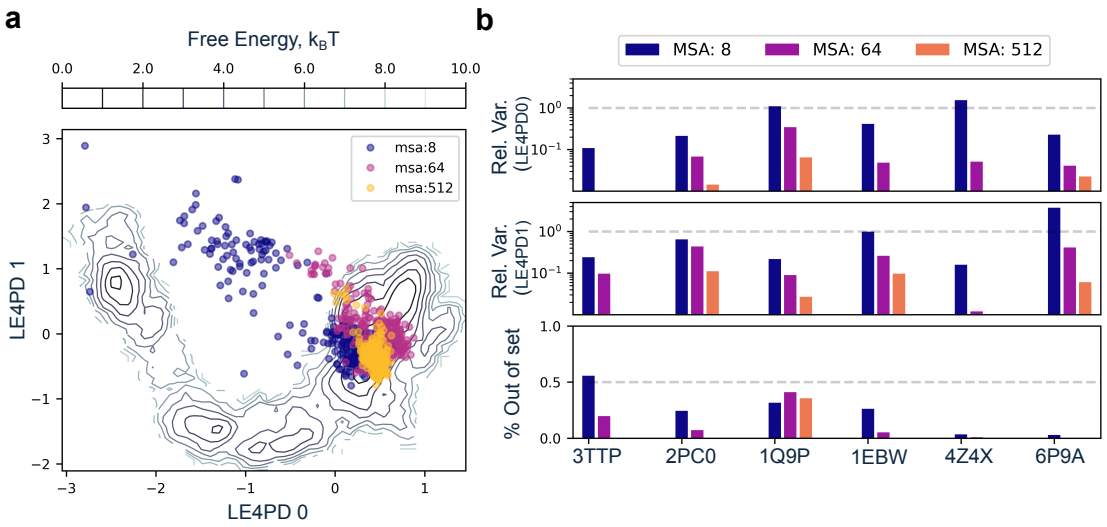


Figure 9.3: Measuring the volume of the space of the slow dynamics spanned by the AlphaFold2 conformational ensembles. a) AlphaFold2-generated ensemble projected onto the free-energy surface spanned by the two slowest LE4PD modes parameterized by a one microsecond MD simulation of the HIV-1 protease sequence encoded by PDB ID: 1EBW. b) Comparing the relative variance in the space spanned by the two slowest LE4PD modes of the AlphaFold2-generated ensembles compared to the one-microsecond MD simulation. I also show in the bottom plot what fraction of structures generated by AlphaFold2 at the indicated MSA depth are outside the support of the one-microsecond dynamics.

The results discussed in the previous section show a general correlation of the relaxation time of collective motions and their variance. So, I expect that, as the MSA depth is decreased, and the amount of volume spanned by the AF2-generated ensemble increases, the timescales of the dynamics generally increase (Figure 9.2) as well. To investigate this claim, Figure 9.3a overlays the structures predicted at MSA depths of 8, 64, and 512 on the free-energy surface of the two slowest LE4PD modes calculated from a microsecond-long MD simulation starting from the structure encoded by PDB ID: 1EBW (simulation details are in the Materials and Methods section). Visually, as the MSA increases from 8 to 512, the predicted structures retreat into the global minimum on the free-energy surface, and the overall variance of the ensemble decreases

(Figure 9.3b). While the global minimum of both the AF2 and MD free-energy roughly coincide for 1EBW, this trend is not general (the SI shows equivalent scatter plots for the other sequences), and will almost certainly not hold for any given protein. Discrepancies between the modes of the ML- and MD-generated distributions are a function of the given sequence, ML architecture, and MD setup (force field, water model, etc.).

The bottom panel of Figure 9.3b shows the fraction of out-of-set or outlying structures for each sequence at three MSA values. I define an outlying structure relative to the projection into the two slowest LE4PD modes calculated from the one microsecond simulation specific to each sequence. The outlying fraction of structures is a function of both sequence and MSA depth.

Generally, as the MSA is increased, the fraction of out-of-set structures decreases, which follows the argument that the MSA serves as a constraint keeping the predicted structures close to the training structures. The exception is 1Q9P, where the fraction of out-of-set structures is approximately constant as a function of MSA since, for this particular sequence, the predicted structures do not collapse to a minimum of the free-energy but rather near the node of the slowest LE4PD mode. This discrepancy could be due to the truncation in the C-terminus of 1Q9P. Since the truncated sequence is dominated by full-length sequences in AF2's training set, it might give the wrong mode for the predicted 1Q9P ensemble compared to MD, which simulates the truncated sequence explicitly.

Generally, the fraction of outlying structures should be a function of sequence, force field, simulation length, and subspace projection. For example, the sequences encoded in both 3TTP and 6P9A are more than 20 mutations from WT, but there are a far larger number of out-of-set structures in the generated ensemble using an MSA depth of 8 for 3TTP compared to 6P9A. In contrast, 2PC0 and 1EBW are both 5 mutations from WT, and they show a similar relative

variance profiles and fraction of out-of-set structures. There does not appear to be a correlation between either the relative variance of the generated ensemble or the fraction of out-of-set structures as a function of Hamming distance from the WT sequence, which implies AF2 is not memorizing the dynamics of WT HIV-1 protease when inferring ensembles.

9.1.4 Quantification of the Timescales Spanned by AF2 Evaluated at Different MSA Depths

From Figure 9.3a, it is clear that, at least for two sequences (e.g., 1Q9P and 4Z4X), the ensembles predicted from an MSA depth of 8 in AF2 structure generation yields an ensemble spanning the space of at least the slowest LE4PD mode, when measured using the fractional variance of that mode. However, assuming that the dynamics and kinetics of the entire protein will match just because the variance is approximately the same between projected AF2 and MD ensembles is questionable. So, I quantify how closely the residue-residue ACFs match between the AF2 and MD ensembles to determine rigorously the timescales of the ML ensembles' dynamics.

Figure 9.4a plots the percentage of each protein's dynamics best explained by the AF2 ensemble generated at the indicated MSA depth. I map the ACFs using the DTW metric defined in the Methods section. Nearly equivalent results are found using the KS instead; the KS results are displayed in the SI. Figure 9.4a clearly shows the lowest MSA studied (8) is, on average, the best at describing the decorrelation timescales of the observed dynamics in the 1000 ns MD simulations. Conversely, using the full MSA depth of 512 allows for the best modeling of the decorrelation dynamics observed in the 10-ns-long MD simulations, on average. For the intermediate timescales in the 100 ns simulations, there is a broad range of well-performing MSAs

from 16 through 128. Roughly, MSA depths between 8 and 16 model well the decorrelation timescales observed in 1000 ns MD simulations; MSA depths between 256 and 512 model best the decorrelation dynamics in 10 ns MD simulations. The in-between timescales in the 100-ns-long simulations are modeled well by MSA depths between 16 and 128.

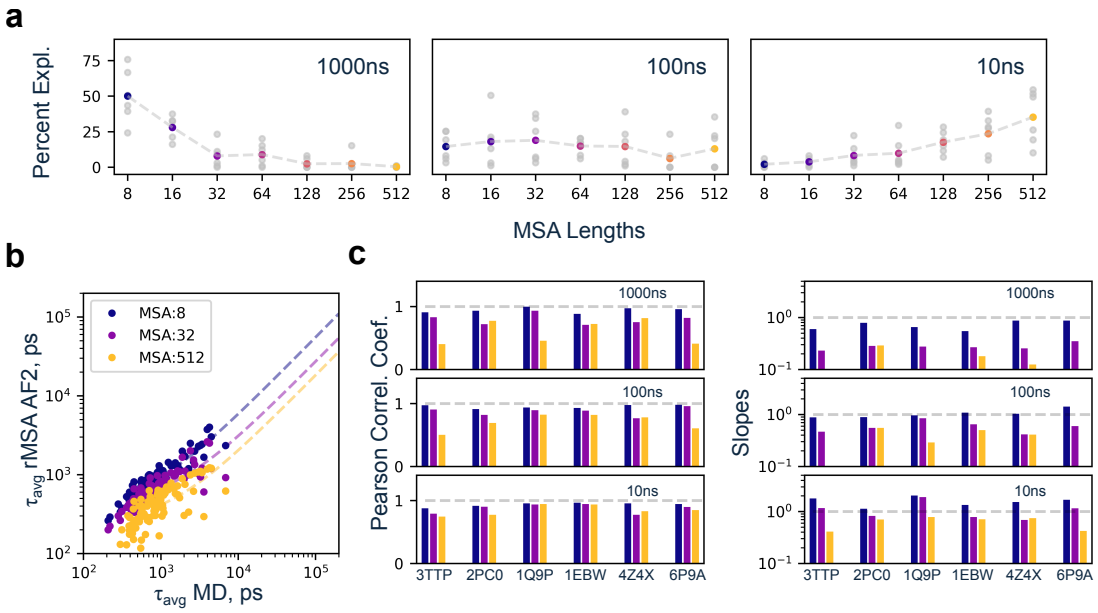


Figure 9.4: Reporting the timescales measured by AlphaFold2 conformational ensembles generated using different MSA depths. a) Percent of each residue's ACF best measured by each of the reported MSAs at three different timescales : 10, 100, and 1000 ns. Colored circles denote the average across sequences at each MSA while grey circle denote the result for each sequence individually. b) Scatter plot and linear fits (colored, dashed lines) of the average integrated correlation time from the reference MD simulations (τ_{avg} MD sim.) and the LE4PD theory parameterized using the AlphaFold2 conformational ensembles at the given MSA depth (τ_{avg} rMSA AF2). Reported are the correlation times from MD and AlphaFold2 for the HIV-1 protease sequence encoded in PDB ID: 1EBW at the 1000 ns timescale. c) Pearson correlation coefficient between the timescales reported in b), but for all six HIV-1 protease sequences studied here and for all three timescales (10, 100, and 1000 ns). Also reported in the second column are the slopes of the linear regression of τ_{avg} MD sim. onto τ_{avg} rMSA AF2 for all six sequences (1Q9P, 2PC0, 3TTP, 1EBW, 4Z4X, 6P9A) and all three timescales (10, 100, and 1000 ns).

As a second approach to quantify the predicted relaxation timescales between the coupled AF2-LE4PD approach and the benchmark MD simulations, Figure 9.4b plots the mode-averaged residue decorrelation timescales defined in Eq. 9.1 from the MD ACFs and the AF2-parameterized LE4PD's theoretical ACFs for the sequence in PDB ID: 1EBW. The scatter of

points for each MSA is fit to a linear model. Figure 9.4c reports the Pearson correlation coefficient and slopes of the linear models fit to each MSA for my six variants.

Figure 9.4c supports the results in Figure 9.4a: the Pearson correlation coefficient for an MSA depth of 8 is closest to unity at the 1000 ns timescale, and it has the slope closest to unity in the linear fit at 1000 ns. Analyzing the results for the 10-ns simulation timescale shows using an MSA depth of 512 gives a linear fit with slope closest to unity. In the limit that the Pearson correlation coefficient and slope approach unity, the predicted decay timescales (kinetics) are identical, less an additive constant, which is why I use both measurables as goodness-of-fit metrics.

9.1.5 Effect of Out-of-Set Sample Removal on the AF2-predicted Dynamics and Kinetics

Generative ML models are prone to infer samples outside the support of known data distributions. For networks attempting to model physics, these output samples generally deviate from known physical reality by violating e.g. conservation of mass, energy, or both. For generative ML models of protein structures, such out-of-set samples[286, 287] may include, but are not limited to including, structures with unphysically large bond lengths, implausible torsion angles, or steric clashes.

Because I assign such conformations null Boltzmann weight, it should be efficacious to remove these generated structures from downstream analyses. In theory, excising structures having infinite free-energy from the ensemble due to the above-mentioned modeling issues should yield a distribution closer to the target distribution, whether it's from simulation or experiment. So,

pruning out-of-set structures should allow for a refined ML ensemble that aligns more closely with a target distribution. In practice, this amounts to a sharply defined maximum entropy approach[264].

My target is the free-energy surface spanned by the two slowest LE4PD modes parameterized using the one-microsecond MD simulations (Figure 9.3). I remove the effect of the outlying structures by projecting the AF2-generated ensemble onto this free-energy surface, then remove from the generated ensemble any structures lying outside the space explored in the simulation. That is, I remove structures with infinitely high free energy i.e. zero Boltzmann weight. I parameterize the LE4PD using the modified AF2 ensemble and perform the calculations as done for the original ensemble. Finally, I examine how removing the out-of-set structures modifies the predicted RMSF profile, decorrelation times, and behavior of the ACFs.

Figure 9.5 shows the results after removing outliers. Figures 9.5a and 9.5b focus exclusively on the sequence encoded by 1EBW; equivalent plots for the other five sequences are given in SI Figures S9, S10, and S11. In Figure 9.5a, the RMSF from the AF2-generated ensemble with an MSA depth of 8 is compared to the reference one-microsecond MD simulation when out-of-set structures are either retained (dashed, blue line) or removed (dashed, red line). For this sequence, there is an almost universal decrease in the predicted RMSF across the entire primary sequence. In particular, there is a notable decrease in the RMSF near residues centered at index 30, with the removal of out-of-set structures pulling the RMSF in that region much closer to the microsecond MD. The same is true around the N-terminus.

Figure 9.5b shows the correlation between the integrated timescales with out-of-set samples included (ordinate) or excluded (abscissa) for my six HIV-1 protease sequences, each demarcated by color. The sequences 2PC0 and 1EBW experience a significant decrease in timescale when

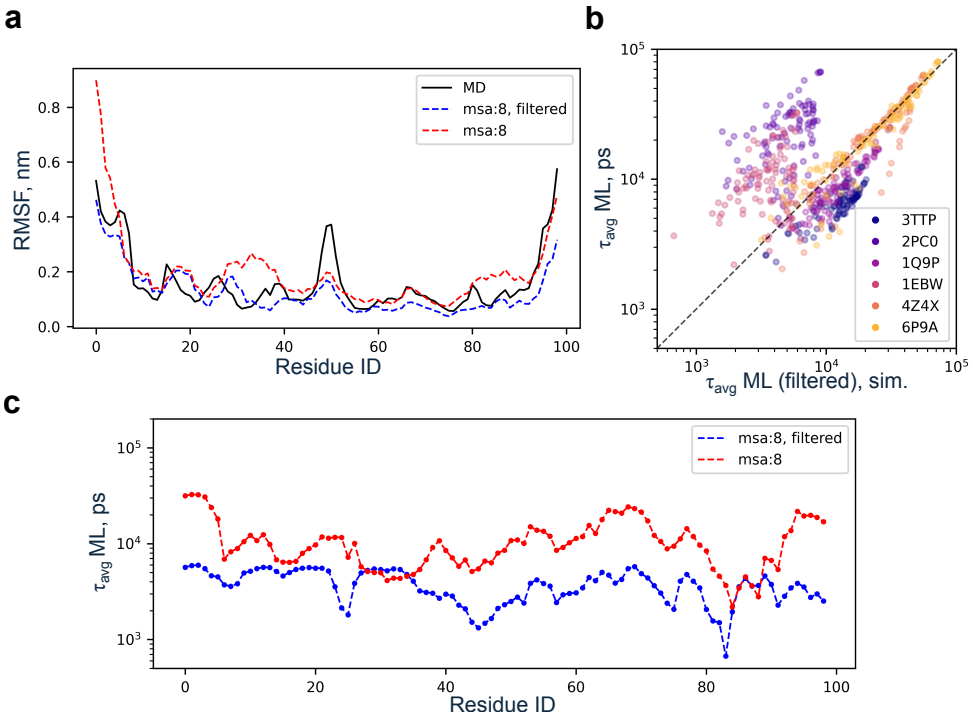


Figure 9.5: Effect of out-of-set structure removal on the predicted dynamics. All comparisons to MD are performed at the 1000 ns timescale. a) Root-mean-square fluctuations along the alpha-carbons predicted from the one microsecond MD simulation of HIV-1 protease with the sequence encoded by PDB ID: 1EBW (black) compared to the AlphaFold2 ensemble generated using an MSA depth of 8 and starting from the same sequence with (red, solid) and without (red, dashed) out-of-set structures in the ensemble input to the LE4PD theory. b) Correlation between the integrated correlation timescale predicted with the entire AlphaFold2 ensemble (τ_{ML}) and the ensemble with out-of-set structures removed (τ_{avg} ML (filtered) for all six sequences. The points are colored by the sequence given in the subplot's legend. c) Integrated timescales for 1EBW τ_{avg} from the LE4PD theory parameterized using the AF2 ensemble generated using an MSA depth of 8 with (blue) and without (red) out-of-set structures included in the ensemble. Including the data points from all six sequences, the Pearson correlation coefficient between the timescales calculated from the ensembles with and without out-of-set structures is 0.912.

out-of-set structures are removed while 4Z4X and 6P9A show little change in timescales; 3TTP and 1Q9P give, on average, slower timescales. Figure 9.5c shows the integrated timescale as a function of primary sequence for 1EBW, which shows clearly how its timescales decrease nearly uniformly as the outlying structures are removed.

Peering at Figures S9, S10, and S11 in the SI explains these disparate trends. Removing of out-of-set structures has a strong sequence dependence on the dynamics, i.e. RMSF, and kinetics in the form of τ_{avg}. The six sequences can be divided into three classes: the first consists of 1EBW

and 2PC0, where removal of outliers depresses the RMSF and timescales. The second is 4Z4X and 6P9A, where removal of outliers has a nearly negligible effect on the dynamics and kinetics. Finally, in the third is 1Q9P and 3TTP, where removing outliers actually increases the timescales nearly uniformly while the RMSF is not significantly altered.

For this last class, the mean of the AF2-learned ensembles lies outside a mode of the simulation-parameterized LE4PD surface. Removing outliers eliminates structures near this mean, increasing the overall variance of the now-outlier-free distribution. Thus, where the predicted mean of the ML-learned ensemble does not overlap with an MD-predicted mode, there is a possibility that removing the outliers in the projected space will increase the variance, as well as, potentially, the RMSF and integrated timescales. When the ML-generated ensemble collapses to a mode of the projected MD-generated distribution, there will be both a reduction in RMSF and timescales because removing the outlying structures reduces the variance and hence the projected volume spanned by the ML-generated ensemble. When the distributions overlap, there will be few outliers, and their removal will have little effect on the results, as seen in the sequences encoded by 4Z4X and 6P9A. The specific sequence effects of outlier removal will depend on the reference free-energy surface.

9.1.6 Comparison of AF2 Results to Other Generative Machine-learned Models

Due to the plethora of methods currently available to generate conformational ensembles of proteins using ML techniques, I compare the AF2 results to three other contemporary state-of-the-art methods: the distributional graphformer (DiG),[272] AF-Cluster (AFc),[95] and the

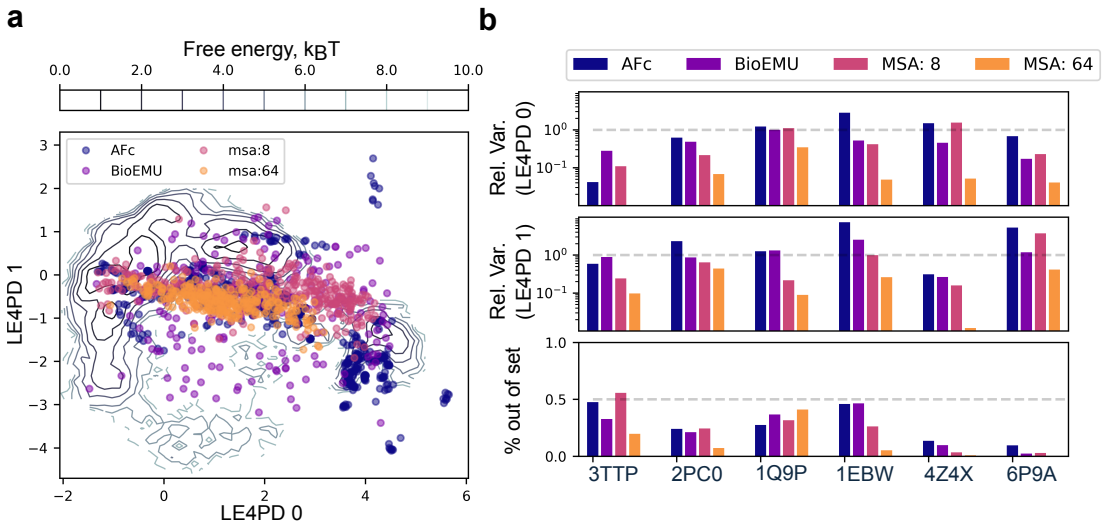


Figure 9.6: Measuring the volume of the space of the slow dynamics spanned by the AF2 conformational ensembles and other competing generative ML methods. a) Free-energy surface spanned by the two slowest LE4PD modes calculated from a one microsecond MD simulation of HIV-1 protease encoded by the sequence PDB ID: 1EBW. Projected onto the surface are the samples from AFc, BioEmu, and AF2 at different MSA depth. b) The relative variance of the two slowest LE4PD modes spanned by ensembles calculated using the same set of methods as in (a) for the six HIV-1 protease sequences. Also plotted in the bottom row is the fraction of out-of-set structures generated using each method. DiG results are omitted because they span a nearly negligible volume in this space, so the relative variance is not visible on the scale provided and none of the structures are outside the support of the one microsecond simulation.

"biomolecular emulator" (BioEmu).[244]

I choose the hyperparameters of the generative models to output 320 structures, the same size as the ensemble generated from AF2, although the exact size of the output ensemble depends on the ML architecture in use. For example, BioEmu automatically removes from the generated ensemble structures possessing steric clashes or bond length violations. For AFc, more than 320 structures may be output due to its clustering protocol. I solve the LE4PD equation separately for each sequence, using the ensemble generated by each method to predict dynamics and kinetics.

I can benchmark the predicted ML ensembles against the reference MD simulations. Figure 9.6a is analogous to Figure 9.3a except the conformational ensembles from the ML methods in addition to AF2 are projected onto the free-energy surface spanned by the two slowest LE4PD

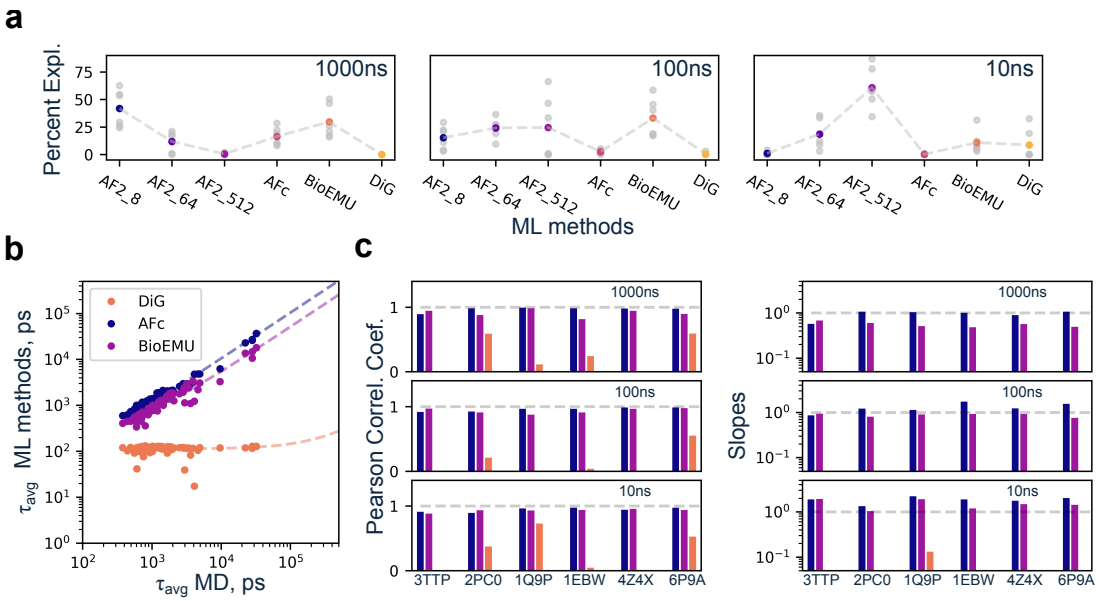


Figure 9.7: Reporting the timescales measured by alternative generative ML methods compared to AlphaFold2. a) Percent of residue ACFs best explained by generative ML methods for the HIV-1 protease sequence encoded by PDB ID: 1Q9P at the 10, 100, and 1000 ns timescales. Colored circles denote the average across sequences at each MSA while grey circle denote the result for each sequence individually. b) Correlation between the timescales(τ_{avg}) calculated from the ACFs from the one microsecond MD simulation and those calculated using the indicated ML ensembles input to the LE4PD theory. Dashed lines shown the best fit linear model to the data. c) Pearson correlation coefficients between the MD and ML integrated timescales for the three methods presented in (b) at the 10, 100, and 1000 ns timescales for the six HIV-1 protease sequences investigated here. Shown in the second column is the fit to the linear regression shown of τ_{avg} MD sim. onto τ_{avg} ML. The units for both τ_{avg} MD sim. and τ_{avg} ML are ps.

modes. The span of each method is quantified, as before, by comparing the relative variance along the LE4PD modes individually in Figure 9.6b. I also display the fraction of structures that are located outside the span of the one microsecond MD ensemble in the bottom row of Figure 9.6b.

Figure 9.6a shows that the competing ML methods BioEmu and AFc span much larger volumes of the slow dynamical subspace compared to AF2. This result is quantified in Figure 9.6b. However, as the last row of Figure 9.6b indicates, this large conformational volume can come at the cost of generating a large number of out-of-set structures. Strikingly, as shown in the SI, even though BioEmu can generate a relatively large number of structures that I quantify as

out-of-set, removing these outlying structures has relatively little effect on either the dynamics or the τ_{avg} values. This result implies that BioEmu learns well the relevant structures in the folded ensemble of HIV-1 protease.

On the other hand, removing out-of-set structures from the AFc ensemble has a more noticeable effect on the predicted dynamics and kinetics. For all sequences, the predicted RMSF values remain greatly elevated nearly universally across the entire primary sequence for each variant even when outlying structures are removed. I conclude AFc is doing a poorer job compared to AF2 (at low MSA depths) and BioEmu at learning the folded free-energy landscape of the HIV-1 protease variants. The results for DiG are not shown because I find that the DiG structures span a negligible space around the crystal structure for all sequences, encoding essentially no functional dynamics.

Figure 9.7 quantitatively evaluates the ML methods' performance. BioEmu and AFc describe the relaxation behavior of the protein's residues at the 1000 ns simulation timescale at a level that is similarly efficacious to AF2 with an MSA depth of 8 (Figure 9.7a). For modeling the dynamics in the 100 ns MD simulation, BioEmu is the most efficacious. At the 10 ns simulation timescale, using AF2 with an MSA depth of 512 is still the best model.

Figure 9.7b shows the correlation between the values of τ_{avg} of each residue in 1Q9P calculated from either the MD simulation or predicted from one of indicated ML methods. Again, I quantify the degree of correlation between the two sets of timescales using the Pearson correlation coefficient and the slope of the linear fit to the points in Figure 9.7c. The results quantify the statements made regarding the modeling efficacy of AFc and BioEmu: the Pearson correlation coefficient is near unity for all sequences at all MD simulation lengths for both methods. Furthermore, for the 1000 ns simulation, the linear fit to the AFc points has a slope near unity, again

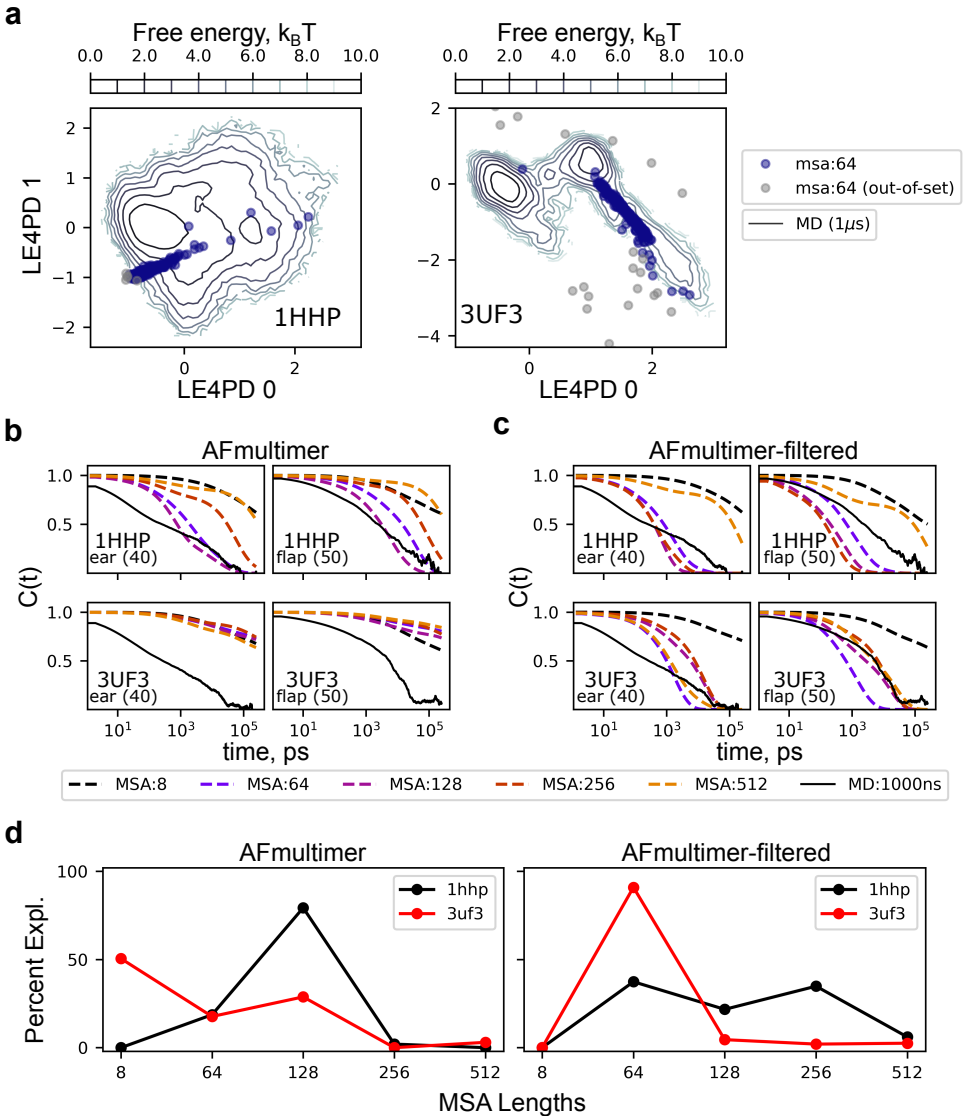


Figure 9.8: Applying my protocol for wildtype (1HHP) and a mutant (3UF3) HIV-1 protease dimer using AFmultimer. a) Projection of AFmultimer structures onto the free-energy surfaces spanned by the two slowest LE4PD modes predicted from a microsecond simulation started after equilibration of the structure deposited under PDB ID: 1HHP (left) or PDB ID: 3UF3 (right). b) Time correlation functions either calculated from MD (black curves) for the 'ear' and flap residues in 1HHP (top panels) or 3UF3 (bottom panels). c) The same time correlation as in b), but with the predictions from the AFmultimer ensembles calculated using the reduced ensemble remaining after removing out-of-set structures. d) Comparison between ACFs from MD and my protocol for both pre- and post- removal of out-of-set structures from AFmultimer.

showing AFc integrated timescales are well-matched to the MD integrated timescales at this simulation length. While the BioEmu slopes differ from unity at the 1000 ns simulation length, they are much improved at modeling the dynamics in the 100 ns long simulations. For the 10 ns long simulations, the slopes for both AFc and BioEmu are significantly greater than unity, indicating an overprediction of the timescales.

I note the negative results when DiG models the dynamics from MD simulation. In Figure 9.7b, the DiG timescales are significant underestimates of the timescales seen in the 1000 ns MD simulation. This effect is quantified in Figure 9.7c, where the slope of the linear fits to the DiG integrated timescales is essentially invisible for all sequences and simulation lengths. These results indicate DiG is learning little to no protein dynamics and is instead returning "ensembles" *very* tightly clustered around a single structure.

9.1.7 Dynamics of HIV-1 dimer from AlphaFold-Multimer

HIV-1 protease functions only as a dimer, and the vast majority of the experimentally resolved structures are of the dimer. Due to spontaneous aggregation in solution, the only NMR ensemble of the monomer available in the PDB is for the 1Q9P sequence.[284, 285] Furthermore, computational studies have shown clear differences in the dynamics between the WT and MUT proteases, especially in the flap region, which are strong evidence for the differences in pharmaceutical efficacy between wildtype and mutant proteases.[288, 289] As such, there have been many experimental[290, 291, 292, 293, 294, 295] and computational[296, 297, 298] studies performed to study the function and dynamics of the dimerized HIV-1 protease.

Thus, analysis of the dimeric HIV protease is important for predicting biologically relevant

dynamics. In contrast to protein monomer ensemble prediction, ML modeling of multimer ensembles remains challenging.[299] The nature of co-folding models can introduce artifacts and may result in spurious outputs. To avoid such complications, I avoid analyzing the dimeric HIV-1 protease using my proposed protocol as the core of this study. However, to provide a more complete picture of the predictions made by ML ensembles, I conclude this work by applying my workflow to wildtype (PDB ID: 1HHP[300]) and a mutant (PDB ID: 3UF3[301]) HIV-1 protease dimer.

I use the same MSA approach as for the monomers but use AlphaFold-Multimer (AFmultimer) [184] instead of AF2 to generate an ensemble of 320 structures at each MSA depth. Since the reduced MSA approach reduces the number of sequence-based constraints, the co-folding models suffer more at lower MSA depths.[152] At lower MSA depths, the fold of the individual chains seems to be preserved; however, the binding orientation and the contact can be altered significantly. As these models are not causal and do not account for environmental conditions, the complex's structure prediction seems to be driven solely to place the two chains in close proximity. Overall, this reality suggests a better MSA depth to account for biologically relevant dynamics is higher than the monomeric case.

To benchmark my results, I perform one-microsecond MD simulations starting from the corresponding crystal structures. A detailed analysis of my simulation is presented in the SI using a set of well-known collective variables[296] specific to this HIV-1 protease system. I also show that the slowest LE4PD modes capture the flap opening relevant to HIV-1 protease function. In Figure 9.8, I demonstrate the results of my protocol applied to my selected dimers. Figure 9.8a shows the overlap between the ensemble predicted by AFmultimer at an MSA depth of 64 and the MD distribution via projection onto the two slowest LE4PD modes from each simulation.

Analysis using other MSA depths is shown in the SI. Even at higher MSA depths, AFmultimer appears to generate a significant proportion of structures that lie outside the MD-sampled region.

Further structural analysis of ML ensembles revealed that AFmultimer prediction contained the dimer with C_{2h} symmetry rather than the well-known experimentally resolved C_2 symmetry. The physical relevance of dimers with C_{2h} symmetry is highly questionable, as it depends on the binding affinity and kinetics of that bound pose; however, C_2 symmetry corresponds to the protease's functional assembly. Due to the presence of such structures in the ensemble, the ACF predictions are well exaggerated when compared to MD in Figure 9.8b. For representative purposes, I chose flap tip residue 50 and 'ear'[296] residue 40. A detailed analysis of other such residues is presented in SI.

Following the reweighting method that penalizes out-of-set samples, I recalculated the ACFs for the reweighted ensemble. Figure 9.8c clearly shows how the relaxation timescales move in line with those calculated from MD. I calculated the DTW metric defined in the Methods section between the predicted ACFs from my protocol, both the raw and reweighted (filtered) ensembles, and the ACF directly calculated from the MD simulation. Figure 9.8d indicates that AFmultimer with an MSA length of 128-64 produces an ensemble that is closer to a 1 μs of MD simulation. Overall, these results: 1) support my initial hypothesis regarding the limitations of co-folding ML, like prediction of C_{2h} symmetry dimers with high confidence; 2) highlight the requirement of a reduced MSA-type treatment to encourage the model to generate more possibly biologically relevant structures; and 3) demonstrate a requirement of a reweighting procedure. These results further show the differences in the MSA depth - timescales relation for monomer and multimer ML methods.

9.2 Discussion and Conclusions

I provide a method qualitatively and quantitatively measuring the timescale of conformational ensembles output from a selection of deep, generative machine learning (ML) models for six sequence variants of human HIV-1 protease. I compute the timescales for the residue displacement autocorrelation correlation function (ACF) by assuming the backbone dynamics of the protein are governed by a coarse-grained, overdamped, and memory-free Langevin equation[274] with a potential of mean force that is acceptably parameterized by the covariance matrix calculated from the generated ensemble. Diagonalizing the Langevin equation produces modes that form a basis set for the calculation of ACFs. This interface between generative ML techniques and the LE4PD methods allows for one-shot generation of dynamics and kinetics from sequence.

The residue-dependent timescales of the variants are benchmarked against each other to compare qualitatively the timescales of the ACFs as a function of sequence. This comparison can help determine the effect of mutations on the relaxation kinetics while the root-mean-square fluctuations calculated from the theory allow for the sequence-dependence of the dynamics. Quantitative benchmarking is performed by brute-force calculation of the analogous ACFs using molecular dynamics (MD) simulation trajectories of a given length. This second step of benchmarking against a known ensemble constructed from a simulation or experiment is not required for the predictive aspect of my method.

I show a correlation between reducing the MSA depth in AF2 for structure prediction and an increase in the timescales of the predicted ensemble due, at least in part, to an increase in the predicted variance of the ensemble along the slow dynamical modes. Reductions in timescale correspond to a reduction in the slow conformational dynamics as the depth of the MSA is in-

creased (Figure 9.3). Decreasing the MSA depth can increase the number of out-of-set structures in the projected LE4PD space (Figure 9.3b). Removing outlying structures affects both the static properties (Figure 9.5a) and decorrelation timescales (Figure 9.5b). I show these results hold for the biologically functioning dimer (Figure 9.8) as well.

Finally, I move beyond AF2 and analyze the conformational ensembles of three more recently developed architectures: AF-Cluster,[95] the distributional graphformer (DiG),[272] and the biomolecular emulator (BioEmu).[244] I find AF-Cluster predicts the slowest dynamics; this prediction is due to a large number of outlying structures. BioEmu produces a much lower number of out-of-set structures, which makes its predicted decorrelation timescales more trustworthy compared to AF-Cluster, and the predicted backbone dynamics, as measured using the alpha-carbon root-mean-square fluctuation, are in excellent qualitative agreement with the 1 μs MD trajectories (Figure S13). Finally, I find that DiG samples only structures tightly clustered around an observed crystal structure, predicting virtually instantaneous and nearly frozen conformational dynamics.

I anticipate the combined use of MD, the structural ensembles generated from ML, and the LE4PD to refine the weights of each ML structure in the ensemble in a manner more precise than that proposed here, e.g. a maximum entropy approach.[264] A re-weighting approach moves the sampled structures closer to a causal Hamiltonian or Langevin model at a given dynamical timescale and lengthscale with sequence specificity. Furthermore, the LE4PD outputs can be interfaced with the ML architecture's loss function to tune predicted ensembles in a subsequent round of training so that the model infers an ensemble with the desired dynamics. I hope this type of approach helps move toward a causal generator of protein ensembles avoiding the computationally expensive querying of an integrator.

9.3 Methods

9.3.1 Using Conformational Ensembles to Parameterize a Langevin Equation to Approximate Conformational Dynamics

The fundamental assumption in this paper is the residue-level fluctuation dynamics of folded, solvated proteins at the microsecond scale and faster obey a memory-free, overdamped Langevin equation[274]:

$$\frac{d\Delta\vec{R}_i(t)}{dt} = -\frac{3k_BT}{\overline{\gamma}}\sum_{j,k} H_{i,j}A_{jk}\vec{R}_k(t) + \vec{\zeta}_i(t), \tag{9.2}$$

where $\vec{R}_i(t) = (x_i(t) - \langle x_i\rangle, y_i(t) - \langle y_i\rangle, z_i(t) - \langle z_i\rangle)^T$ is a vector measuring the fluctuations of the Cartesian coordinates of residue i from its equilibrium position $(\langle x_i\rangle, \langle y_i\rangle, \langle z_i\rangle)^T$, k_B is the Boltzmann constant, T is the temperature in Kelvin, $\overline{\gamma} = 1/N\sum_i \gamma_i$ is the average friction coefficient, and $\vec{\zeta}_i(t)$ is a white noise term modeling solvent collisions that obeys the following fluctuation-dissipation theorem: $\langle\vec{\zeta}_i(t)\cdot\vec{\zeta}_j(t')\rangle = 6k_BT\delta_{i,j}\delta_{t,t'}$. Finally, H_{ij} is the i,j^{th} element of the hydrodynamic interaction matrix giving long-ranged hydrodynamic interactions between residues i and j while A_{ij} is the i,j^{th} element of the potential of mean force matrix giving the pairwise interactions between residues i and j, which are, in general, long-ranged. This potential of mean force matrix is defined as the pseudo-inverse of the covariance matrix, $\mathbf{C} := \langle(\mathbf{R} - \langle\mathbf{R}\rangle)^T(\mathbf{R} - \langle\mathbf{R}\rangle)\rangle$ in the space spanned by the residue trajectories[274]: $\mathbf{A} := \mathbf{C}^{\dagger}$. For more detailed explanations of the $\mathbf{H}$ and $\mathbf{A}$ matrices, the reader is invited to read Ref. [274] and its associated supplementary material.

The coupled dynamics at the residue level is separated by applying a similarity transformation to Eq. 9.2, giving the following equation of motion[274]:

$$\frac{d\vec{\xi}_a(t)}{dt} = -\sigma_a \vec{\xi}_a(t) + \vec{\zeta}_a(t), \tag{9.3}$$

with $\vec{\xi}_a(t) = \sum_i Q_{ai}^{-1} \Delta\vec{R}_i(t)$, $\lambda_a := (\mathbf{Q}^{-1}\mathbf{HAQ})_{aa}$, $\sigma_a := \frac{3k_BT}{\bar{\gamma}}\lambda_a$, and $\vec{\zeta}_a(t) = \sum_i Q_{ai}^{-1}\vec{\zeta}_i(t)$ is the white noise term in the space of Langevin modes. The friction coefficients γ_i for each residue are calculated using the solvent accessible surface area and assuming it decomposes into solvent-dependent and a solvent-independent terms.[274]

9.3.2 Calculating Time Correlation Functions from Structural Ensembles

The mode solutions to the Langevin equation given in Eq. 9.3 decay exponentially with a characteristic timescale given by $\tau_a = \sigma_a^{-1}$ and lengthscale $l_a = \mu_a^{-\frac{1}{2}}$, and the ACFs in the space of residue fluctuations is given as a linear combination of the ACFs for the Langevin modes:

$$\langle \Delta\vec{R}_i(\tau) \cdot \Delta\vec{R}_i(\tau + t) \rangle = \sum_{a>6} \frac{Q_{ia}^2}{\mu_a} e^{-\frac{t}{\tau_a}}. \tag{9.4}$$

In Eq. 9.4, the summation skips the first six modes as they correspond to the three translational and three rotational modes of motion, which are irrelevant for modeling the internal conformational dynamics of interest here. Furthermore, $\mu_a := (\mathbf{Q}^{-1}\mathbf{AQ})_{aa}$. The integral of eq. 9.4, which

is eq. 9.1, can be written in the mode coordinates as follows:

$$\tau_{\text{avg}}(i) = \int_0^\infty \sum_a {Q_{ia}}^2 \mu_a \exp\left[-t/\tau_a\right] dt / \sum_a {Q_{ia}}^2 \mu_a$$
$$= \sum_a {Q_{ia}}^2 \mu_a \tau_a / \sum_a {Q_{ia}}^2 \mu_a = \langle \tau_a \rangle_i. \tag{9.5}$$

Defining the weight of mode a's contribution to $\tau_{\text{avg}}(i)$ as $w_a := {Q_{ia}}^2 \mu_a$ is used in transitioning from the left to the right side of the second equality in Eq. 9.5. As such, $\tau_{\text{avg}}(i)$ is the average mode decorrelation time for residue i, as predicted using the LE4PD.

Finally, since the Langevin equation is coarse-grained, it displays accelerated dynamics on the correspondingly smoothed free-energy surface. [302] To account for this speed-up, I approximate the missing barriers *a posteriori*[303] using an approximate scaling law for proteins derived in Ref. [304]: $\tau_a \rightarrow \tau_a \exp[\epsilon\sqrt{\lambda_a}/k_B T]$, where $\epsilon := 6.5$ kcal / (mol nm) is an energy per unit length. I use this approximation since it is found, in Ref. [304], to accurately describe the timescales of protein dynamics between the sub-Ångstrom to the nanometer lengthscales, which are the same as those examined here.

9.3.3 Measuring the Distance between Time Correlation Functions

To quantify the distance between the MD- and AF2-generated ACFs, I use two distinct methods: the Kolmogorov-Smirnov (KS) test[305] and dynamical time warping (DTW).[306, 307] The details on how these methods are used are given below.

To perform distance calculations between ACFs using the KS test, each ACF is converted into its corresponding survival function, $S(t)$. Defining the ACF $C(t) := \langle \Delta\vec{R}_i(\tau) \cdot \Delta\vec{R}_i(\tau + t) \rangle$,

I have $S(t) := 1 - C(t)$. This survival function is equivalent to the cumulative distribution functions normally input to a two-tail KS test, and it is used to calculate the usual KS distance between the two survival functions: $d_{\text{KS}} = \sup_t |S_{\text{MD}}(t) - S_{\text{AF2}}(t)|$.

Similarly, I use DTW in the usual manner: I calculate a mapping between the MD- and AF2-generated ACFs such that the path-warping distance is minimized: $d_{\text{DTW}}(C_1, C_2) = \min_p(c_p(C_1, C_2))$, given that p is a suitably defined warping path between ACFs C_1 and C_2 while the cost of an suitable warping path between C_1 and C_2 is given by $c_p(C_1, C_2) = \sum_{l=1}^{L(p)} c(C_1(l), C_2(l))$. In calculating the cost for taking an individual step along the warping path $c(C_1(l), C_2(l))$, I use only a single time series index l since the compared ACFs from simulation and theory are of the same length and with the same lag between each sequential value.

The percentage of each protein's dynamics best explained by the AF2 ensemble at the indicated MSA depth is calculated using the DTW distance matrix $\mathbf{D}$, where $\mathbf{D}_{m,n}$ represents the DTW distance between the ACF calculated by brute-force from the MD and from the LE4PD theory parameterized using the AF2-generated ensemble at MSA depth m for residue n (these matrices are shown in SI). This matrix is calculated at the indicated timescale. The percent explained is calculated as:

$$\mathbf{k} = \arg\min_m \mathbf{D}_{m,n},$$

where $\mathbf{k}$ is a vector of length N that contains the values of MSA depth at which the ACF from the LE4PD is closest to the MD, N is the number of residues, and $m \in \{8, 16, 32, 64, 128, 256, 512\}$.

Then

$$\% \text{ explained at } m = \frac{100}{N} \sum_{n=1}^{N} \delta(\mathbf{k}_n - m),$$

which counts the percentage of residues' ACF best matched by LE4PD parameterized from AF2 ensemble at MSA depth m. This calculation was performed for all three MD timescales (10, 100, and 1000 ns).

Chapter 10: Conclusion and Future Works

10.1 Summary

In this thesis, I have developed and demonstrated methods that work under the "sequence–ensemble–function" paradigm. I used the conformational selection hypothesis to combine AI-based and physics-based methods into a computationally efficient drug discovery protocol from a target assessment perspective. Further, I have demonstrated how such methods can be leveraged to enhance immunotherapies such as CAR T-cell therapy. Overall, this thesis has led to the development of the protocol called AlphaFold2-RAVE (AF2-RAVE). Applying AF2-RAVE to different problems led to the development of hierarchical AF2-RAVE and CAR-BOOST (Chimeric antigen receptor biophysically oriented optimization of synaptic topology). Further investigation of AF2 and other generative AI methods for biomolecular structural ensemble prediction led to the development of the Langevin equation approach for protein dynamics from generative AI methods. A summary of all these developments is shown in Fig. 10.1. From these model developments, the following conclusions were made:

- A post hoc integration of physics-based sampling methods is still necessary when using a data-based model like AlphaFold2.
- The multiple sequence alignment (MSA) dictates AF2's structure prediction to some extent.

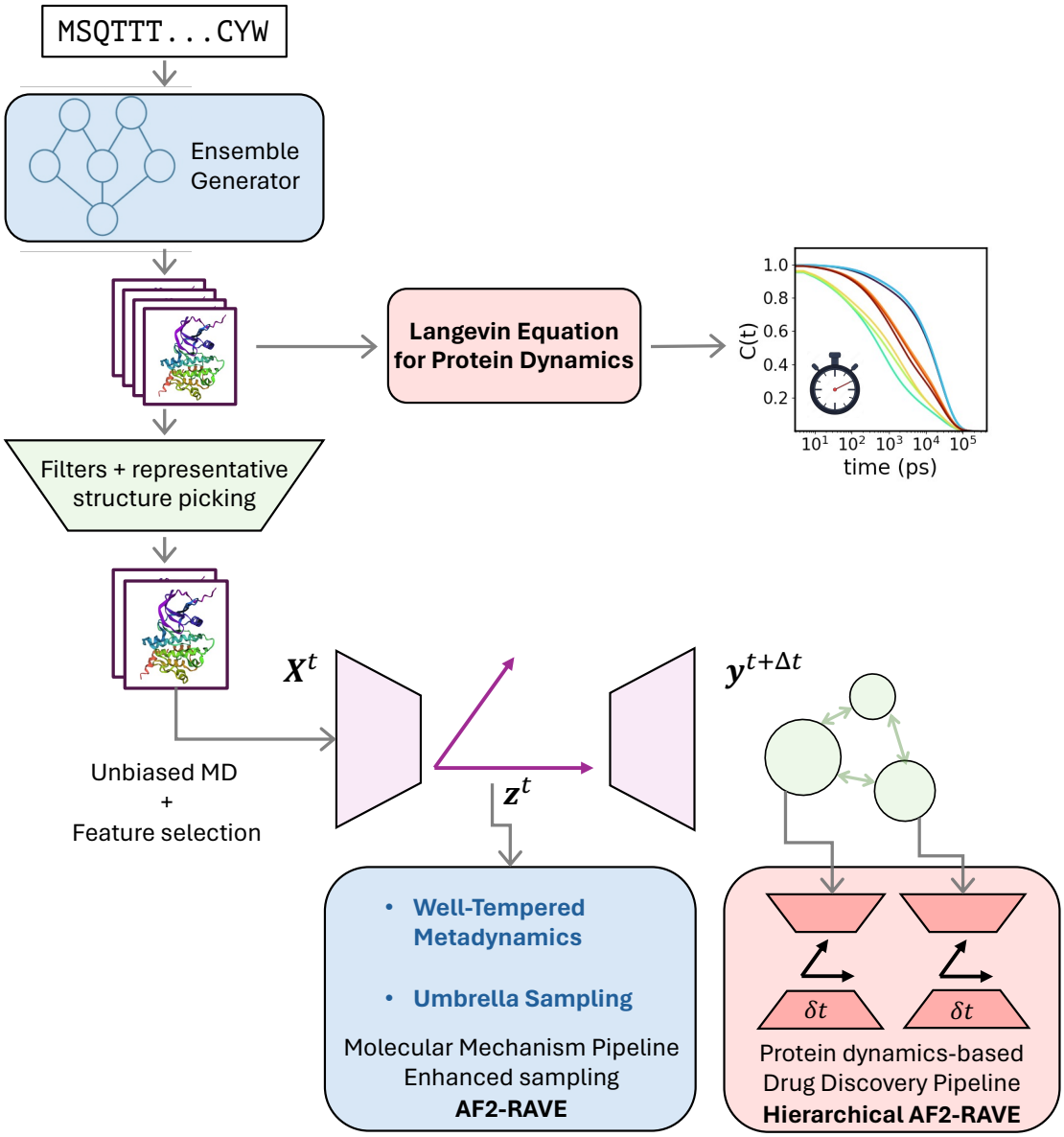


Figure 10.1: A summary picture of this thesis

The reduced-MSA approach for AF2 remains a reliable method for generating multiple non-Boltzmann-weighted conformations. The gap profile ('-') or variance in the sequence profile can make rMSA-AF2 produce a structurally diverse ensemble.

- By fitting the AF2-generated ensemble to a Langevin equation, I have demonstrated that reducing the MSA creates an ensemble with large-timescale changes. The lowest rMSA parameters yield an ensemble closer to the microsecond timescale, whereas mid-range rMSA parameters

produce a 100-nanosecond timescale. Further, I have demonstrated that recent methods, such as BioEmu, also produce an ensemble closer to the microsecond timescale. However, these BioEmu models suffer from the bias present in AF2's Evoformer blocks.

- Along with the state predictive information bottleneck (SPIB), AlphaFold2-RAVE can capture slow-timescale changes in biological systems. By cleverly selecting input features for SPIB, I have shown that a low-dimensional representation can be transferred to structurally similar proteins.

- AF2 provides docking-friendly structures; thus, a templated AF2 approach is required for the final docking studies.

- To capture changes across different timescales, I have shown a hierarchical approach for AF2-RAVE. One key point here is that a different set of input features is necessary for training SPIBs corresponding to different timescales.

- I have also shown that AF2-RAVE can be used to assess protein multimers predicted from a co-folding model.

- Inferring from the canonical T-cell synapse, I have demonstrated that CAR engineering can be enhanced by picking a linker based on geometrical constraints for an optimal synapse.

- By reusing the AF2-RAVE-generated data for CAR hinges, I have shown that the CAR-BOOST approach can be transferable to any target receptor in a computationally efficient manner.

10.2 Future directions

Figure 10.1 represents all the elements of this thesis. In addition to describing the methods, Fig. 10.1 also shows places where this framework can be improved.

10.2.1 Interpretation of AF2 for a given MSA

The coevolution signal attribution for AF2 remains a challenging question. The big-picture questions here are how much coevolutionary signal is required, for which specific region it is required, whether there is a sequence-pattern-to-fold memorization learned by AF2 without an MSA due to the masked language model loss in the Evoformer, and how to know when it has memorized the training set. Since AF2 is decomposed into initialization, Evoformer, and structure module components, another important question here concerns information flow. Perhaps understanding this can answer the three questions raised. Currently, the MSA block perturbation trick seems promising; however, an extended analysis with more proteins is required. In addition to this, interpretation of model embeddings would also be useful. For example, Simon et al. showed that a sparse autoencoder could be used to interpret protein language model (PLM) weights [308]. Their work showed that certain neurons can be associated with disulfide bonds, structurally proximal regions, binding domains, etc. Additionally, Zhang et al. used an embedding-analysis approach called "categorical Jacobian" to analyze a PLM [309]. They showed that the PLM indeed captures coevolutionary signals and relies on motif pairings. In a similar spirit, I have also contributed to interpreting RNA language models through embedding perturbation, specifically through layer-wise interpretation of the models[310]. Including the conclusion of Chapter 2, it can be said that such sparse autoencoders can be useful for AF2 or

AF3 model interpretation and additionally for separating the coevolutionary signals coming from the MSA and the signals that the model has learned.

In addition to that, there is a very general question I am interested in. Given that there is no inclusion of environmental conditions and that a static structure alone has been used for AF2, the question of the most probable or most stable structure based on free energy is out of the picture. However, based on all motif pairings, does this model understand basic questions like whether positive and negative charges attract each other, whether cysteines can form disulfide bonds (assuming oxidizing environment), and so on? Perhaps building a simple language model induced with structure[311] or residue-interaction bias can lead to a model that understands these chemical principles. The reason why this question is useful is that this analysis can perhaps lead to better models to address systems without any coevolutionary signals.

10.2.2 Better initial guesses with experimental hints

While these AI models produce conformationally distinct structures, the representative-picking step drops a lot of information. This bias can lead to errors in downstream prediction. To avoid this, clever representative picking is required. A better solution for this is the inclusion of experimental observables. Thus, moving forward, picking the best set of structures from the complete generated structure set that describes the experimental observable can lead to better results in the AF2-RAVE pipeline.

While this approach can help, if the generated structure set does not contain the states of interest, then this approach would also fail. Thus, guided generation would be the next step in this direction. There have been many methods that suggest guided generation of structures

[312, 313, 314]. The conclusion from all these papers is that perturbing the representation is by far the best solution for guiding this AF2 model. In light of these observations, perhaps an ensemble-based constraint can be used for guiding AF2. In the case of time-dependent experimental data, the Langevin equation fitting can be utilized as well[315, 316, 317].

10.2.3 Path-representing information bottleneck for non-overlapping initial trajectories

In the AF2-RAVE protocol, the hierarchical approach does not require overlaps between metastable states at large timescales. The reweighting or ranking approach using umbrella sampling requires overlap, but the relative rank can sometimes be calculated even with inexact umbrella sampling. However, for metadynamics (MetaD), since the simulation explores the information bottleneck space, the arrangement of the metastable states is quite critical. Generally, this would not be an issue if the simulations contained some overlaps. Practically, this requires a huge number of simulations from a large number of initial points. Perhaps a better stitching procedure could be used in such cases. The path-variable approach of defining a progress coordinate and a restriction coordinate can be highly useful because it guides the MetaD on where to go and where not to go[318]. Thus, incorporating these ideas to learn a path-variable space[319] based on the state predictive information bottleneck principles can lead to better reaction coordinate design if the initial trajectories do not have much overlap or if the initial trajectories are from biased simulations.

10.2.4 SPIB for protein subsystems

Analysis of large protein complexes is generally complicated due to the system size and the large number of metastable states. Hempel et al.[320, 321, 322] have shown that decomposing protein complexes into subsystems can be used to describe the dynamics of a complex system in an efficient manner. The same approach of iVAMPnets[321] can be used for SPIB, wherein the complex is divided into subsystems and features from each subsystem are passed through a weighting/attention matrix that contains the coupling information between the subsystems. Further, the unified SPIB will have different decoders for each subsystem. In this way, multiple-timescale changes in different subsystems can perhaps be modeled such that each IB space can encode information about a different subsystem at a different timescale. For example, in a protein kinase, the DFG side-chain flip and the activation loop can be separated into two subsystems to obtain a SPIB space where one or a set of coordinates encodes information about the activation loop and the other coordinates encode information about the DFG flip.

10.2.5 Context-dependent AF2-RAVE

Currently, the last part of the AF2-RAVE protocol contains ranking from US, exploration using MetaD, and structure picking using hierarchical AF2-RAVE for structure-based drug discovery. However, changes in environment are not directly learned in one run. That is, for different environmental conditions, the pipeline, mostly MetaD, has to be rerun. This requires a well-defined force field at intermediate temperatures and pressures. Methods like Boltzmann generators, Thermodynamic Maps, and Latent ThermoFlow have shown that perhaps training a generative model at two different temperatures can be used to extrapolate for any tempera-

ture [323, 324, 325, 326]. This approach can potentially avoid the overlapping issues mentioned above and provide a model that can reweight the distribution across experimental conditions.

10.2.6 Application perspective: Family-specific AF2-RAVE

One key question from my thesis is transferability. In the case of globular proteins, one possible way to approach a transferable model is to have a model that can work on a particular family of proteins. In my thesis, I have shown this for protein kinases, where I studied one kinase in detail and used the learned reaction coordinate for other kinases like Abl1 kinase and Src kinase. From this, one can obtain type-II-inhibitor-associated, classical DFG-out structures for any kinase given a sequence. Thus, for each protein family, a specific system that is as "flexible" as DDR1 can be found and studied in detail. Then, the learned knowledge can be used for other protein systems from its family. This provides computationally efficient transferability while not removing the human interpretation.

10.2.7 Application perspective: Acceleration-based designing of peptides and quantifying antigenicity

In Chapter 7, I showed how to assess TCR-pMHC complexes from co-folding models. The existence of multiple metastable states itself brings up an interesting perspective on these complexes. Perhaps for certain peptide mutants, such metastable states vanish, or they form newer or more stable metastable states to induce activation. Additionally, an infrequent MetaD-type approach can be used to calculate k_{off} values for the complex [327]. A comparative study on the sequence dependence of the antigen peptide for a longer residence time can perhaps be

related to the activation of T cells.

Appendix A: Supporting Information for Chapter 2

A.1 Evoformer Layer-Wise Interpretation

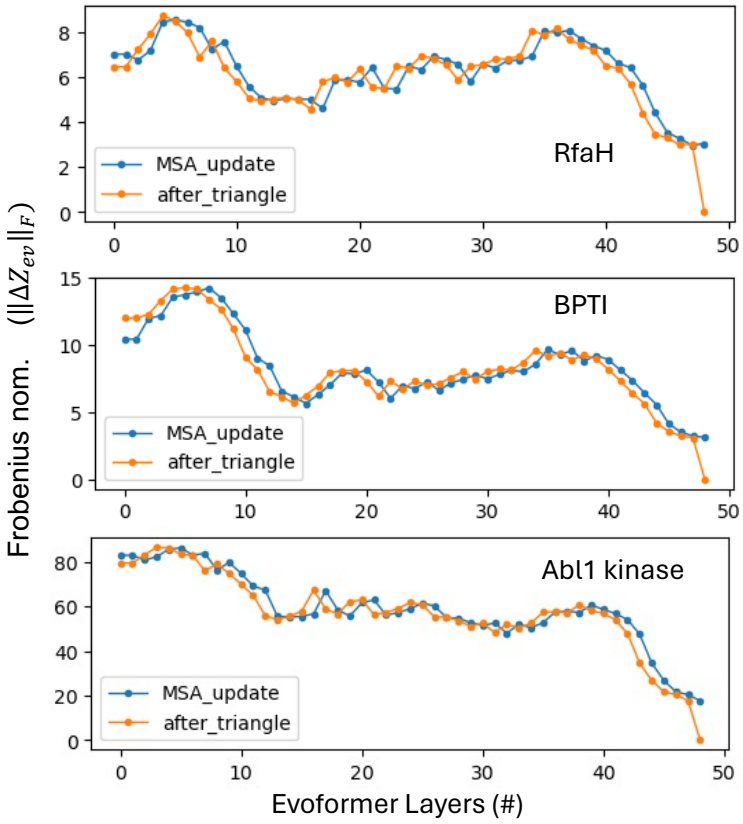


Figure A.1: Evoformer layer-wise analysis was performed for three distinct protein systems: BPTI, a protein containing multiple cysteine residues; RfaH, a fold-switching protein; and the Abl1 protein kinase domain, a large multidomain protein. The Frobenius norm was calculated between the pair representation at each Evoformer layer and the final pair representation. Each curve corresponds to the representation before or after the triangle attention/update operation. For all three proteins, the profiles show a two-step transition, with one transition near layer ∼8 and another near layer ∼46. This suggests that the final few Evoformer layers undergo a sharp transition toward the final prediction. However, a larger-scale study is required to make a more general statement. Such information may eventually help guide improved fine-tuning strategies for AF2-like models.

Appendix B: Supporting Information for Chapter 3

B.1 Methods

B.1.1 AlphaFold2

To generate a diverse structural ensemble, we made a key change to the publicly available Colabfold notebook[11]. The input featurization for the AlphaFold2 algorithm is in the form of two MSA clusters. The first is used as the input to the main AlphaFold2 network, and the second is used to initialize a pair representation matrix, which is the input for a second channel in the AlphaFold2 network. We decrease the size of both MSA clusters to (16, 32) and (8, 16), to obtain local conformational changes in CSP and SSTR2. For UBA2, to get more global changes, we further decrease MSA length to (4, 8) and (2, 4). As also reported in Ref. 91, we find that restricting the information in this featurization leads to revealing structures that are significantly different from the crystal structure, while providing hints to metastability. The choice of this MSA length is problem specific. In the case of CSP, we find that to obtain rotameric diversity , an MSA length of (16, 32) and (8, 16) is sufficient, while lower MSA lengths resulted in globular changes, as seed through RMSD. Since we were looking for unfolding, we used MSA lengths of (4, 8) and (2, 4) for UBA2. We find that lowering MSA lengths does give rize to unphysical structures that we have to remove, as well as structures that have no obvious unphysicality on inspection (e.g. broken bonds), but are thermodynamically unstable. The other computational choices and parameters are as follows. We use the model dropout feature, set number of recycles to 1, and use 128 random seeds– since AlphaFold2 generates 5 structures per random seed, we obtain 640 structures for each set of MSA hyperparameters. We then remove structures that are obviously unphysical (for instance, broken bonds).

The pLDDT score is sometimes suggested as heuristic for thermodynamic stability, however, this is not what it was trained for, and is not an ideal indicator of stability. Considering the lDDT[170] (Local Distance Difference Test) or, in the case of the AF2 output, its learned estimate, the pLDDT score is a distance metric from the crystal structure, it is not surprising that some of the structures that we need to sample to see conformational diversity are low in pLDDT scores. We analyzed the metric for GPCRs by running the reduce-MSA AF2 protocol by itself 7 times and computed average pLDDT scores. The scores ranged from 85.3 to 47.65, with a mean and standard deviation of 76.22 and 6.74. We then filtered those structures that do not overlap with our sampled structures, as shown in Fig.B.1, to find that they have a much lower pLDDT mean of 69.84. However, we note that the structures that we use to initialize our protocol have pLDDTs in the range of 68.42 to 85.3, we cannot use this has a metric to remove structures, either from training, or as thermodynamically irrelevant. Moreover we have found that the range for ”unphysicality” is different across both proteins and problem sets, for instance for UBA2, we

needed extremely low pLDDT scored structures to learn a reasonable CV.

B.1.2 Metadynamics

Algorithms developed to enhance sampling can be characterized in two ways: driving sampling by adding "biases" in the form of energetic potentials, or by splitting and stratification methods that focus on simulating multiple trajectory fragments and resampling them with biased probabilities[328]. In practice, the former is found to be preferable for enthalpic barriers, while the latter is more effective for entropic barriers. Our scheme is in theory agnostic to the sampling method used, as long as it is possible to recover unbiased statistics to obtain potentials of mean forces using the method. Here, we choose to use an easily implementable and relatively fast method, metadynamics[64]. Metadynamics functions by depositing Gaussian biases intermittently along the simulation so that regions of the collective variable space that have been traversed become less probable the more they are sampled, and the system is forced to sample rarer regions. This results in a time dependant bias that is recorded and can be used to generate a PMF along the collective variables used for bias. Additionally, it has been shown that an on-the-fly bias can be computed independently[6] to calculate a weighted histogram along any arbitrary collective variable. However, while essential, in practice the choice of collective variable is difficult. Biasing irrelevant collective variables or missing crucial slow collective variables[329] could result in observing no significant motion away from initial metastable state, or to sampling unphysical regions of configurational space.

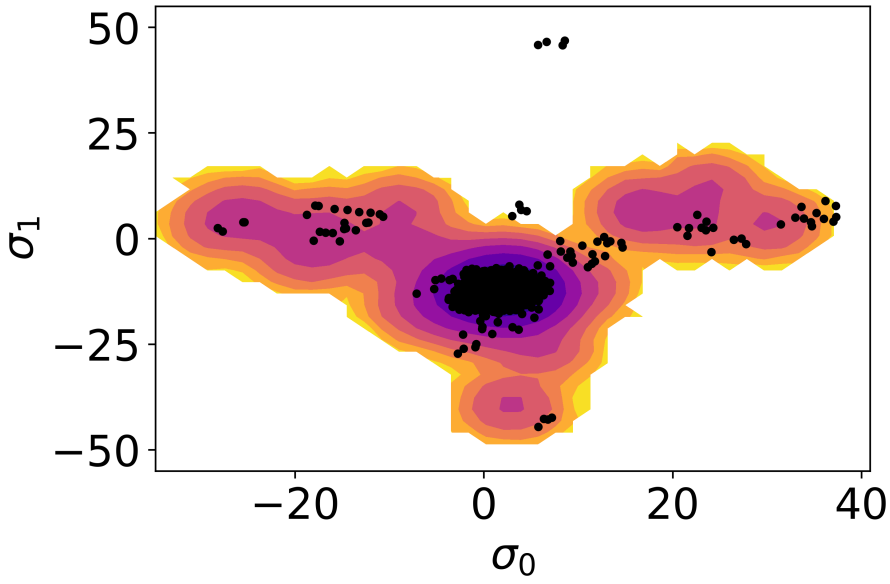


Figure B.1: Reduced-MSA AF2 structures projected on SPIB obtained from AF2-RAVE, used to compute average pLDDT scores. We use structures that do not overlap with regions we sample to compute these scores.

B.1.3 State predictive information bottleneck

Here, we use SPIB iteratively with biased or enhanced dynamics to learn the reaction coordinate[5]. SPIB is essentially a past-future information bottleneck protocol which takes time resolved trajectory data in a high dimensional order parameter space. It then predicts a latent space of desired dimensionality and identifies states in this space. Here we used a 2-dimensional information bottleneck which we denote $\{\sigma_1, \sigma_2\}$. The protocol requires an initial identification of state labels which is then refined iteratively. The model learnt through SPIB consists of an encoder, which transforms the high dimensional input into a latent space, and a decoder, which uses this latent space to predict state labels after a pre-set time-lag. Ideally, the latent space identified as the information bottleneck would correspond with our traditional understanding of reaction coordinates. This was demonstrated to be true for a model system in Ref. 5, where the latent space corresponds very well with the committor as defined in transition path theory. The information bottleneck is found by optimizing a loss function that maximizes the latent space's ability to predict the state of the system after a pre-set time-lag, while reducing the input dimensionality to the given bottleneck dimensionality. SPIB is capable of using both a linear or a non-linear encoder. In this work, we use a linear encoder with a non-linear decoder. We have found that when there are regions of state space that have zero sampling that we may aim to push the simulation into, a linear encoder is far superior as a non-linear encoder will often overfit to sampled regions while producing unphysical results in unsampled regions. This is important, since our inputs are unbiased trajectories initialized from cluster centers of the AlphaFold2 outputs. To remain as generalizable as possible, for our initial state labels we use time based labelling, i.e. the frames are labelled so that the first and second half of each trajectory individually form an initial state.

B.1.4 Other simulation details

The protein is represented by the AMBER03 force field[330]. The simulations are performed at 300 K with the leap-frog integrator in GROMACS 5.1.4[331]; LINCS was used to constrain the lengths of bonds to hydrogen atoms[332]; Particle Mesh Ewald was used to calculate electrostics[333]; the step size was 2 fs. The systems were solvated with TIP3P water models and equilibrated under NVT and NPT for 200ps and 300ps respectively. All simulations were run in water, as we intend this to be a demonstrative paper for this protocol, though we have found that it is suitable for simulations in other environments as well. We used PLUMED 2.4 [334] to extract collective variables. Other specific details about the systems, such as input collective variables, number of trajectories and number of RAVE iterations are noted in their respective sections below.

B.1.5 Collective variable choices

An important feature of our protocol, both in its versatility and adaptability, is the choice of input order parameters or collective variables in the clustering and SPIB learning stages. On one hand, SPIB's ability to handle a large number of inputs lends itself to highly exploratory work, the careful choice of CVs also lends itself to our ability to study extremely different scientific questions. In our work, we present three unique classes of protein structural changes, and our choice of CVs is key.

For CSP, we opt for the most generalized choice of collective variables to study rotameric states: all sidechain dihedral angles. For the 66-residue peptide, this results in 125 dihedral angles.

Similarly, for UBA2 to retain maximum information while exploring partially disordered states, we opt for the per residue Q value[72] and inter-helical angles as our input collective variables, defined for a given structure as follows:

$$Q(\theta_I) = \frac{1}{N}\sum_{\{i,j\}}^{N} \frac{1}{1+\exp\beta[r_{ij}-\lambda r_{ij}^o]},$$

where θ_I is the Ith residue, r_{ij} is the distance between atoms i and j for the specified structure, and r_{ij}^o is the distance in the reference (native-like) structure. The atom pairs i, j are chosen so that i belongs to the residue I, and $r_{ij}^o < 4.5$Å, i.e. all pairs of atoms that are geometrically in "contact" in the reference structure. We choose the single structure Alphafold2 outputs as our reference both wild-type and mutant. The parameters β and λ were taken from Ref. 72. For a notion of inter-helical angles, we use the α-carbons of residues {R9, F15, I21} and {A23, K29, A35}.

PHE92	GLN102	ASP122	GLN126
PHE127	ARG184	SER192	THR194
ILE195	ASN196	TRP197	GLY199
TYR205	PHE208	ILE209	THR212
PHE272	PHE275	ASN276	PRO286
LYS291	PHE294	TYR302	

Table B.1: SSTR2 residues used to compute pairwise distances as inputs for RAVE.

Our approach for SSTR2 is more targeted, as we specifically aim to sample conformations that change ligand binding affinities, so as to facilitate drug discovery. SSTR2 has a recently resolved experimental structure, and like most other GPCRs, the residues involved in ligand binding are known[335]. We begin by narrowing our CVs to this set of residues, tabulated in Table 1, then selecting all residue pairs that are (1) within 4.5Åof each other and (2) do not belong to the same transmembrane helix or connecting loop. We use the Cα distances between these residues as our initial OPs to input into RAVE. One key difference in protocol is that to select initial structures for unbiased sampling, we use regular space clustering[336], which results in a more diverse and less redundant initial guess, as AlphaFold2 tends to overamplify the probability of native-like or inactive structures.

B.2 Benchmarking the protocol using cold shock protein

To benchmark our protocol, we also run a control simulation on the cold shock protein (CSP; PDB ID: 1HZB). In this work, we present sampling on the State Predictive Information Bottleneck (SPIB), a 2-d linear combination of all 125 sidechain dihedrals to show the existence of several rotameric states. Here, we ensure that our method both samples the correct thermody-

namic profile for the dihedrals of the specific residue of interest, while also increasing ergodicity in sampling in different degrees of freedom.

CSP has known rotameric metastability in its eighth residue (Trp8), exhibiting 6 metastable states in its χ_1 and χ_2 (the first two dihedral angles following the backbone) torsions as per fluorescence spectroscopy [102]. We show representative structures for these 6 states with the relevant residue highlighted in purple in Fig. B.2.

To demonstrate the power of our method, we assume no knowledge of Trp8, or any other residues with metastable rotamers, and at every stage of the algorithm, we use all available 125 protein χ torsional angles. To benchmark thermodynamic weights, we run metadynamics[6] on the same system, but biasing along the Trp8 χ_1 and χ_2 angles so as to generate a PMF in the spaceB.3. We show that without using any *a priori* information regarding residues that are known to exhibit conformational metastability, we are able to obtain different competing conformations together with their accurate Boltzmann weights. Without using any specific preference towards the χ_1 and χ_2 dihedrals of Trp8, our protocol correctly ranks the 6 metastable states along these two dihedrals. We also rerun the same protocol using only the crystal structure initialized unbiased runs, to show that the reducedMSA AF2 structures are indeed useful. While we see some increase in sampling in the χ_1 direction, there is no sampling in χ_2. It is likely that on running for longer, we will begin to see sampling in χ_1 as well. However, the bottleneck learnt from this

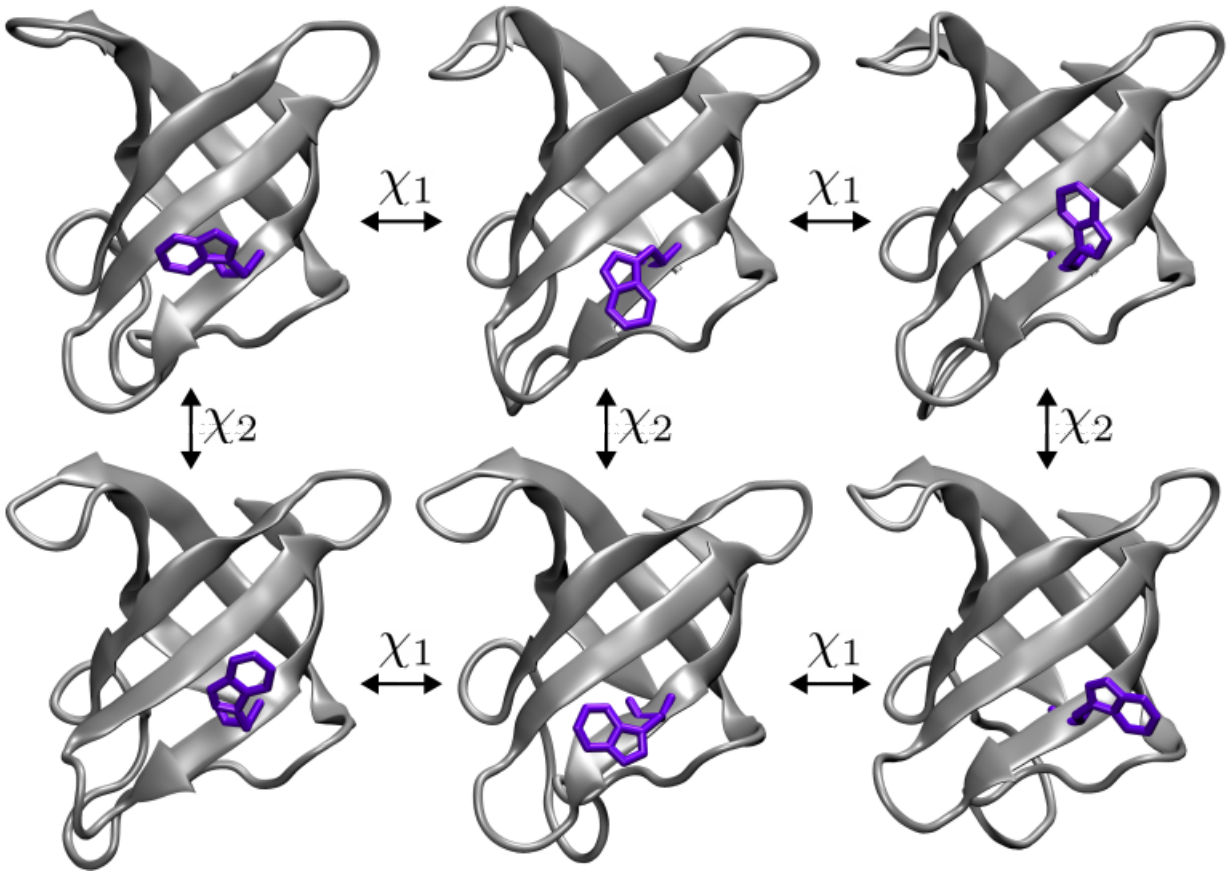


Figure B.2: Representative structures for the six metastable states in Trp8 orientations for cold shock protein 1HZB, with Trp8 highlighted in purple.

trajectory has very low emphasis on diversity in the Trp8 conformations, as those barely fluctuate in the initial input trajectory.

Further, we also show that our protocol is more general than the focused metadynamics run, and samples conformations along not just Trp8 but along all 53 residues that contain sidechains. To quantify our improvement in generalized sampling, we define sidechain dihedral space as the space of all configurations in all dihedrals, and measure volume explored in this space by an n-dimensional histogram, where n is number of sidechain angles for each residue. We define a measure of ergodicity enhancement as the increase in volume of dihedral configuration space explored. In Figure B.4, we see that our protocol not only samples the Trp8 residue conformations comparably to metadynamics directly on that residue, but also samples more extensively for all residues with sidechains, which directed metadynamics fails to do.

Finally, we consider the PMF on our SPIB coordinates in Fig. 2 (main text), which shows a much larger number of states than anticipated. Since the final SPIB is non-linear, it is difficult to confidently assign physical meaning to all the "states" it predicts. However, we can attempt a careful analysis of the trajectory, and we find that there are also some non-Trp8 residues that clearly change with the SPIB states in clean transitions, suggesting that SPIB is learning rotameric metastable states in multiple residues. In Fig. B.5, we plot our trajectories on the SPIB coordinates, colored as per the angles of these specific dihedrals on a periodic colormap. These are distinct from the bulk of the dihedrals where there is no correlation with the information bottleneck or its predicted states. While in this case, we are using CSP as a model, in cases where we are interested in the structural biology of a protein, such analysis would be very useful.

B.3 Conformational diversity in SSTR2

As seen in [2], the conformational diversity in GPCRs that is most important for drug design purposes is in interhelical angles and and distances, specifically those of TM6 and TM7.

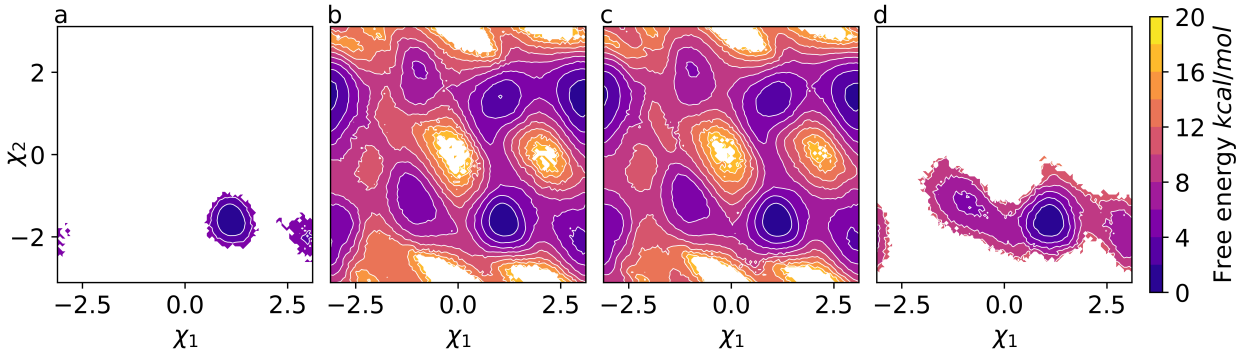


Figure B.3: PMFs along sidechain angles of the Trp8 residue which exhibits six metastable states. In (a) we show calculations from 20 ns long unbiased trajectory initialized from XRD structure. In (b) we show results of metadynamics performed on sidechain angles of the Trp8 residue. In (c) we show results from metadynamics performed using information bottleneck obtained without *a priori* knowledge of Trp8 metastability. Finally, in (d) we show a results from using our protocol *without* reduced-MSA AF2 inputs, learning a bottleneck with only the crystal structure initialized unbiased trajectory as an input. Colorbars for all 4 figures are shown with energies in units of $kcal/mol$.

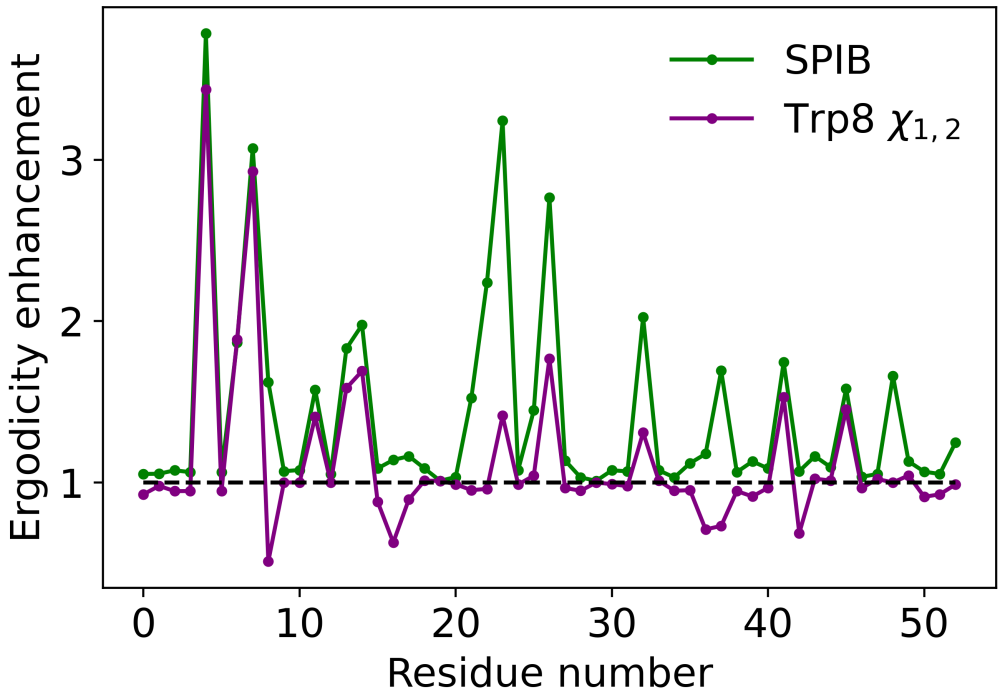


Figure B.4: Ergodicity enhancement in CSP residues. To examine the improvement in sampling of sidechain configurations, we compute the volume sampled in χ spaces for each residue and plot this for metadynamics performed by (i) directly biasing the sidechain dihedrals χ_1 and χ_2 of the Trp8 sidechains (purple) and (ii) biasing along the information bottleneck our methodology obtains (green). Both volumes are normalized by volume sampled by unbiased trajectories of the same length.

Since we assume no prior information, our collective variables are in terms of α-carbon distances of specific residues. However, we can use specific residue distances to establish that interhelical motions are key in our sampling.

Distances with highest weights in our SPIB are PHE92-ASN196 and PHE92-TRP197 (representing TM6-TM7 interactions), GLN126-THR212 (representing TM5-TM6 interactions, and PHE127-ILE209 (representing TM7-TM4) interactions. We also observe significant dihedral motions in helical residues in the ligand binding core of the protein.

B.4 Functionality and structural changes in UBA2

UBA2, present in the human Rad23 protein [104], and is a 3-helix bundle consisting of 41 residues, where the hydrophobic patch formed between the first and third helix is functionally crucial for binding to ubiquitin, and the constituent helices are extremely stable [337]. Studies show that L355A is primarily a disordered structure[104, 338] with loss of functionality. Our enhanced sampling order parameter is a learnt SPIB, which has non-uniform weights on per-residue Q values, representing a principled approach to combining residue contacts. Using this order parameter, we launched 4 500ns-long well-tempered metadynamic trajectories for both WT and L355A. We use total Q values as a way to characterize the disorder-order separation in UBA2 across both WT and mutant, and to support the stabilization of a partially disordered non-native

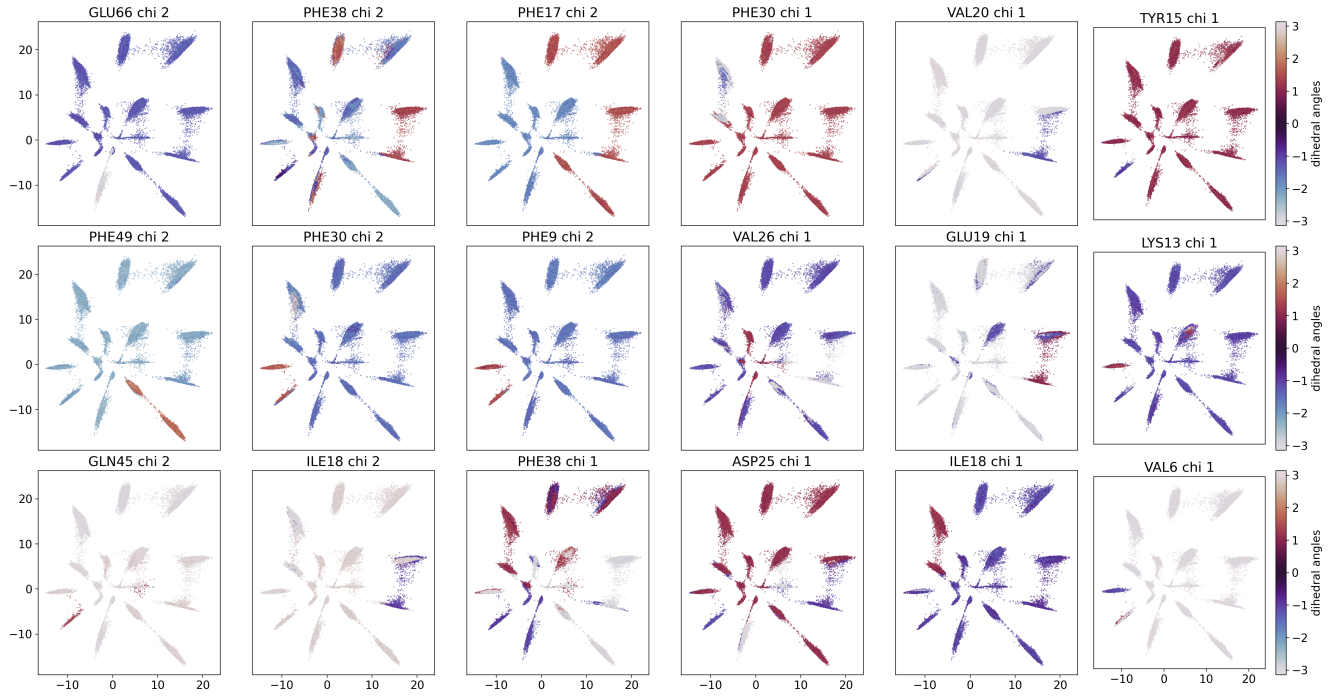


Figure B.5: The variation (represented by a periodic colormap) of some carefully chosen sidechain angles in our trajectory plotted on the SPIB coordinates. We see that the states are detecting rotameric transitions in several residues aside from Trp8 as well.

form to a missense (L355A) mutation. This is a convenient coordinate since it can be decomposed into per-residue Q values and reassembled as a linear combination representation of the SPIB.

For a stability analysis of UBA2's unfolding, we calculated the free energy difference between the folded states and unfolded states in Q-value space. To define the folded states we looked at the fractional helicity (discussed below) and used a Q-value of 0.9 (Q_c)to differentiate the folded vs unfolded. We justify the choice of this value through structural analysis. We find that the basin at Q-value 0.8 (Fig 3c in main text) contains conformations with the hydrophobic patch disconnected. The free differences for both WT and L355A are calculated using the equation

$$\Delta G_{U->F}(Q_c = 0.9) = G_F - G_U = k_B T ln(\frac{\int_0^{Q_c=0.9} e^{-\beta G(q)} dq}{\int_{Q_c=0.9}^1 e^{-\beta G(q)} dq})$$

An important point to be noted is that the $\Delta\Delta G$ calculated as the difference between ΔG^{L355A} and ΔG^{WT} is positive suggesting the point mutation is destabilizing.

Here, we present some more analysis on more structure based collective variables to interpret the mutant's observed loss of functionality. Fig. B.6 shows the PMF of both WT and L355A in a 2D space defined by both RMSD and Q-value, showing that our pipeline resolves the structural separation well. The PMF is constructed by reweighting 4 separate trajectories apiece, and we only present the PMF in regions of state space that are sampled by all 4 trajectories so as

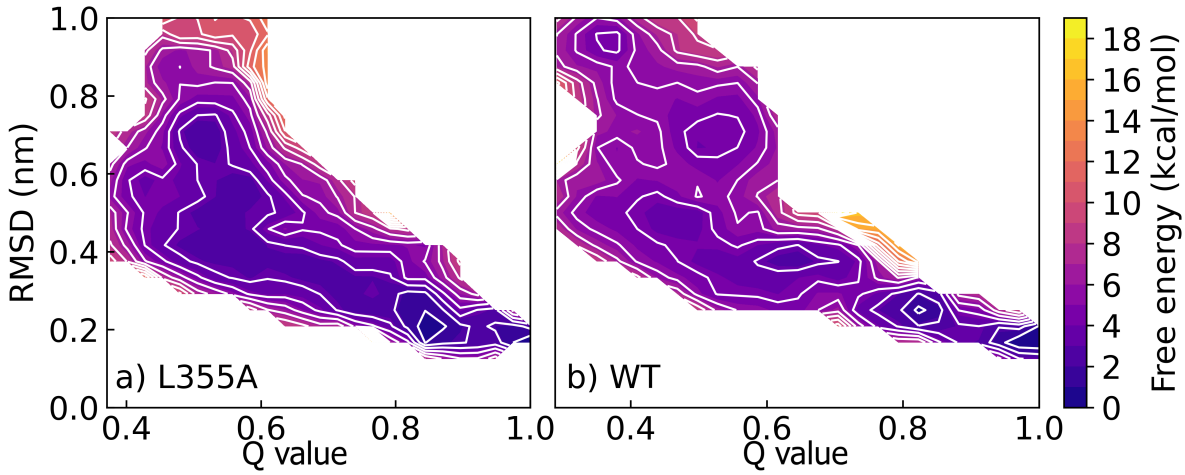


Figure B.6: PMFs plotted on Q value-RMSD space to represent global structural changes found by AF2-RAVE for a) L355A and b) WT. Here, for reference structure, we use an equilibrated structure from classical AlphaFold2.

to minimize error.

Fig. B.7 shows fractional helical content for the L355A and WT, i.e. the fraction of residues with backbone torsions corresponding to helices[339] on the Ramachandran plot. These results are presented as a weighted histogram, using weights from our RAVE-based simulations. Fig. B.7a shows that about 70 percent of the WT remains helical with little deviation, while the distribution for L355A is wider and slightly shifted. This corresponds with results in Ref. [103] which show that the protein retains helicity even when partially disordered. All these results can be refined by performing multiple rounds of RAVE, but here we show results from a single iteration.

Figs B.7 (b,c,d) show fractional helicity per individual helix (H1, H2, H3) as defined in the protein's WT native form. L355A residues from H1 and H3 have a broader distribution compared to the bimodal WT distributions, while the mutation causes a nearly complete loss in H2 helical content. This could suggest that the loss of helical content in H3 may be the cause for the loss of functionality due to mutation, as residues in the H3 along with H1 and loop connecting H1 and H2 form the hydrophobic patch that is crucial for ubiquitin binding.

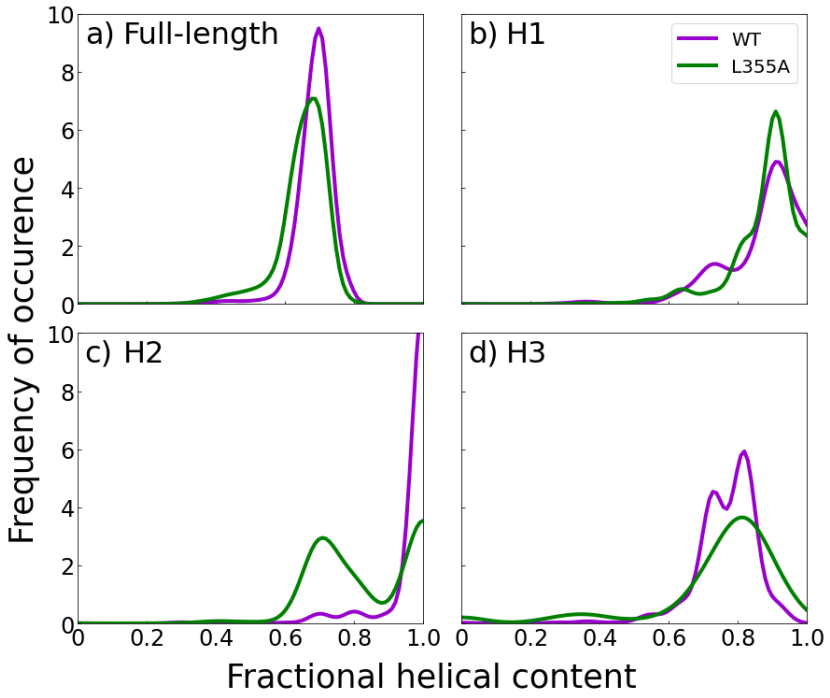


Figure B.7: Fractional helicity for L355A and WT reweighted by metadynamics time-independent weights[6] : a) residues from the full protein structure, b), c), d) residues belonging to helix1 (H1), helix2 (H2) and helix3 (H3), respectively. Helices are defined by structure in the protein's wild type native fold.

Appendix C: Supporting Information for Chapter 4

C.1 Methods

C.1.1 Training State predictive information bottleneck

Since we aim to make our protocol generalizable, we start with a large set of input CVs. However, biasing using metadynamics on a function of a large set of CVs is hard to control and not always statistically stable. To alleviate this to some extent, we adopt a basis CV refinement step as a stand-in for regularization in our OP learning protocol, wherein we run SPIB three times, each time discarding features with weights lower than 0.25 of the maximum weight. The intuition for SPIB here is to learn coordinates that separate the inputs and learn those coordinates that *are not* well sampled in the unbiased trajectories. For this, it is best to use long timescales, around $0.25 - 0.5$ times the length of our input trajectories

C.1.2 Simulation details

The protein is represented by the AMBER03 force field[330]. The simulations are performed at 300 K with the BAOAB integrator[340] in OpenMM[341]; LINCS is used to constrain the lengths of bonds to hydrogen atoms[332]; Particle Mesh Ewald is used to calculate electrostatics[333]; the step size was 2 fs. The systems are solvated with TIP3P water models and equilibrated under NVT and NPT for 200ps and 300ps respectively. To prevent melting during biasing, we also restrain the conserved αC-helix of the N-lobe. Large scale motion for this motif is essential, both to see a transition and for drug unbinding pathways[3, 342]. However, biasing CVs that include distances from this helix often leads to irreversible disordering, and we find that applying torsional restraints on residues 65 to 81 is sufficient to allow for smooth upward motion. In Fig. 4.1, we show the αC-helix and its position with respect to the DFG loop.

Supporting Information

C.2 Collective variables

Collective variables described in the main text are defined in detail in tables Table C.1,Table C.2, and visually represented in Fig. C.1. Distributions obtained from reduced MSA AF2 are shown in Fig. C.2.

Universal kinase motif	Conserved residue	ID for DDR1	Description
HRDxN	Asn	173	Asn of the HRDxxxxN motif
DFGAsp	Asp	186	Asp of DFG motif
DFGPhe	Phe	187	Phe of DFG motif
DFGGly	Gly	188	Gly of DFG motif
ChelE	Glu	74	C-helix conserved Glu
ChelX	-	78	Four down from the C-helix conserved Glu
X2	-	81	C-terminal of C-alpha helix
X3	-	165	Midpoint of section in Clobe roughly parallel to DFG motif
X4	-	189	Three down from DFG
sbridgeR	Arg	172	Conserved Arginine that makes a salt bridge with C-helix Glu
sbridgeK	Lys	57	Conserved Lysine that makes a salt bridge with C-helix Glu
PloopN1	-	18	First residue of the P-loop
PloopC1	-	23	Last residue of the P-loop

Table C.1: Residues included in collective variable definitions

Distances			
Kinase motif	**Atom name**	**Kinase motif**	**Atom name**
HRDxN	CA	DFGPhe	CA
ChelE	CA	DFGPhe	CA
PloopN1	CA	PloopC1	CA
PloopN2	CA	PloopC2	CA
DFGAsp	CG	HRDxN	CA
DFGPhe	CG	X2	CA
sbridgeR	CZ	ChelE	CD
sbridgeK	NZ	ChelE	CD
DFGGly	O	sbridgeR	N
ChelX	CA	DFGPhe	CZ
sbridgeK	CA	DFGPhe	CZ
sbridgeK	CB	ChelE	CB
sbridgeK	CB	DFGAsp	CB
X2	CA	X4	CA

Table C.2: Distances used as input collective variables for SPIB

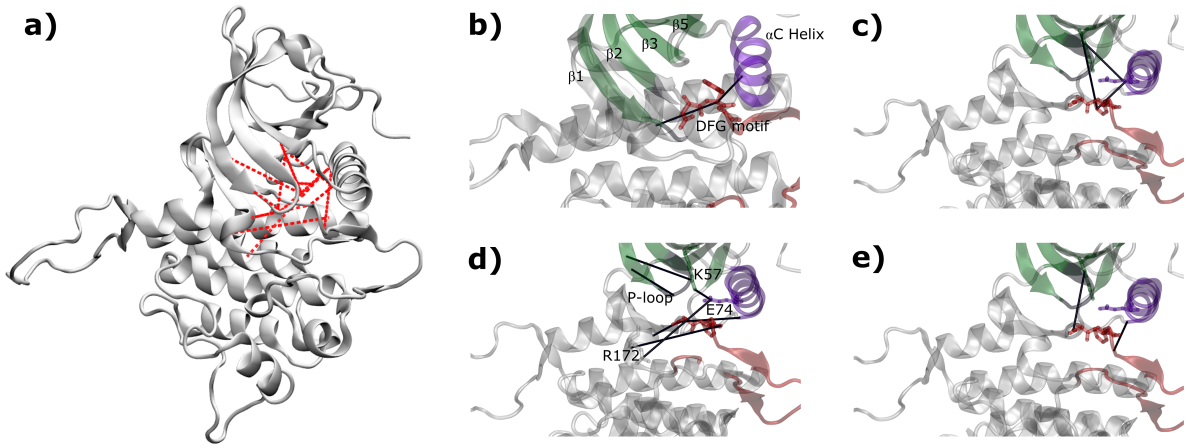


Figure C.1: a) Visualization of input CVs for SPIB in red dashed lines, detailed visualization of the CVs in b)-e) in black lines. b) CVs corresponding to the A-loop distance with N lobe αC helix and C lobe N173[7], c) CVs used for Dunbrack classification of DFG in/inter/out conformations that primarily focus on distances between DFG Phe and the conserved salt bridge between K57 and E74 present in the N lobe[8], d) CVs corresponding to A-loop DFG motif distances with to N lobe αC helix and C lobe R172. These CVs also represent the dissociation of the conserved salt bridge by including the salt bridge distance, as in c), and the structural integrity of the P-loop. e) CVs corresponding to A-loop DFG motif distance with N lobe αC helix and β3.

C.3 Comparison of AF2 results across kinases

In Fig. C.3, we show structural diversity obtained from reduced MSA AF2 for two other kinases in comparison with DDR1, specifically the Abl and Src kinases. We specifically choose these kinases as they are well known to have very different stabilities for the active and inactive states, are well studied, and are of medical importance[120]. It is clear that there is very no qualitative difference and very little quantitative difference in AF2's predictions for these three kinases– suggesting that for homologous molecules, AF2 learns very limited structural information. At low MSA depths (e.g. 8), AF2 generates diverse conformations almost agnostic to specific kinase identities, while at even slightly higher MSA depths (e.g. 16), AF2 does not see statistically significant diversity in either kinase sequence, with a strong bias towards the DFG-in active conformation. A detailed analysis of the effect of the MSA length on the conformational diversity of all three kinases is presented in the SI. As suggested in Monteiro et al.[150], varying MSA length can provide a qualitative understanding of the relative population for some sequences. Here, we see that for DFG-in dominant kinases like Abl and Src, this hypothesis is true. However, the relative populations are flipped for DFG-out dominant kinases like DDR1, and it is furthermore evident from Fig. C.3 that relative populations for all kinases are uncorrelated with Boltzmann stabilities. Reweighting these populations with their corresponding Boltzmann weights is the primary motivation for this work.

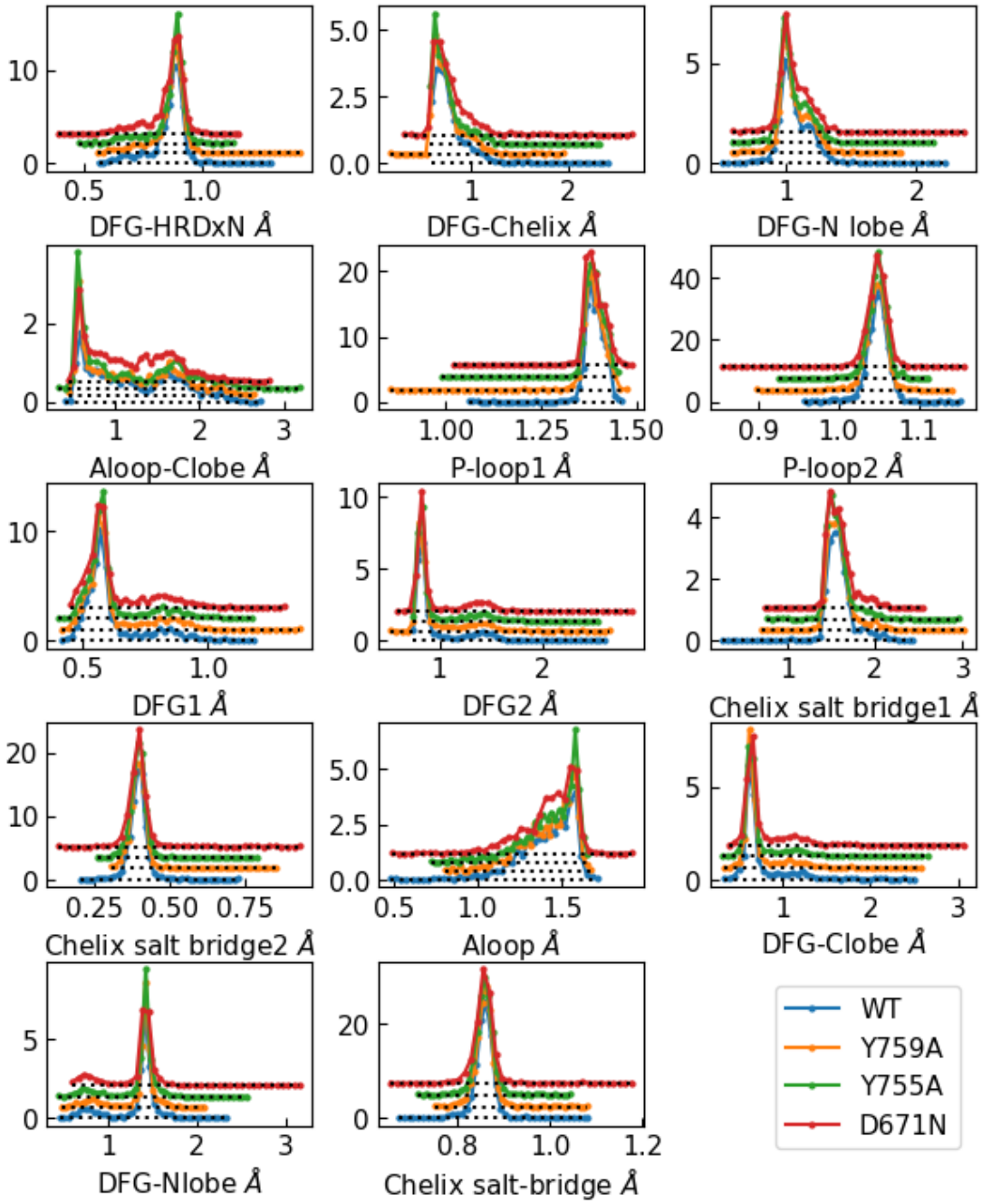


Figure C.2: Collective variable distributions for reduced MSA AF2 structures for wild type and mutants D671N, Y755A, and Y759A. These CVs are described in Table 1, and represent a large number of important collective motions in the molecule. The distributions have been plotted with offsets since these are highly overlapping and similar, indicating that AF2 even with reduced MSA can not distinguish between conformational stabilities of these sequences. The zero lines are marked with dashed black lines.

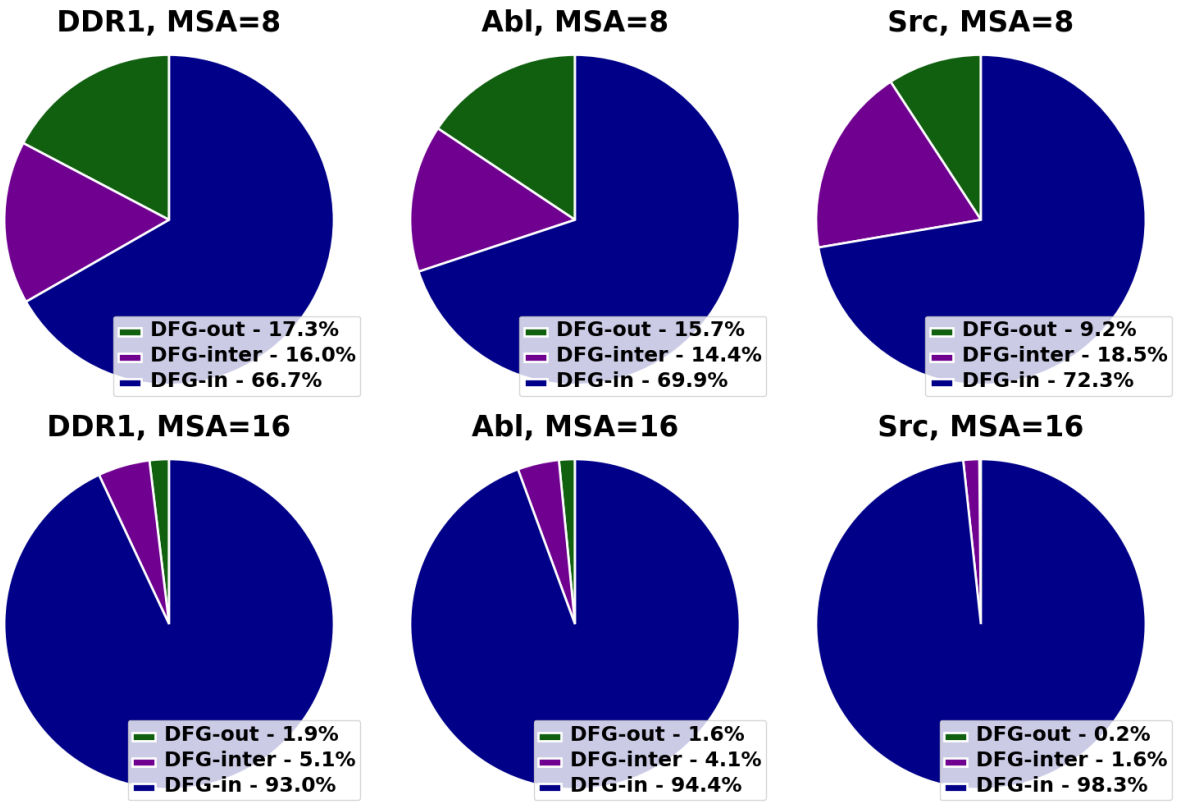


Figure C.3: Statistics of structural diversity obtained from AF2 reduced MSA on three kinases: DDR1, Abl, and Src, showing that AF2 does not really distinguish between them.

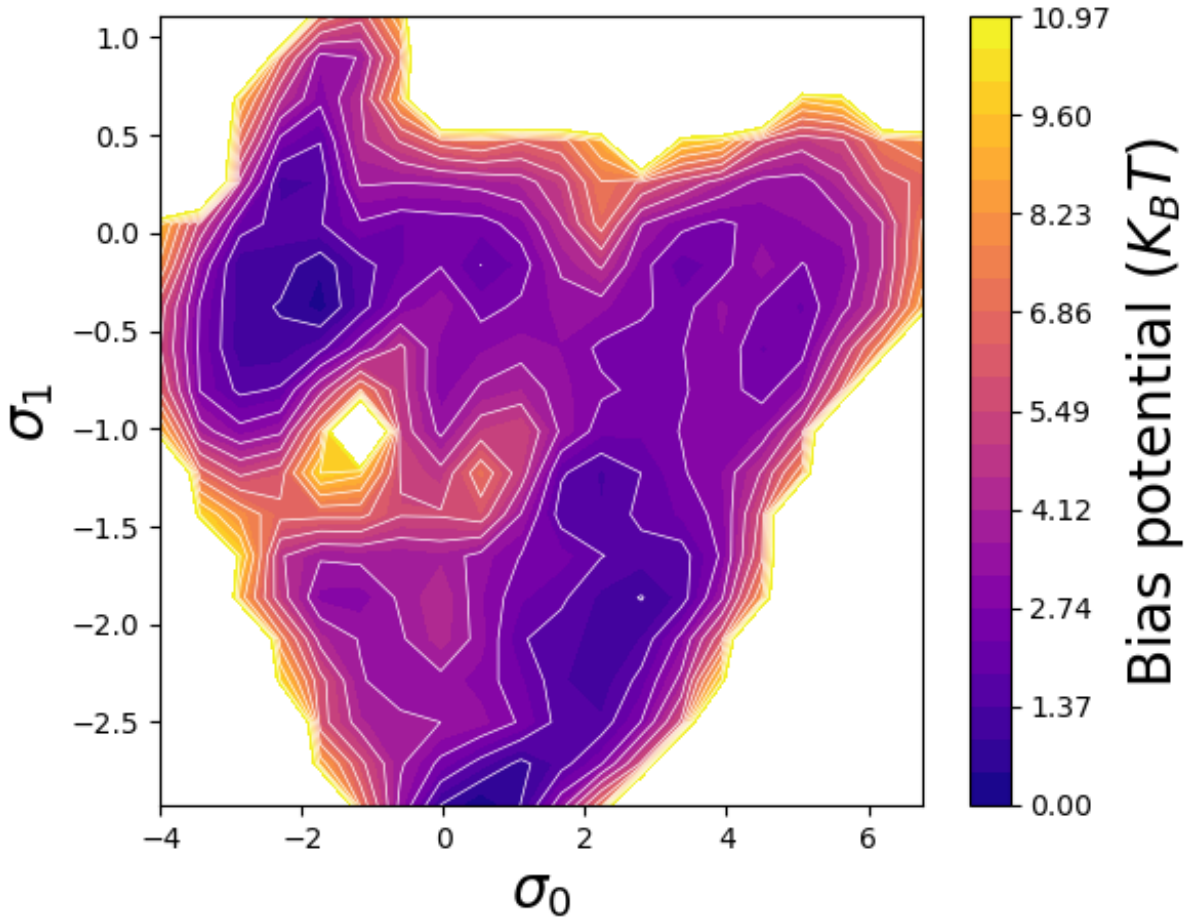


Figure C.4: Bias potential learnt from metadynamics on SPIB learnt from unbiased AF2RAVE seeded trajectories. The potential is plotted on SPIB coordinates $\sigma_{0,1}$

C.4 Sampling results

The bias potential learnt to perform our final calculations is shown in Fig. C.4. Our final combined PMFs are shown in Fig. C.5. To compare, we also show unbiased sampling trajectories on the same axes in Fig. C.6, noting that there is little overlap in trajectories between DFG-in and DFG-out states, with a clear separation in this CV space, hence necessitating the need for enhanced sampling.

In Hanson at al. the same four kinases that we study are simulated for extremely long time intervals using unbiased molecular dynamics. A large set of collective variables is extracted from these trajectories, which are then Processed using TICA. A two-dimensional TICA sata set is then clustered, and analyzed to provide free energy differences. While our method is different, we are still able to compare our results, and in Fig C.7, we provide a graph that is roughly parallel to theirs.

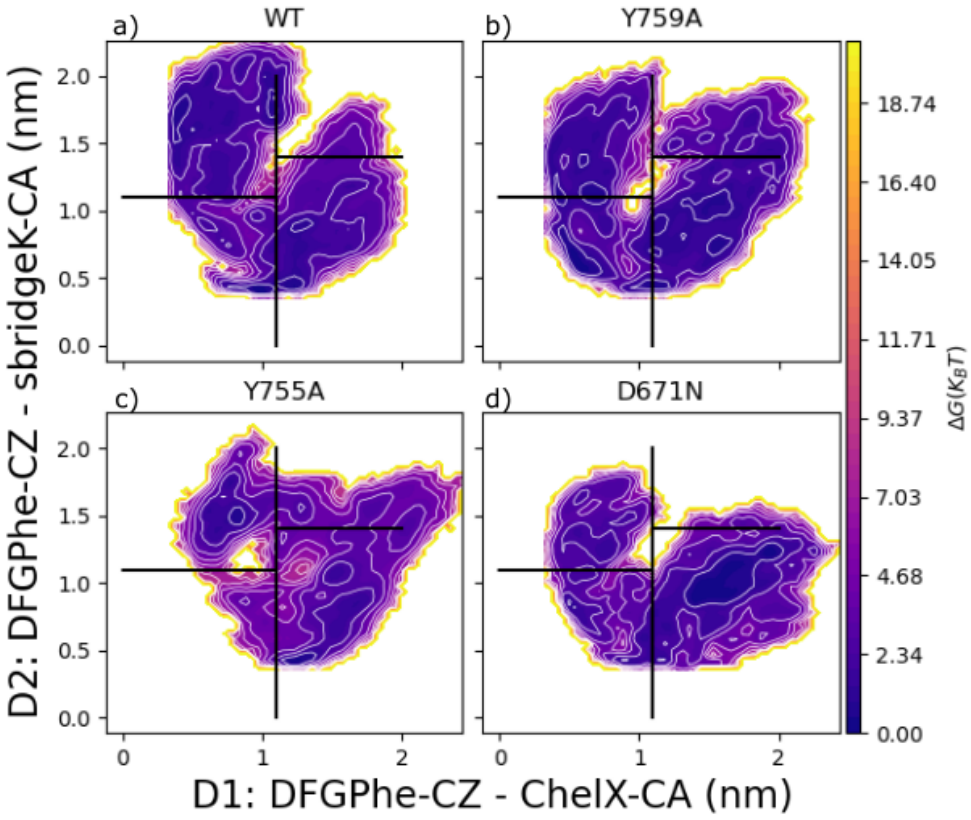


Figure C.5: PMFs sampled for all four systems projected on Dunbrack co-ordinates, a) Wild type, b) Y759A, c) Y755A and d) D671N. The solid lines differentiate the DFG loop conformations. The top-left quadrant describes DFG-in, the bottom-left quadrant describes DFG-inter, the bottom-right quadrant describes DFG-out and the unassigned belongs to the top-right quadrant.

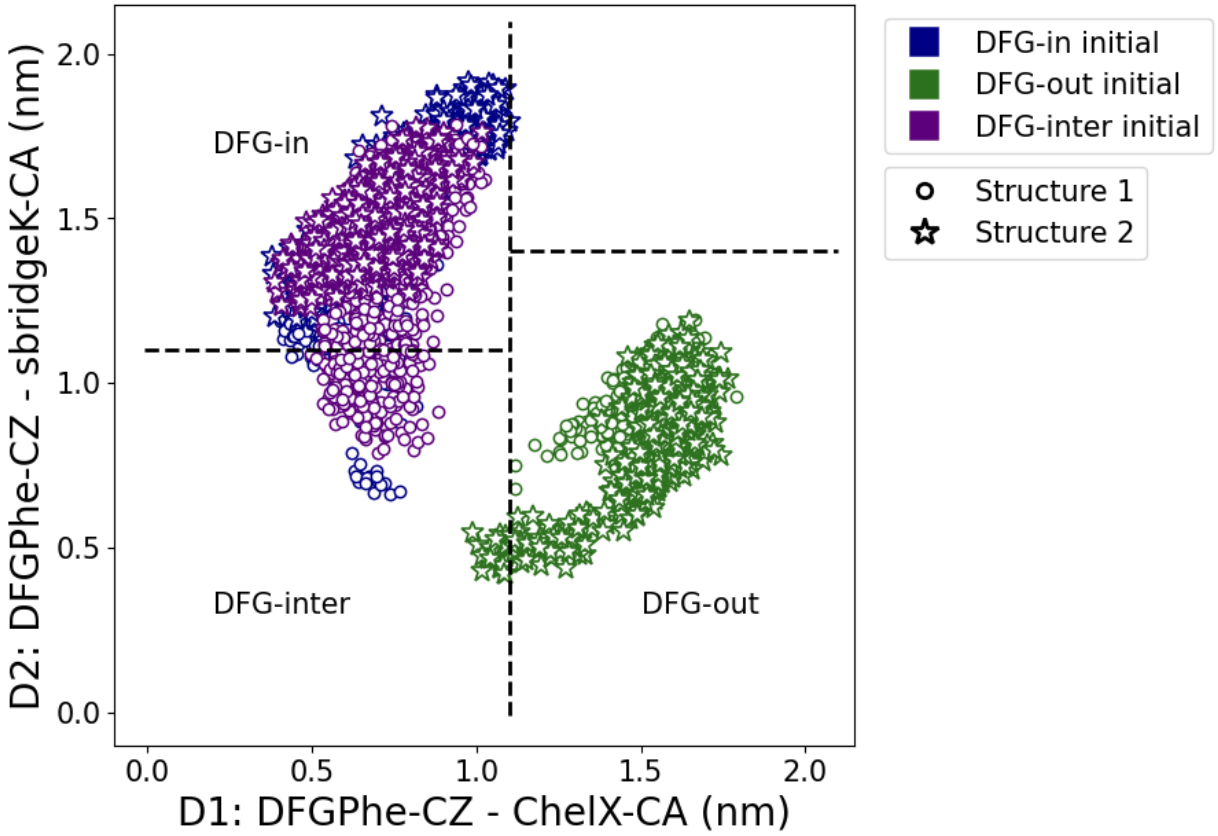


Figure C.6: 50ns unbiased trajectories seeded with initial structures from AlphaFold2 reduced-MSA

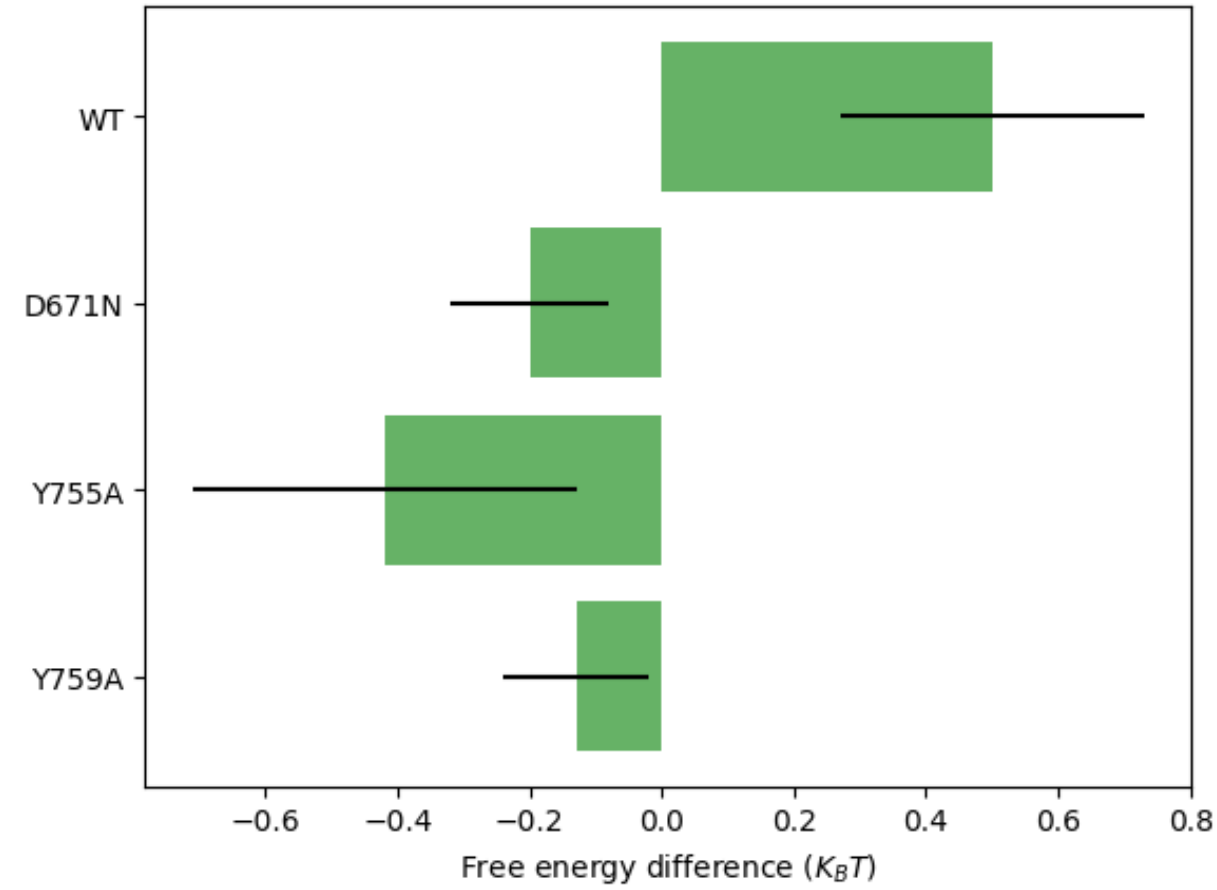


Figure C.7: Visualization of free energy differences for DFG-in to -out transition in bar chart form.

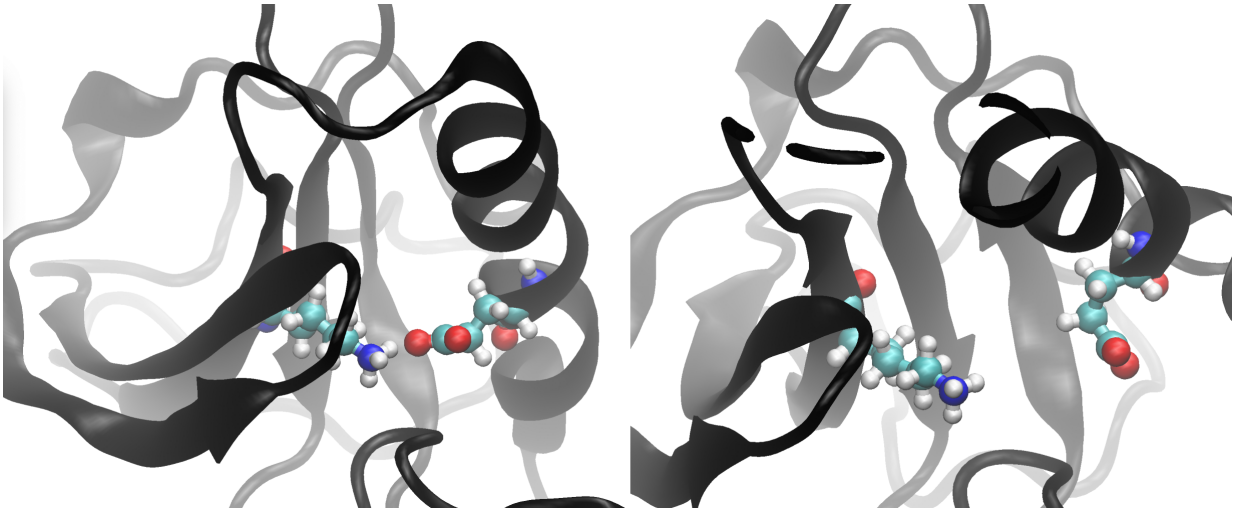


Figure C.8: Visualization of the E57-K74 salt bridge forming and helix upward motion observed during sampling.

C.5 Using GrASP on our results

In Fig. C.9 we show cryptic pocket predictions using GrASP[124] for various free energy minimum structures in DFG-in and DFG-out conformations for wild type DDR1. It is clear that we uncover new pockets by increasing sampling, and several of these correlate with druggable regions in other kinases, but further work must be done to investigate the usefulness of these pockets.

C.6 Code Availability

The code to run AF2-RAVE in a seamless manner is available at `https://github.com/tiwarylab/alphafold2rave`. This can be run on Google Colab using GPUs. Using Colab Pro is advised. Codes, parameters, and bias files used to specifically run the simulations from this protocol can be found at `https://github.com/tiwarylab/kinase_Aloop`. These will currently only work for the sequences used in this paper, but most files are easily adaptable for use in other kinases with some specific changes, which are marked. An example video is also in the folder, and while full trajectories are too large to upload, they can be made available on request.

C.7 Data Availability

All data associated with this work is available through `https://github.com/tiwarylab/kinase_Aloop`.

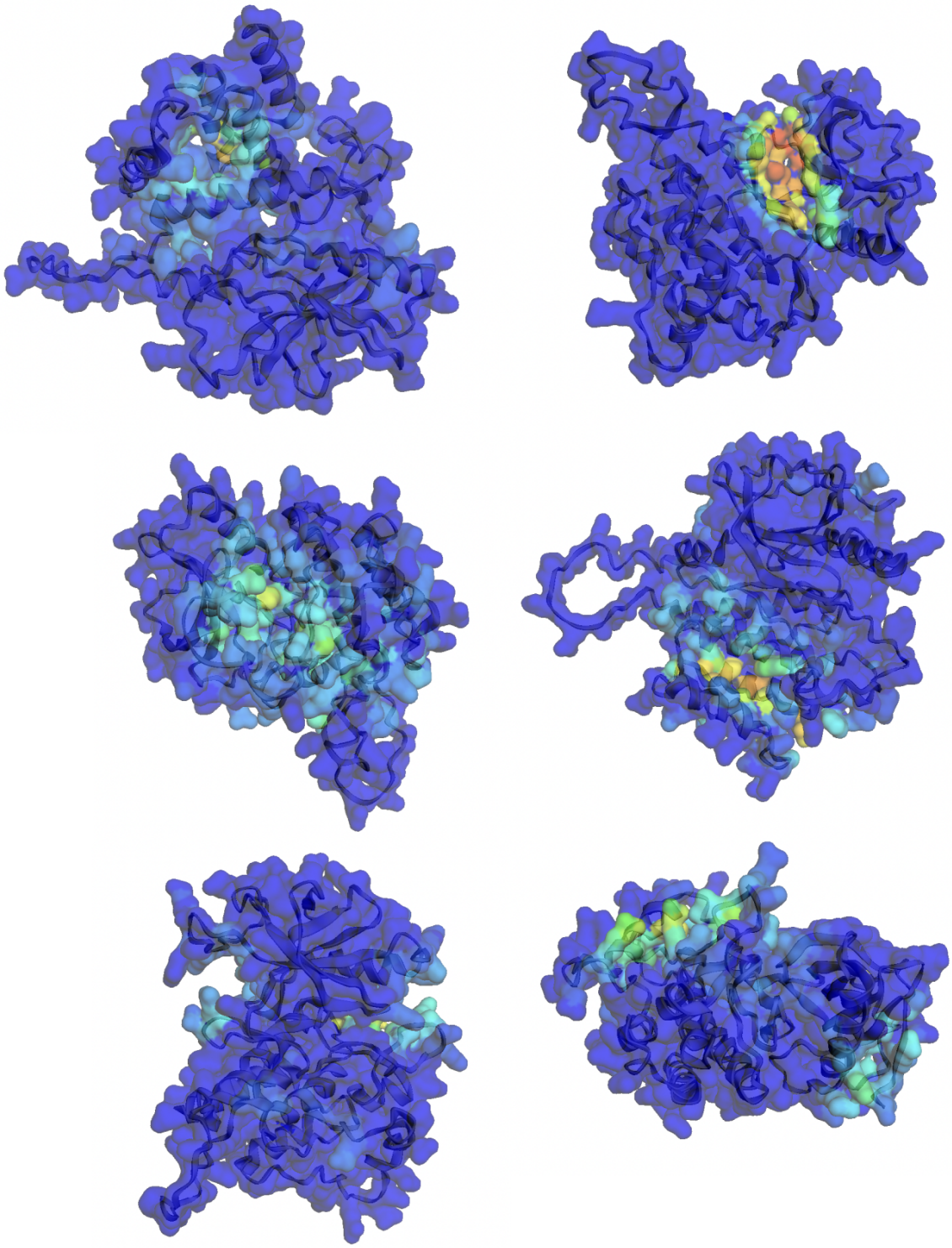


Figure C.9: Cryptic pocket predictions for various free energy minimum structures in DFG-in and DFG-out conformations for wild type DDR1.

Appendix D: Supporting Information for Chapter 5

D.1 Methods

D.1.1 Summary of systems tested and tools used

In this work we tested AF2RAVE-Glide protocol on active and inactive conformations of DDR1, Abl1 and Src kinases against type I inhibitors VX-680 and Dasatinib, and type II inhibitors Imatinib and Ponatinib. The top-ranked structures obtained from AF2RAVE-Glide were compared against publicly available crystal structures deposited in the Protein Data Bank (PDB), with PDB codes provided in the main text.

To generate diverse structural ensembles for DDR1, Abl1, and Src kinases, we primarily used the reduced MSA AF2 (rMSA AF2) method. For comparison, we also produced AF2-cluster ensembles for these three kinases.

For MD simulations, all the systems in this paper were parameterized with the Amber99SB*-ILDN force field [343, 344] with the TIP3P water model[345] and neutralized with Na^+ ions or Cl^- ions. The simulations are performed at 300 K with the LangevinMiddleIntegrator[346] in OpenMM[341] with the step size of 2 fs. Particle Mesh Ewald[333] is used for calculating electrostatics and the lengths of bonds to hydrogen atoms are constrained using LINCS[332] throughout all simulations. Before performing MD simulations for analysis, energy minimization is conducted for all initial structures, followed by equilibrium runs under NVT and NPT for 500 ps and 1 ns respectively.

To account for the induced-fit effect, Induced-Fit Docking (IFD) from the Schrödinger suite is the primary docking method used in this work. For comparison, we also tested the Glide XP docking from the Schrödinger suite and DiffDock to dock the two known type II inhibitors (Imatinib and Ponatinib) against DDR1 AF2 structure and the 15 classical DFG-out structures in the DDR1 rMSA ensemble.

Code and data for this paper are available in `https://github.com/tiwarylab/AF2RAVE_Glide-kinase`

D.1.2 Generation of rMSA AF2 ensemble, AF2-cluster ensemble and AF2-template structures

In this work, Colabfold [11] was employed to generate all the AF2-based ensembles and structures. The multiple sequence alignments (MSAs) were produced using mmseq2. For rMSA AF2 ensembles, the MSA depth was reduced to either 16 or 32 for each kinase. For each depth, 128 random seeds were utilized, and each seed produced 5 structural models via Colabfold, resulting in a total of 1280 rMSA AF2 structures per kinase. The structure with the highest pLDDT

score among all 1280 was identified as the AF2 structure (native structure). Subsequently, any unphysical structures with an RMSD greater than 7 Å from the AF2 structure were discarded, culminating in the final rMSA AF2 ensembles.

AF2-cluster for DDR1, Abl1 and SrcK were run with default setups as provided in the ColabFold notebook in the original paper of AF2-cluster[95].

The AF2-based homology modeling protocol, AF2-template (tAF2) method is implemented using Colabfold. For a given query sequence, tAF2 structures are generated by Colabfold upon uploading desired template structure and deactivating the Evoformer module. For each template, 5 AF2-template models are generated, and the last 3 structures exhibiting lower pLDDT will be discarded. The AF2-template (tAF2) method is employed here to transfer structure sets between homologous systems. To generate tAF2 structures for Abl1 in the classical DFG-out state, each of the 15 classical DFG-out structures from the DDR1 rMSA AF2 ensemble is used as a template, resulting in 30 tAF2 Abl1 structures in total. For Src kinase, a single representative tAF2 structure is generated using the "holo-model" DDR1 structure as the template.

D.1.3 AF2RAVE protocol

D.1.3.1 Regular space clustering on the rMSA AF2 ensemble

We used the same 14 collective variables (CVs) for regular space clustering as in the previous AF2RAVE work on kinases[114]. These CVs are pariwise distances, selected to describe the kinase conformations around the ATP-binding pocket and the A-loop.

D.1.3.2 Unbiased MD and SPIB

50 ns unbiased MD simulation was run for each AF2RAVE initial structure of DDR1 from Figure E.8. The standard deviations of the 14 CVs are calculated after concatenating all the 12 unbiased MD trajectories. 8 CVs with standard deviations larger than 0.25 of the maximum standard deviation remain as the input features of the SPIB model. We conducted a parameter screening of the SPIB time lag, ranging from 1 ns to 40 ns with intervals of 1 ns. Eventually, we selected a time lag of 16 ns based on the performance of SPIB coordinates in representing physical features, including DFG-type and A-loop position.

D.1.3.3 PMF calculations from umbrella sampling

2D umbrella sampling is conducted along the two learnt SPIB coordinates, employing 11x11 windows. The bias potential equilibrium points are uniformly distributed in the SPIB latent space, with σ_1 ranging from -0.8 to -0.1 and σ_2 ranging from 0 to 0.8. The strength of the bias potential is set to 1000 $kJ/mol/nm^2$. Each window originates from the structure closest to the window's equilibrium point in Euclidean distance within the latent space among the 12 AF2RAVE initial structures and lasts for 100 ns (with the first 10 ns discarded in PMF calculation). For Abl1 kinase, we noticed a substantial proportion of the αC helix breaking in the umbrella sampling trajectories (Figure E.19). This finding aligns with earlier enhanced sampling investigations on DDR1 using metadynamics[114], where the authors imposed restraints to prevent αC helix breakage. In our study, we opted to exclude all umbrella sampling windows where

the ratio of broken αC helix exceeded 20% for the Abl1 PMF calculation (Figure E.20). The WHAM algorithm is applied to bin and reweight the biased trajectories and compute the final PMF.

D.1.3.4 PMF calculations from unbiased MD

50 ns unbiased MD simulation was run starting from each structure in the 15 classical DFG-out structures in DDR1 rMSA AF2 ensemble. Upon discarding first 10 ns, all the unbiased trajectories are simply concatenated to calculate the Boltzmann distribution and PMF for each bin around the classical DFG-out basin in the latent space. For Abl1, unbiased simulations start from 30 AF2-template structures to calculate the PMF around the classical DFG-out basin.

D.1.3.5 Boltzmann ranks assignment for structures in AF2-based ensembles

After calculating the PMF value for each bin in the latent space, we projected AF2-generated structures into latent space. We then directly assign the PMF values of the corresponding bins to these AF2-generated structures.

We must acknowledge the limitations of the way we assigned PMF values to AF-generated candidate holo structures. First of all, the free energy profiles are derived from MD simulations, and the PMF values directly correspond to the MD structures. Here, we assumed that the latent space adequately represents the conformational changes of protein pocket within specific metastable states. Additionally, while the enrichment of holo structures in the top Boltzmann-ranked structures persists, the absolute PMF values and Boltzmann ranks of proper holo structures may fluctuate with the umbrella sampling setups, as depicted in Figure E.11. Theoretically, the number of umbrella sampling windows and the simulation length should be sufficiently large for PMF convergence. However, there is always a trade-off between PMF accuracy and computational costs, so we opted to stick with the current setups.

D.1.4 Docking details

All the input structure for our docking experiments were first relaxed in solution with a MD energy minimization step. This work investigates two type I inhibitors (VX-680 and dasatinib) and two type II inhibitors (imatinib and ponatinib) by docking. For Glide XP docking or IFD, we used Ligprep in Maestro to prepare the ligand inputs from SMILES files. For DiffDock, the ligand inputs were directly provided as SMILES files.

D.1.4.1 Glide XP Docking

Glide XP docking experiments in this work were run with default setups in the Maestro software. Glide XP docking was performed for all four ligands on the AF2 structures of DDR1 or Abl1, as well as on the 15 classical DFG-out conformations of DDR1 in the rMSA AF2 ensemble.

D.1.4.2 Induced-Fit Docking (IFD)

For Induced-fit Docking, we used the Glide XP for initial docking, followed by Prime relaxation and final Glide XP docking. Parameters remain default in Maestro IFD. IFD was performed only for the type II ligands on the AF2 structures of DDR1 or Abl1, on the classical DFG-out conformations of DDR1 or Abl1 in the rMSA AF2 ensembles, as well as tAF2 structures of Abl1 and SrcK.

Given ponatinib's backbone features, notably its lengthy and slender carbon-carbon triple bond, it exhibits reduced sensitivity to steric clashes, resulting in successful docking with the DDR1 "holo-model" structure at a ligand RMSD of 0.89 Å. Conversely, the "holo-model" structure struggles to accurately dock with imatinib using IFD (Figure E.5 B). We then employed an extended-sampling IFD approach by initially trimming the DFG-Phe residue from the "holo-model" structure. The trimmed residue is temporarily mutated to alanine during initial docking and later restored in subsequent Prime relaxation and final docking steps. Here, in the "holo-model" structure, we manually chose the DFG-Phe residue which is significantly hindered by holo-imatinib (Figure E.6 B). For generic systems lacking groundtruth information, broader screening of single residue trimming for the protein pocket may be necessary for this extended-sampling IFD method. In essence, it's a trade-off between the quality of holo-like structure to dock with and the accuracy/complexity of the docking method.

D.1.4.3 Diffdock performace on DDR1 classical DFG-out in rMSA AF2 ensemble

DiffDock docking experiments in this work were run in the webserver with default setups in the version before 3/8/2024 (`https://huggingface.co/spaces/simonduerr/diffdock`)[131]. DiffDock was performed only for type II ligands on the AF2 structures of DDR1, as well as on the 15 classical DFG-out conformations of DDR1 in the rMSA AF2 ensemble.

Appendix E: Supplementary Figures and Data

Supplementary Data

Supplementary File 1. Confidence score for the DiffDock pose aligns with AF2RAVE pmf values. The DiffDock confidence score of the pose with the lowest ligand RMSD (marked in red/bold) from each classical DFG-out structure in DDR1 rMSA AF2 ensemble is compared with the AF2RAVE pmf value for corresponding strutcure (marked in red/bold).

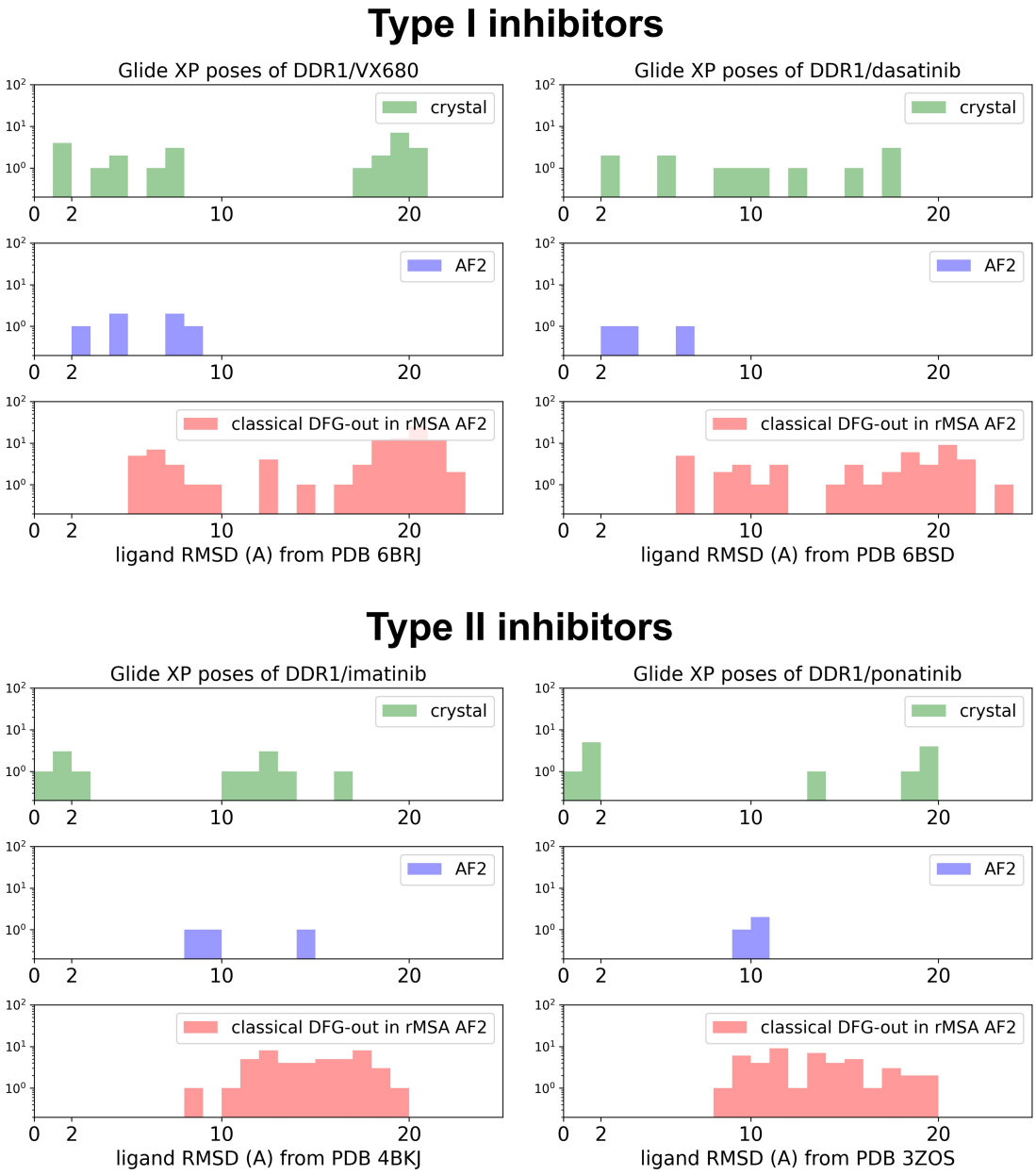


Figure E.1: The distributions of ligand RMSDs for Glide XP docking poses of DDR1 and type I/type II inhibitors (upper/lower panel). Results from cross-docking against 4 crystal holo structures, docking against the AF structure, and docking against 15 classical DFG-out structure in rMSA AF2 ensemble are shown as green, blue, and red, separately.

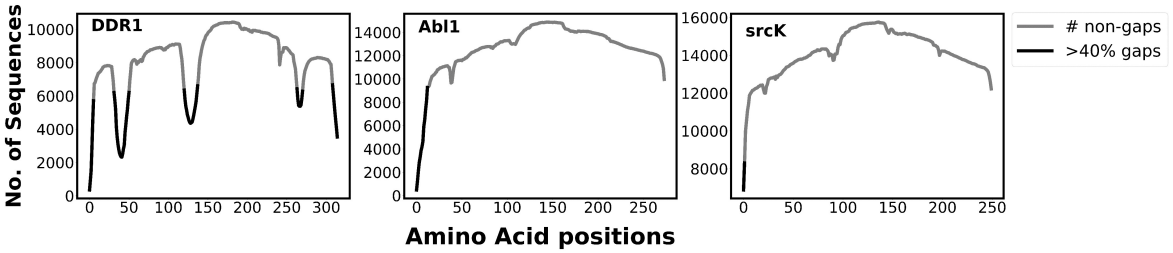


Figure E.2: The plot illustrates the number of gaps in the multiple sequence alignment (MSA) generated by mmseq2 (using Colabfold [11]) for different kinases. The non-gap count describes the coverage of each position in the MSA. The presence of residue positions with gap counts higher than 40 per cent of the total sequence in DDR1 implies that it has fewer conserved regions than abl1 kinase and src kinase. This characteristic of DDR1 MSA enables the rMSA AF2 protocol to generate multiple conformations for DDR1, including the classical DFG-out conformation, by initializing it at various states. However, the highly conserved nature of abl1 and src makes it challenging for the rMSA AF2 to initialize at a state that can lead to a classical DFGout conformation. Therefore, we used the AlphaFold template protocol to overcome this initialization issue with rMSA AF2.

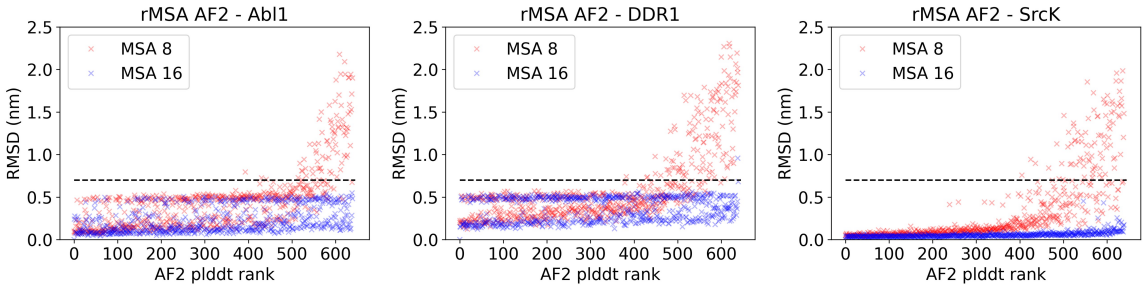


Figure E.3: The AF2 pLDDT rank is plotted against the CA RMSDs from the AF2 structure (the one with the highest pLDDT) for each structure in the rMSA AF2 ensemble for Abl1, DDR1 or Src kinase. A RMSD cutoff of 7 Å (dashed black line) is applied to filter out unphysical structures with large RMSD from the native structure. Each rMSA AF2 ensemble is consist of 1280 structures, 640 for MSAs of depth 8:16 (red) and 16:32 (blue), separately.

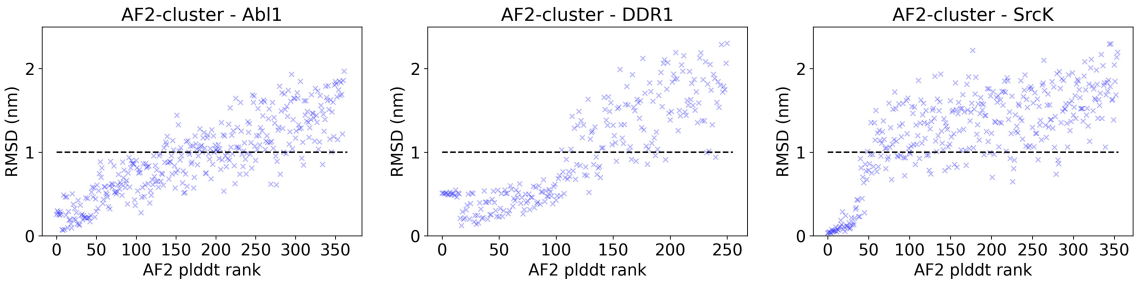


Figure E.4: The AF2 pLDDT rank is plotted against the CA RMSDs from the AF2 structure for each structure in the AF2-cluster ensemble for Abl1, DDR1 or SrcK. A RMSD cutoff of 10 Å (dashed black line) is applied to filter out unphysical structures with large RMSD from the native structure. After the RMSD filter, 197 out of 362 structures remain for Abl1, 134 out of 251 structures remain for DDR1, and 93 out of 355 structures remains for SrcK.

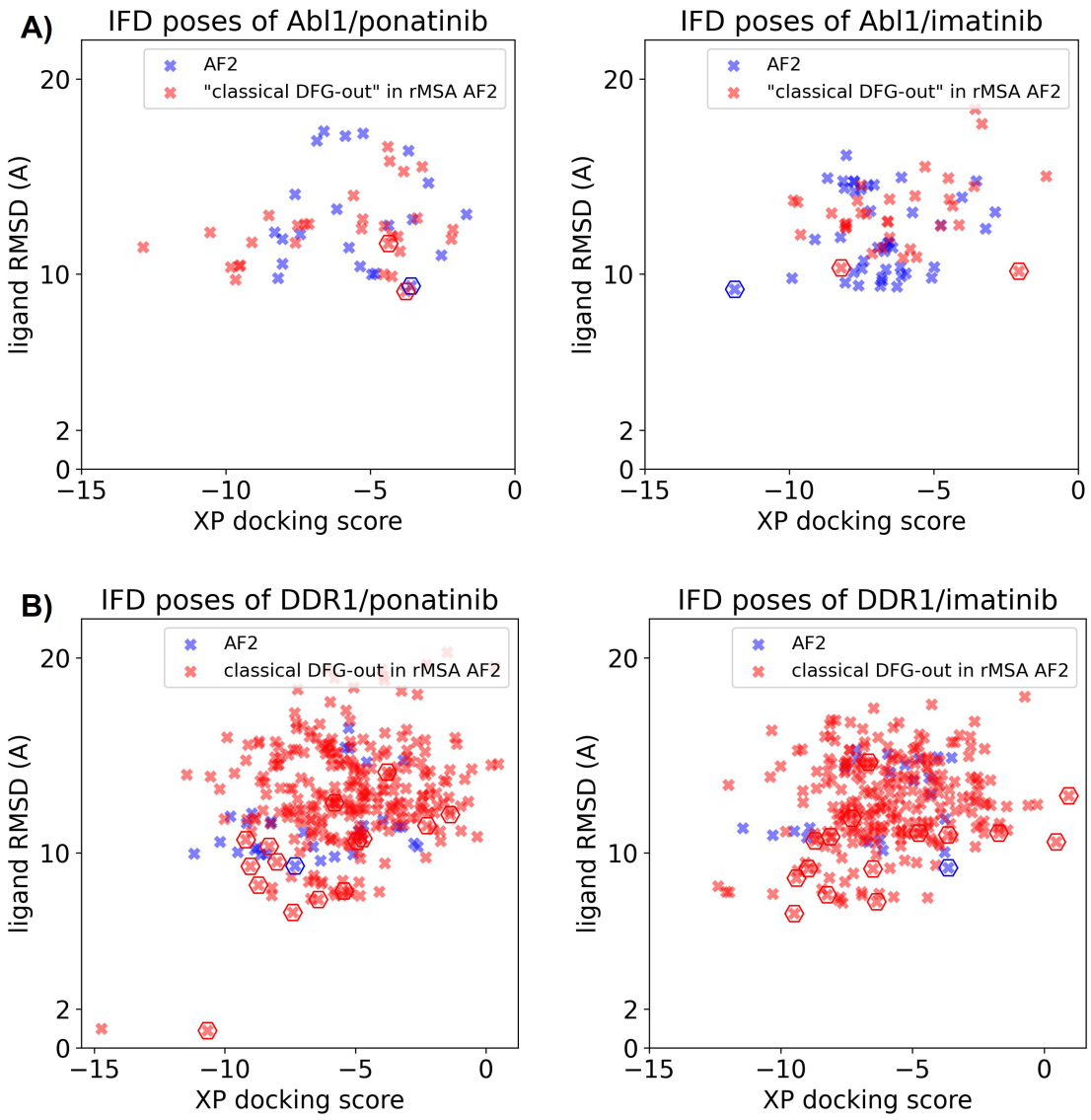


Figure E.5: Ligand RMSDs are plotted against the docking scores for the IFD docking poses of type II inhibitors (ponatinib and imatinib) against AF2 structure (blue) or classical DFG-out structures in rMSA AF2 ensembles (red). A) IFD docking results for Abl1. B) IFD docking results for DDR1. The pose with the lowest ligand RMSD from each input structure is marked by hexagon.

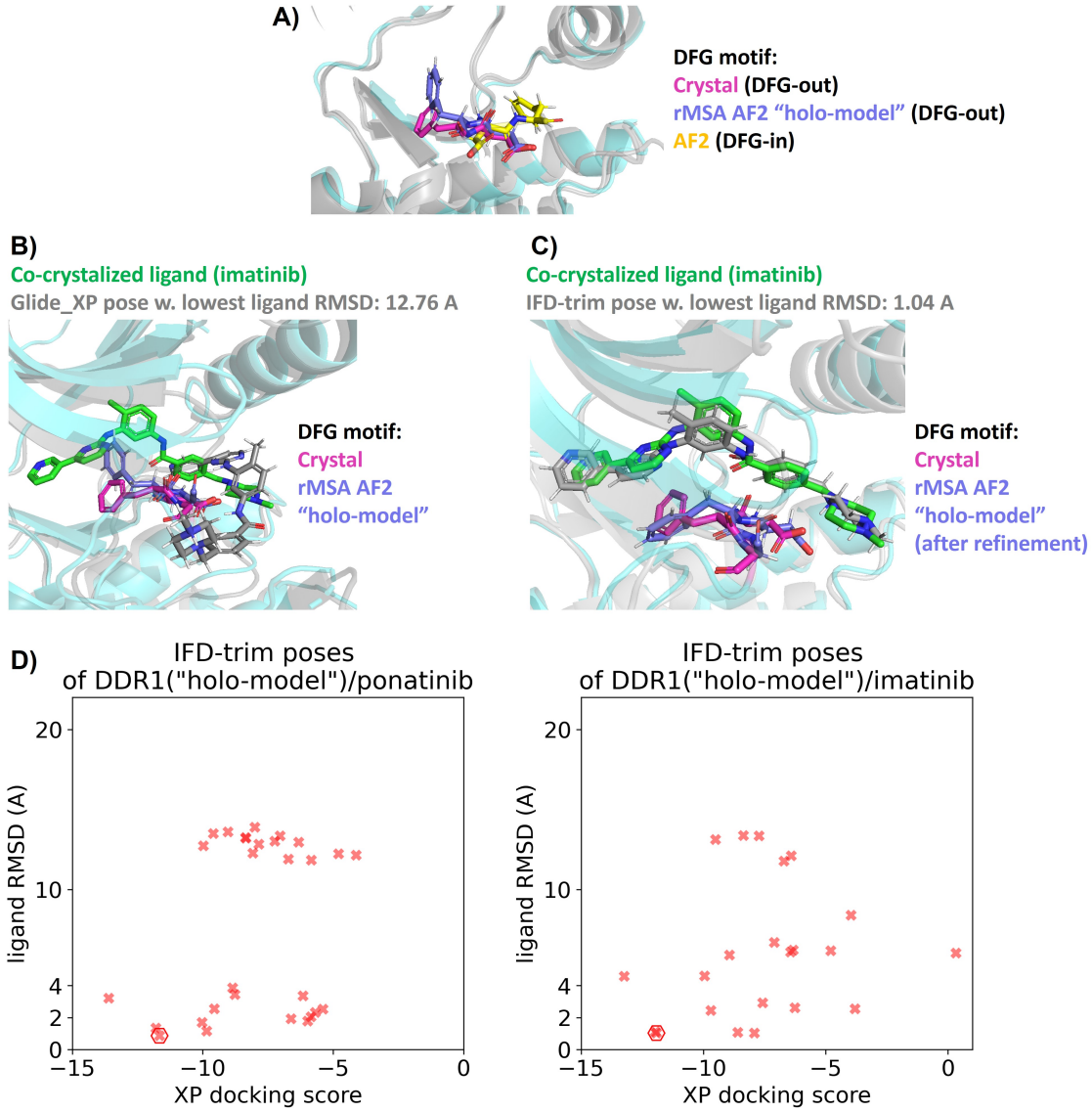


Figure E.6: A) Comparison of the DFG motif for DDR1 in its co-crystalized structure with imatinib (PDB 4BKJ), its "holo-model" structure and its AF2 structure. B&C) In the "holo-model" structure, the Phe residue in the DFG-motif requires rotation to prevent steric clashes with imatinib. proteins from crystal structure are shown as cyan cartoon, while all the other proteins are shown as grey cartoon. D) Ligand RMSDs are plotted against the docking scores for the IFD-trim docking poses of type II inhibitors (ponatinib and imatinib) against the "holo-model" structure in DDR1 rMSA AF2 ensembles. The pose with the lowest ligand RMSD is marked by hexagon.

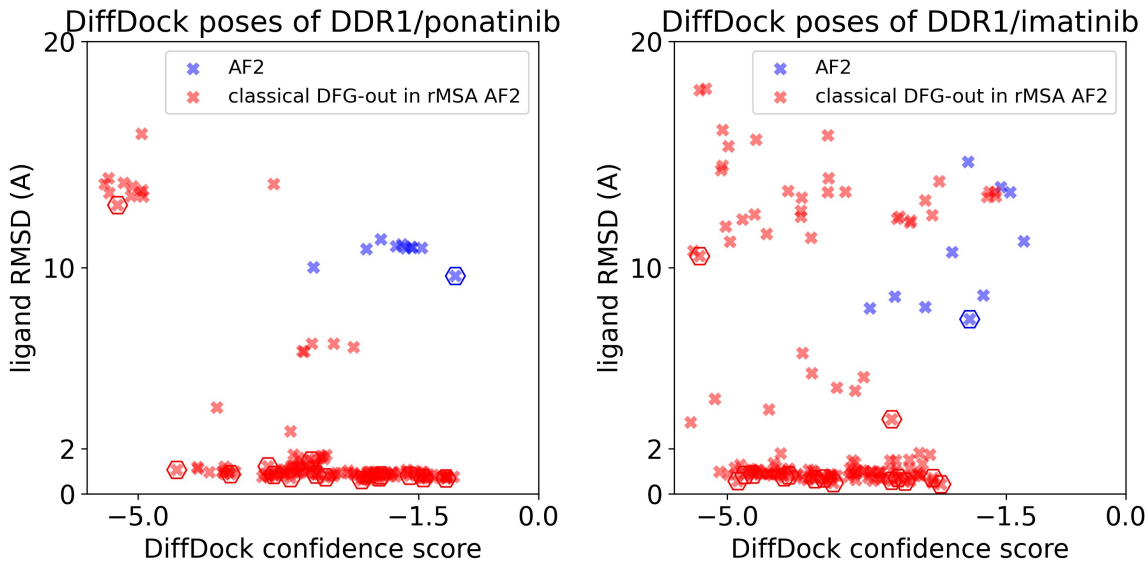


Figure E.7: Ligand RMSDs are plotted against the DiffDock confidence scores for the DiffDock poses of type II inhibitors (ponatinib and imatinib) against DDR1 AF2 structure (blue) or the classical DFG-out structures in DDR1 rMSA AF2 ensemble (red). The pose with the lowest ligand RMSD from each input structure is marked by hexagon.

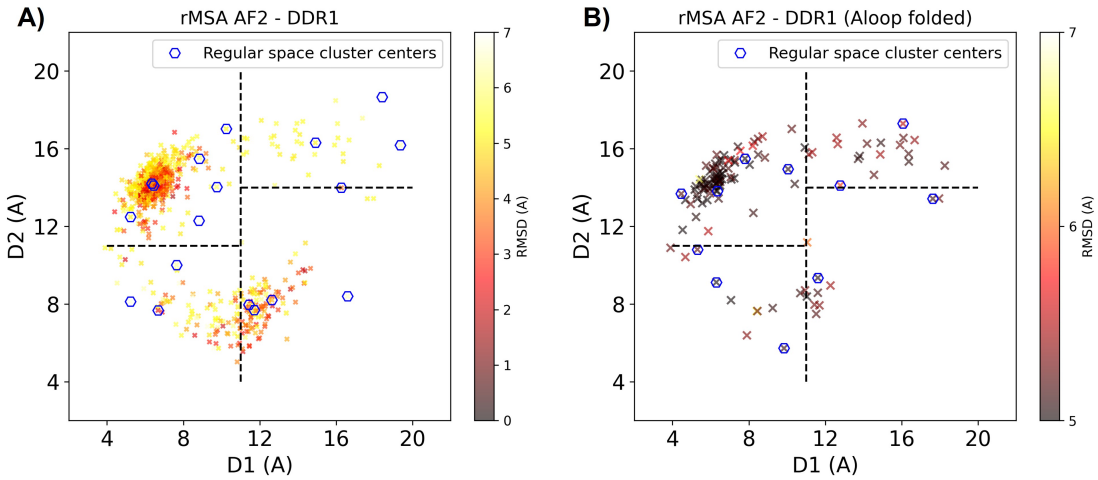


Figure E.8: A) the rMSA AF2 ensemble for DDR1 is projected in the Dunbrack space. Sample points are color-coded based on the CA RMSD from the AF2 structure with the highest pLDDT. Regular space cluster centers are marked by blue hexagons. For each DFG-type (in, inter or out), top two cluster centers with the lowest CA RMSD are selected as AF2RAVE initial structures. B) To take account of the underrepresented A-loop folded configurations, an extra regular space clustering is conducted only for the A-loop folded structures in the rMSA AF2 ensemble. The color code, notation and the way to select initial structures are the same as plot A. Combining AF2RAVE initial structures from both plot A&B, there are 12 initial structures in total.

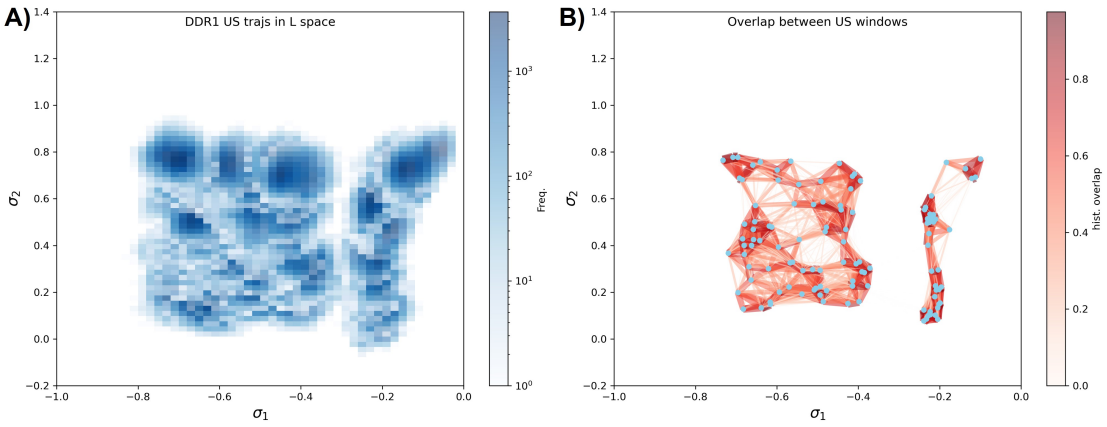


Figure E.9: A) Distributions from different umbrella sampling windows in the latent space. B) The distribution overlap graph for all the umbrella sampling windows. The mean value of each distribution is shown as blue dots. Each distribution's 2D histogram is flattened into 1D vectors, and the cosine similarity between two distributions is then indicated by the width and color of the edge connecting the respective dots. Windows from the A-loop folded region are not overlapped well with the windows from the A-loop extended region, while windows inside the A-loop folded region (the left part of the graph) are well connected and are used for the local PMF calculation in Figure 5.4 D.

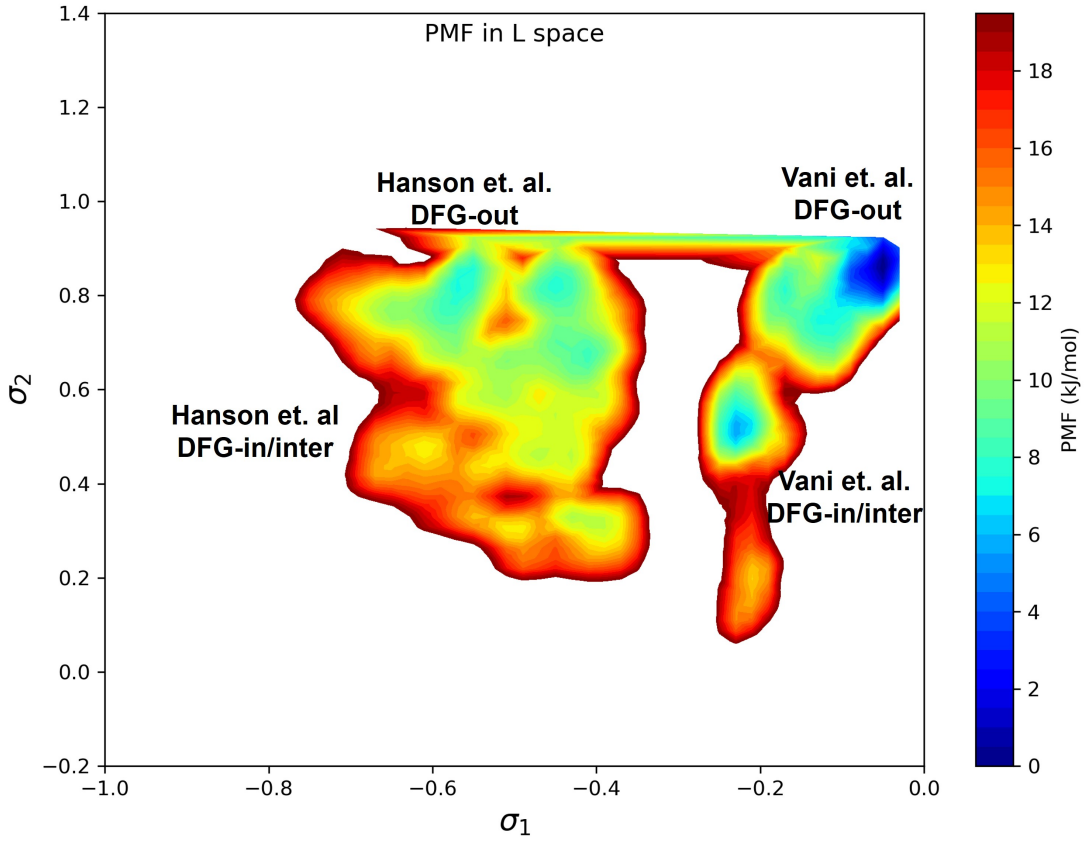


Figure E.10: DDR1 PMF calculated with all the umbrella sampling windows. Hanson et al.[4] found the A-loop folded DFG-out state to be more stable than the A-loop folded DFG-in/inter state for DDR1; Vani et al.[114] reported that the A-loop extended DFG-out state is more stable than the A-loop extended DFG-in/inter state for DDR1. Although our umbrella sampling setup is not sufficient to sample the A-loop movement, the observed relative stability corresponds with the findings of Hanson et al. and Vani et al.

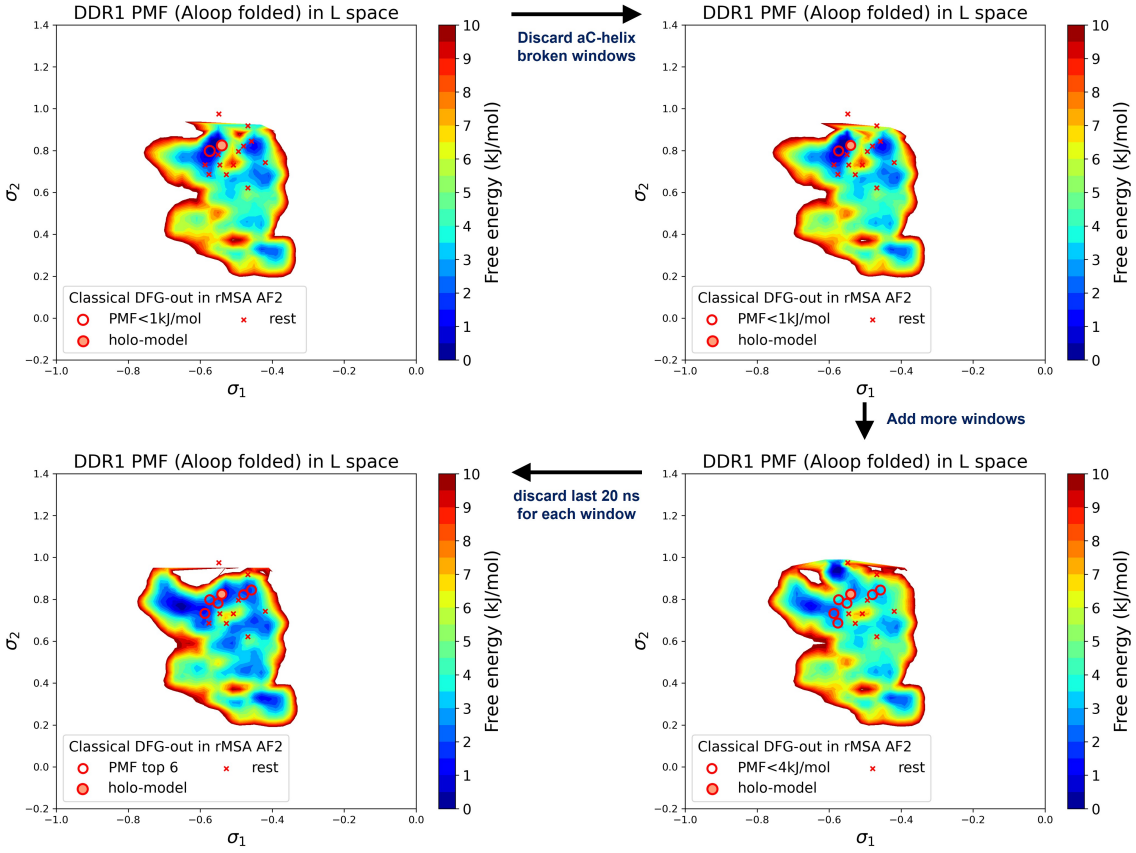


Figure E.11: PMF values and Boltzmann ranks of candidate structures fluctuate with the selection of the umbrella sampling windows and the simulation length of umbrella sampling trajectories, demonstrated with the DDR1 system.

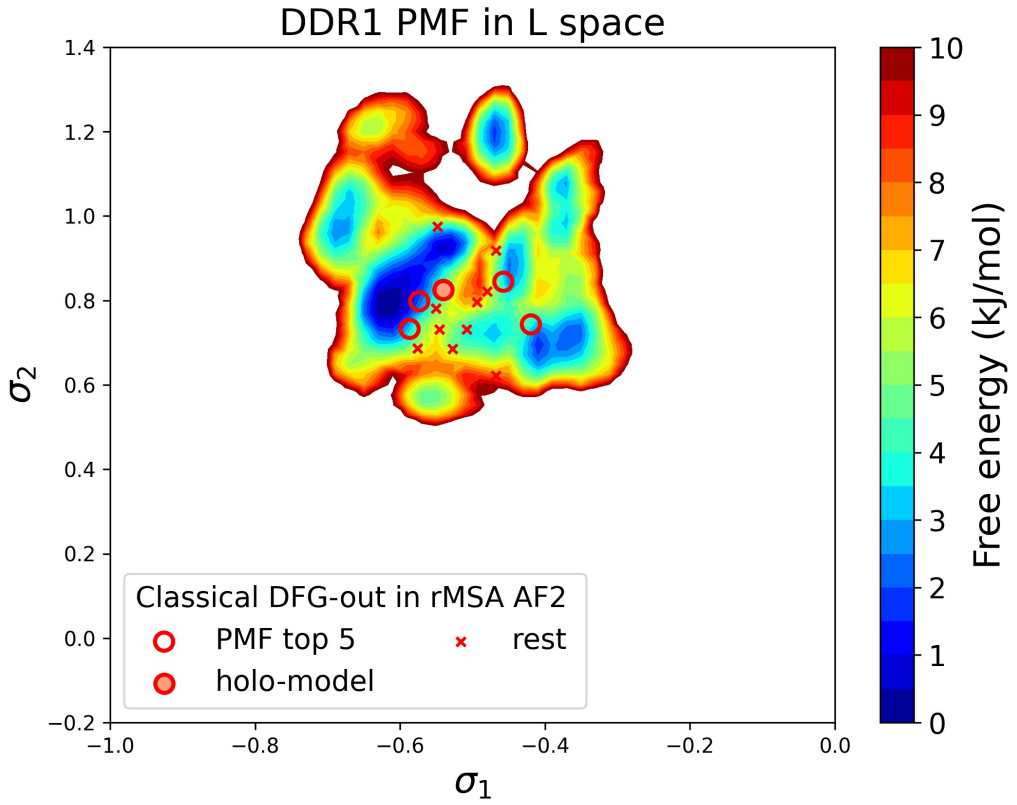


Figure E.12: Free energy profile for DDR1 in the latent space, calculated from unbiased MD simulations. The 15 DDR1 classical DFG-out structures in rMSA AF2 are shown as red cross and circles (top 5 structures ranked by free energy values). The "holo-model" structure is emphasized using a red circle filled with red.

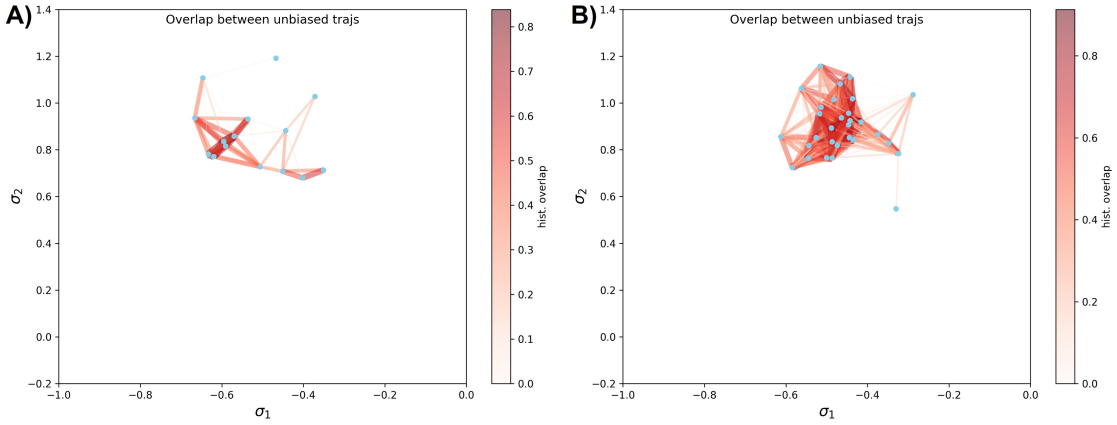


Figure E.13: A) The distribution overlap graph for all the unbiased MD trajectories starting from 15 classical DFG-out structures in DDR1 rMSA AF2 ensemble. B) The distribution overlap graph for all the unbiased MD trajectories starting from 30 Abl1 tAF2 structures in classical DFG-out state. The color-code is the same as Figure E.9.

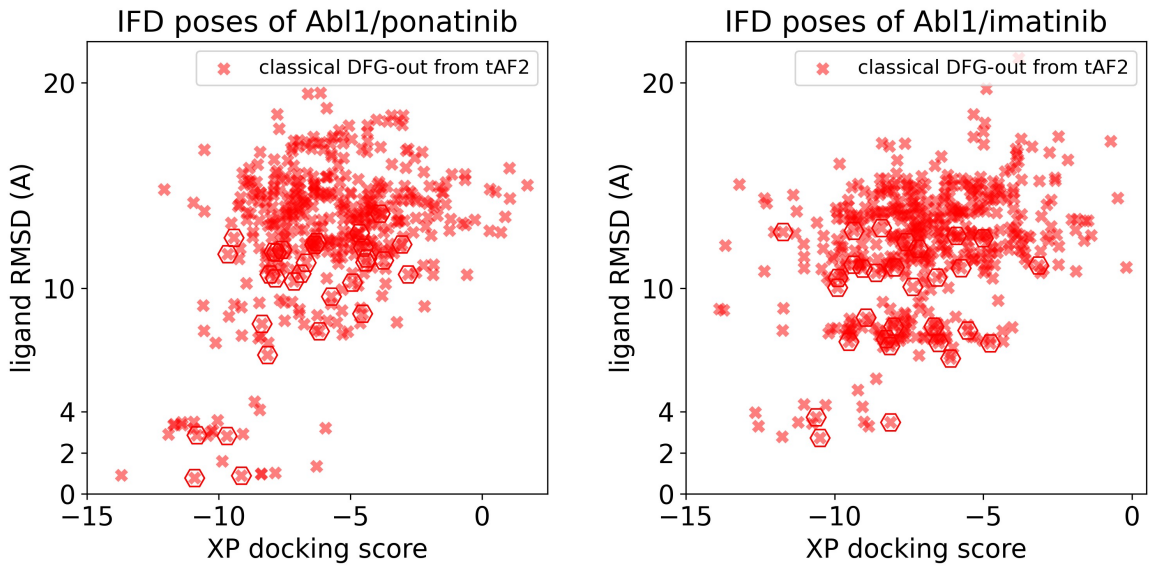


Figure E.14: Ligand RMSDs are plotted against the docking scores for the IFD docking poses of type II inhibitors (ponatinib and imatinib) against Abl1 tAF2 structures. The pose with the lowest ligand RMSD from each input structure is marked by hexagon.

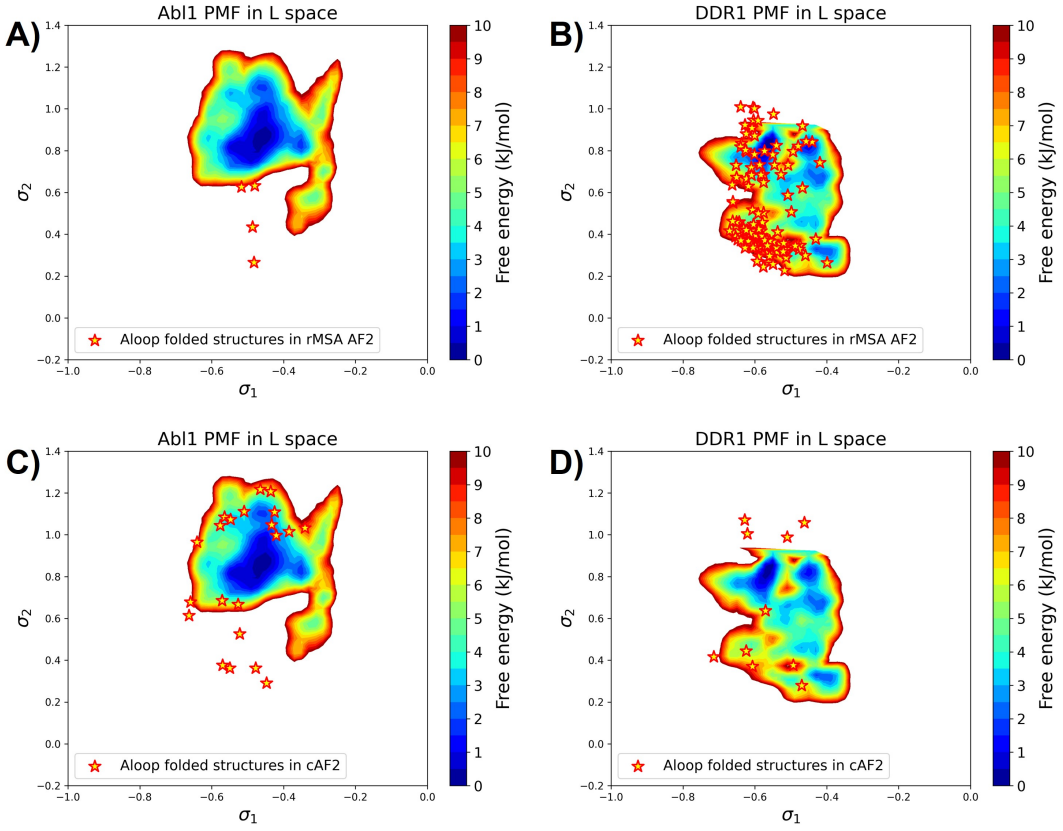


Figure E.15: The projection of A-loop folded structures from the rMSA AF2 ensemble or the AF2-cluster ensemble (cAF2) on the AF2RAVE PMF for Abl1 or DDR1.

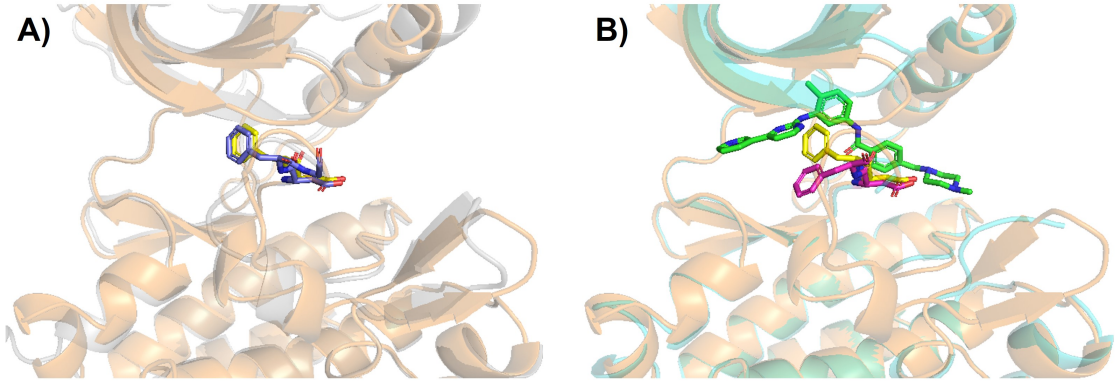


Figure E.16: A) the AF2-template structure for Src kinase is superimposed with its template structure (classical DFG-out in DDR1 rMSA AF2 ensemble, "holo-model"). The tAF2 structure of Src is shown as light-orange cartoon (protein) and yellow sticks (DFG motif), while DDR1 template is shown as light-gray cartoon (protein) and blue sticks (DFG motif). B) the AF2-template structure for Src kinase is again superimposed with Src/imatinib co-crystallized structure (PDB 2OIQ). Crystal structure is shown as light-cyan cartoon (protein), green sticks (ligand) and magenta sticks (DFG motif).

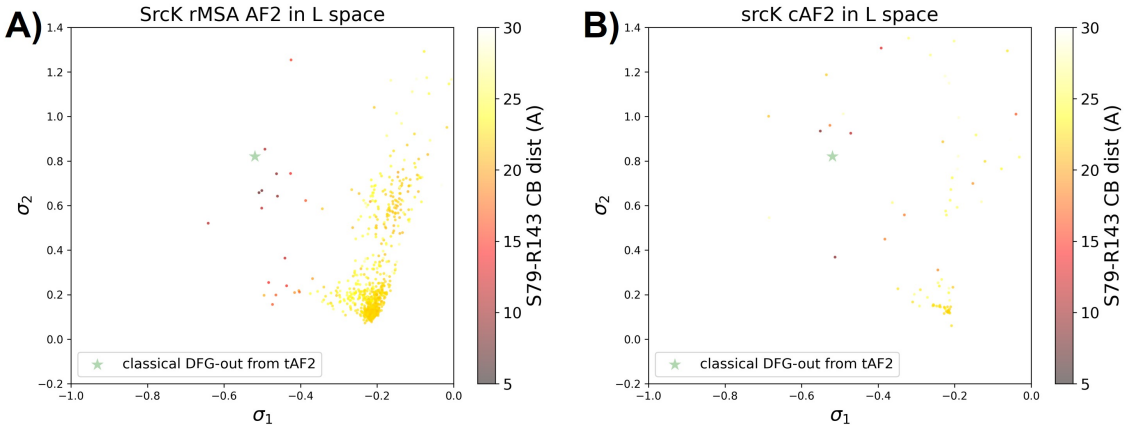


Figure E.17: The projection of A) the rMSA AF2 ensemble or B) the AF2-cluster ensemble on the AF2RAVE latent space for SrcK. The classical DFG-out SrcK structure generated from AF2-template in Figure E.16 is shown as the green star. The color-code shows the A-loop location.

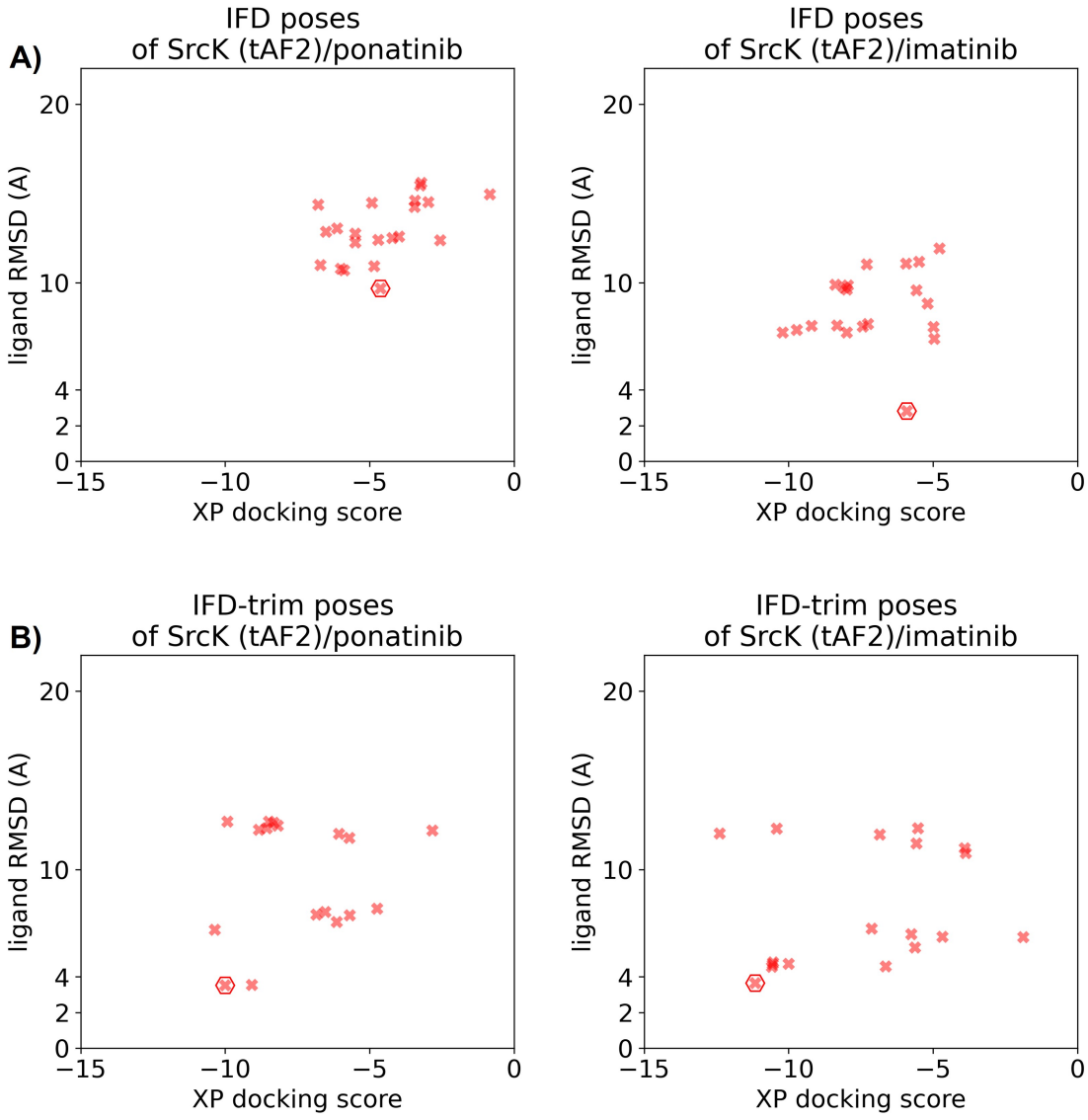


Figure E.18: Ligand RMSDs are plotted against the docking scores for the IFD/IFD-trim docking poses of type II inhibitors (ponatinib and imatinib) against the SrcK tAF2 structure. The pose with the lowest ligand RMSD from each input structure is marked by hexagon.

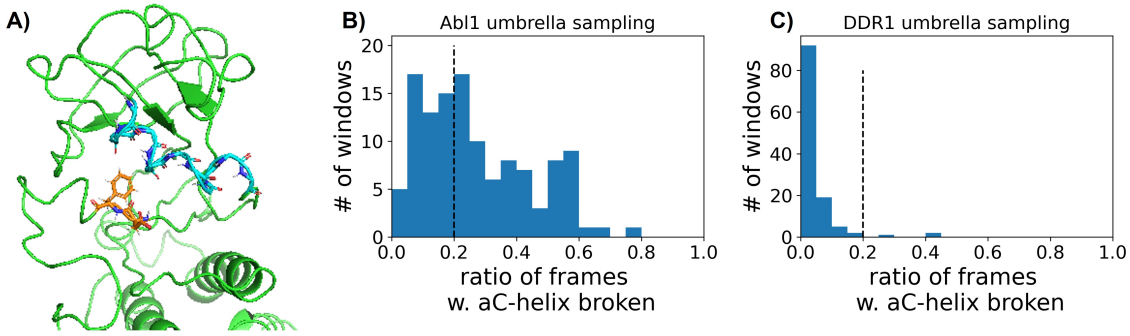


Figure E.19: A) one representative frame with αC helix broken in Abl1 umbrella sampling trajectories. The backbone of the αC helix is shown with cyan sticks, while the DFG motif is shown as orange sticks. B) or C) The distribution of the ratios of frames with αC helix broken in each umbrella sampling window for Abl1 or DDR1.

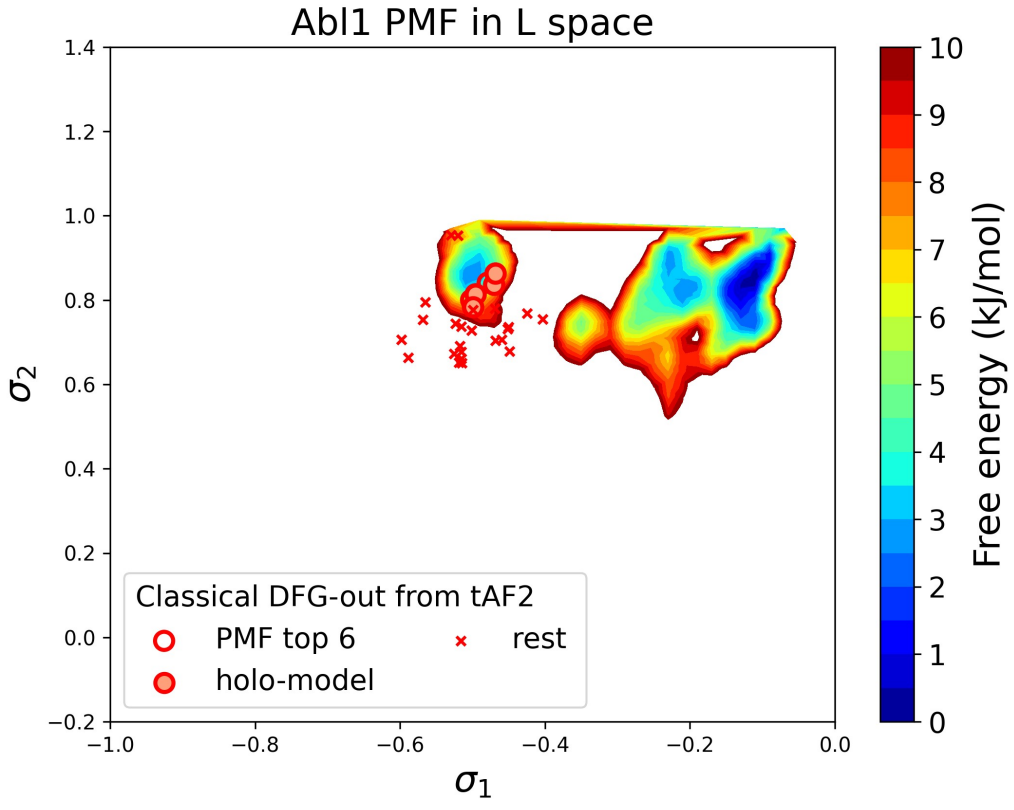


Figure E.20: Abl1 PMF calculated from umbrella sampling after discarding windows with αC helix broken. The four holo-like structures ("holo-models") are enriched to the top six based on PMF values.

Appendix F: Supporting Information for Chapter 6

F.1 MD simulations

F.1.1 Unbiased MD simulations

F.1.1.1 Short MD for AF2RAVE

All the systems were neutralized with Na^+ ions or Cl^- ions and parameterized with the Amber99SB*-ILDN force field [343, 344] with multivalent cation potential optimized by experimental ion-oxygen distance (IOD) [347] and the TIP3P water model[345]. Particle Mesh Ewald[333] was used for calculating electrostatics and the lengths of bonds to hydrogen atoms were constrained using LINCS[332] throughout all simulations. Simulations were performed at 300 K with the LangevinMiddleIntegrator[346] in OpenMM[341] with the step size of 2 fs. Energy minimization was conducted for each initial structure, followed by equilibration runs under NVT and NPT for 500 ps and 1 ns respectively. Production simulations were then carried out under the NPT ensemble for 100 ns. To prevent Ca^{2+} ions from dissociating from EF-hands, truncated harmonic constraints between Ca^{2+} ions and coordinating oxygen atoms were applied during equilibration runs and production runs, as shown in Fig. F.5C. Coordinating oxygen atoms were identified based on their pairwise distances with calcium ions across NMR structure ensembles, which consistently remain below 5 Å for both Ca^{2+}-S100A1 (PDB 1ZFS) and Ca^{2+}-S100B (PDB 1MWN), as shown in Fig. F.5A-B. For pseudo EF-hands, four coordinate oxygen atoms are identified: SER19-O, LYS27-O, GLU32-OE1, and GLU32-OE2. For typical EF-hands, five coordinate oxygen atoms are identified: ASP62-OD1, ASP66-OD2, GLU68-O, GLU73-OE1, and GLU73-OE2. For each production run, restrained calcium-oxygen pairs with average distances over 5 Å were considered as unstable coordinate bonds, and production trajectories with more than two such unstable coordinate bonds were discarded, as shown in Fig. F.5D, leaving 9 production runs for Ca^{2+}-S100A1 and 12 for Ca^{2+}-S100B for further SPIB analysis.

F.1.1.2 Long 1 μs MD from another study

In an unpublished study, we performed 1 μs long MD simulations for six Ca^{2+}-S100 systems to investigate inhibitor selectivity. Here, we analyzed the fluctuations from those long-timescale simulations to identify rigid domains within the Ca^{2+}-S100 proteins.

To set up the initial structures for MD simulations, XRD structures were used for the S100B/SBi5369 and S100B/SBi5345 complexes. Since no XRD structures are available for the S100A1/SBi5369 and S100A1/SBi5345 complexes, homology modeling was performed using AlphaFold2 with the S100B/SBi5345 XRD structure as a template. For the compound-free

S100B and S100A1 systems, the compounds were simply removed from the S100B/SBi5369 or S100A1/SBi5369 structures mentioned above, respectively. We employed the CHARMM36m force field [348, 349], with ligand parameters generated using the CHARMM General Force Field (CGenFF). No restraints were applied to the Ca^{2+} ions. All initial structures were prepared using CHARMM-GUI [350, 351], which included solvation with a water box extending 10 Å from the protein surface, neutralization with Na^+ and Cl^- ions, and adjustment to a physiological salt concentration of 0.15 M. The protonation states of all residues were maintained as assigned in the original XRD structures. Simulations were conducted using the OpenMM platform, with default input scripts provided by CHARMM-GUI. Prior to production runs, all systems underwent energy minimization, followed by a 150 ps equilibration in the NVT ensemble using a 1 fs time step. Production simulations were then performed in the NPT ensemble for 1 μs using a 2 fs time step.

To identify rigid domains and global features, we computed structural variance across the entire amino acid sequence using a sliding window of 12 residues. The structural variance was quantified using the order parameter Q[352], with the Ca^{2+}-S100B/SBi5369 XRD structure serving as the reference:

$$Q_{ref} = \frac{1}{N} \sum_{|i-j|>2}^{N} e^{-\frac{(r_{ij}-r_{ij}^{ref})^2}{2\sigma^2}},$$

where $\sigma = (1 + |i - j|)^{0.15}$Å r_{ij} and r_{ij}^{ref} are C_α atoms distances between residue i,j in the MD frames and reference structure, respectively. The Q value ranges from 0 to 1, with higher values indicating greater structural similarity. Q values were averaged over all six long MD trajectories for each 12-residue peptide segment. Adjacent segments with average Q values greater than 0.9 were merged to define rigid regions. Based on this analysis, we identified four rigid helices and therefore eight terminal anchor residues per monomer, as illustrated in Fig. F.3.

F.1.2 Restrained MD simulations

For the restrained MD simulation, the simulation and initial structure setup follow the same protocol as the unbiased MD section, except for the modifications detailed below. For both the equilibration run and the production run, an additional constraint potential is applied in the following form: $V_{template} = k_{template} \sum_{ij} \sum_{temp} e^{-\frac{(r_{ij}-r_{ij}^{temp})^2}{2\sigma^2}}$, where $k_{template}$ is the restraint strength, set to 100 kJ/mol, and $\sigma = 2$ Å, r_{ij} are CVs selected based on their contributions to the SPIB latent variable, specifically those with absolute weights exceeding 50% of the maximum value. The reference CV values, r_{ij}^{temp} are corresponding CV values from the top 10 lowest free energy structures among the 20 MD template structures. To preserve the crystal-like characteristic in tAF2 configuration for rotamers with fast dynamics, the simulation is performed at 200 K with a short 5 ns production run. The final frames from these production runs are compiled into the AF2RAVE structure set.

F.2 SPIB setup

The State Predictive Information Bottleneck (SPIB) model is a variational autoencoder architecture that takes high-dimensional features $\{X_t\}$ from MD trajectories as encoder input and predicts the future state label $\{y_{t+\Delta t}\}$ at a time lag Δt as decoder output. The low-dimensional latent space $\{z\}$ serves as an information bottleneck, compressing the high-dimensional input while retaining sufficient information to construct metastability in the latent space. To achieve this, the SPIB objective function maximizes the mutual information between the latent space and future state labels while minimizing the mutual information between the input and latent space: $argmax[\mathcal{L} = I(z; y_{t+\Delta t}) - \beta I(X_t; z)]$, where β is a positive hyperparameter that balances regularization ($I(X_t; z)$) and reconstruction ($I(z; y_{t+\Delta t})$). This encourages the latent space to approximate reaction coordinates for transitions between long-lived metastable states—those with residence times longer than Δt.

During training, the SPIB model iteratively refines the state labels until the state density converges. After training, features that fluctuate much faster than the chosen time lag Δt are assigned lower weights by the encoder, preventing them from influencing state assignments. As a result, most frame pairs separated by Δt are assigned to the same state, effectively filtering out fast intra-state fluctuations. Therefore, the choice of both the time lag Δt and the input features should be aligned with the timescale and structural regions relevant to the dynamics of interest.

F.2.1 Global SPIB

For the 28 global features previously used in regular space clustering, we calculated their standard deviation after concatenating all 21 unbiased MD trajectories for both Ca^{2+}-S100 systems. To refine the feature set, we retained 14 of them with standard deviations larger than 0.25 of the maximum standard deviation as input collective variables (CVs) for global SPIB. For each S100 isoform (Ca^{2+}-S100A1 or Ca^{2+}-S100B), we first performed TICA with a time lag of 5 ns, followed by k-means clustering with 20 centers to generate initial state labels for SPIB. We then conducted a hyperparameter sweep for the SPIB time lag, ranging from 5 ns to 50 ns in 5 ns intervals, and observed that the SPIB latent space remained largely consistent across different time lags. Based on this, we selected a representative latent space trained with a 25 ns time lag on the Ca^{2+}-S100B system, as shown in Fig. F.6. Since the input global features are transferable among Ca^{2+}-S100A1 and Ca^{2+}-S100B, the global SPIB latent space is also transferable.

F.2.2 Local/pocket SPIB

Given trajectories within a certain backbone mode, we trained the SPIB model based on the 53 local features. Unlike backbone motions, sidechain rotamers usually fluctuate on much faster timescales. To capture various metastable packing modes of pocket residue sidechains given a certain backbone mode, we first performed TICA with a time lag of 50 ps, followed by k-means clustering with 20 centers for initial state labeling in the SPIB model. We then conducted a hyper parameter sweep of the SPIB time lag, ranging from 50 ps to 5 ns in 250 ps intervals, and observed that the latent space remained largely stable across time lags ranging from 50 ps to 800 ps. Based on this, we selected a representative latent space trained with a 300 ps time lag, as shown in Fig. F.8.

F.3 Dataset for ground truth holo structures of Ca^{2+}-S100A1 and Ca^{2+}-S100B

XRD-PDB24 structure set for Ca^{2+}-S100B: 24 XRD holo structures from PDB: 3RM1, 5CSJ, 5CSI, 4PE0, 3LLE, 4FQO, 5DKR, 5CSF, 4PE4, 5ER4, 5CSN, 5DKQ, 3CR4, 4PDZ, 3IQQ, 3HCM, 5DKN, 4PE7, 3LK0, 3GK4, 4PE1, 4XYN, 5D7F, 5ER5.[174, 177, 353, 354, 355, 356, 357, 358, 359, 360, 361]

XRD-SBi5345/XRD-SBi5369: 2 unpublished XRD structure with experimentally designed inhibitors SBi5345 and SBi5369 bound with Ca^{2+}-S100B.

For Ca^{2+}-S100A1, no XRD holo structures with compound bound is reported in PDB, as far as we know. However, there are two NMR structures with peptide bound with Ca^{2+}-S100A1: 2K2F, 2KBM.[362]

F.4 Docking setup

F.4.1 Retrospective docking with Glide XP

Retrospective docking was investigated on two inhibitors, SBi5345 and SBi5369. We used Ligprep to prepare the ligand inputs from SMILES files, Protein Preparation Wizard to prepare the protein structures, and Glide XP docking to generate docking poses, in the Schrödinger Maestro software[363].

For Glide XP docking, the inner and outer box sizes were set to 15 Å and 35 Å respectively, with the center of each box defined as the centroid of the binding-site residues. Two binding-site setups were used for statistical analysis: one broadly defined as the C-terminal half (residues 41–85, including the hinge region and the typical EF-hand) of the entire protein single chain, and a more precisely defined binding site that includes all pocket residues involved in deriving the 53 local features. To soften the potential and reduce penalties for close contacts, a van der Waals radii scaling factor of 0.50 was applied to both ligands and receptors. A maximum of 20 poses per ligand was generated for each receptor box, and the resulting docking poses were compiled for each structure set. The XRD structures of the Ca^{2+}-S100B/inhibitor dimer are not perfectly symmetric, resulting in two reference ligand binding poses per XRD structure. For each docking pose, the ligand RMSD is calculated against both reference poses following C_α alignment of residues 41–85, with the lowest RMSD value reported.

Since ligand RMSD does not capture the atomic environment of proteins, it can remain low even when protein sidechain rotamers are incorrect and protein-ligand interactions are disrupted. To address this, we also computed Protein-Ligand Interaction Fingerprint (PLIF) bitstrings using PyPLIF-Hippos[364]. According to its documentation, residues within 5 Å of the holo-ligand in the XRD structures were probed to generate the bitstrings. To compare docking poses with reference XRD holo poses, we used the Tanimoto coefficient (Tc) and McConnaughey coefficient (Mcc) as similarity metrics for the PLIF bitstrings. However, since PLIF bitstrings do not account for the atomic environment of ligands, we determined that PLIF similarity coefficients provide a meaningful assessment of binding pose similarity only for docking poses with sufficiently low ligand RMSD (below 9 Å in this case).

F.4.1.1 DOCK6.11 docking setup

All structures used in the enrichment test shared nearly identical backbone conformations and were aligned to chain A of the X-ray crystal structure of Ca^{2+}-S100B/SBi5369 for reference. The docking region was selected using SPHGEN provided by DOCK. All spheres within 5.5 Å of the known ligand were used and any spheres outside the known binding site were excluded manually. The search volume for docking was constructed to be the box enclosing the selected spheres with a small margin of 2.0 Å using SHOWBOX, and the precomputed docking grid was constructed using GRID with a grid spacing of 0.2 Å. SPHGEN, SHOWBOX and GRID programs are available at http://dock.compbio.ucsf.edu accompanying the standard DOCK 6.11 distribution. While the exact parameters are quite system-dependent this protocol for setup is fairly standard[365].

F.4.1.2 DOCK6.11 pocket preparation

S100B crystallizes as a dimer, with both monomers potentially contributing to ligand binding. For all docking evaluations presented here, both potential binding sites were considered, and the final docking score for each structure was computed as the average across the two sites. The combined BEDROC profiles are then constructed by ranking ligands accordingly.
Additionally, in the crystal structure of SBi5369/Ca^{2+}-S100B dimer, the FF-gate region was not resolved in one of the chains, and corresponding pocket was excluded from the enrichment analysis due to poor performance, as shown in Fig. F.15f.

F.5 Detailed Docking and Enrichment Protocols

F.5.1 Docking validation

Retrospective docking with Glide XP[126, 127, 128] using the Schrödinger Maestro software[363] and enrichment docking using DOCK 6.11[159, 160] were performed to evaluate the effectiveness of the AF2-RAVE structure set for rigid docking. Details of the docking setups are provided in the SI. Glide XP excels in reproducing ligand orientations and scoring accuracy; however, its high computational cost limits its applicability for high-throughput virtual screening of large compound libraries. In contrast, DOCK is specifically designed for efficient screening of ultra-large ligand databases and is used for studying enrichment.

Three structural sets were compared across all docking experiments, including XRD: X-ray holo structures of Ca^{2+}-S100B bound to newly designed Ca^{2+}-S100B inhibitors SBi5345 and SBi5369, with holo ligands removed; AF2: AlphaFold2 reference structure of S100B with calcium ions added to EF-hands; AF2-RAVE: AF2-RAVE structure set generated from the Ca^{2+}-S100B native global basin. Since all structures are dimers, each was treated as two binding pockets in the docking experiments.

F.5.1.1 Ligand library selection and preparation for enrichment testing

Experimental data identified approximately 85 ligands with measurable binding affinity to S100B, which I considered as true inhibitors. However, many of these ligands are structurally

similar, as determined by Tanimoto similarity[366] using 2048-bit fingerprints generated with RDKit[367]. To reduce redundancy, I selected a random subset of 10 ligands, with all pairwise similarities below 0.8. The similarity distributions before and after filtering are shown in Fig. F.13. Although some similarity remains among the selected ligands, my primary objective is to assess the quality and diversity of the protein structures. Any bias introduced by ligand similarity is expected to affect all structures equally.

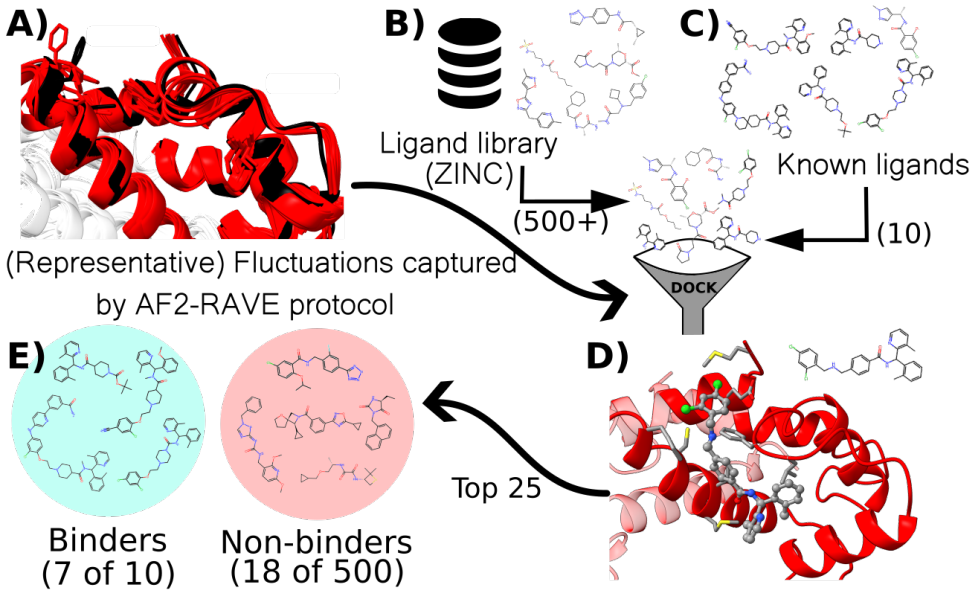


Figure F.1: Pipeline for enrichment test using DOCK. Known inhibitors are diluted with a large set of property-matched random molecules in a ratio $>$ 50:1 (50 decoys per true inhibitor). A) AF2-RAVE allows capturing key side-chain fluctuations needed for DOCK to make the required interactions. B) Ligands from the ZINC database are used as decoys. C) 10 known inhibitors are used as true actives D) DOCK is used to score these ligands to the same pocket from diverse conformations of each representative structure E) This test evaluates how effectively DOCK scoring can distinguish true actives from decoys for a given structure. The AF2-RAVE ensemble demonstrates strong performance, achieving approximately 14-fold enrichment over random selection.

The objective of this analysis is to assess whether DOCK can effectively distinguish true S100B inhibitors from decoy molecules, and to evaluate the advantage of structural diversity in AF2-RAVE compared to using only AlphaFold2 or crystal structures. The molecular weights of true actives have a range of 325–728 a.m.u., while ligands in ZINC database are typically smaller (125–450 a.m.u.). To account for this, I set a lower molecular mass threshold of 325 a.m.u.—corresponding to the smallest known actives—and divided the 325–425 a.m.u. range into 10 bins. From each bin, 120 neutral ligands were selected. The final set of 596 decoys was randomly sampled from this pool, resulting in a total of 606 ligands including known binders and decoys. To reflect the hydrophobic nature of the pocket and the neutrality of the known actives, all decoys were also chosen to be uncharged, giving me ligands with properties, including charges and molecular weights, matched to true actives. Initial and final molecular weight distributions are shown in Fig. F.14. All 606 ligands were then scored against each protein structure using DOCK, and enrichment performance was evaluated.

F.5.1.2 Enrichment factor calculation

The primary objective of docking in this context is early enrichment—that is, ranking true actives before the decoys. In addition to direct enrichment plots, which show the number of decoys encountered before each true active is identified, I also report the standard BEDROC (Boltzmann-Enhanced Discrimination of the Receiver Operating Characteristic) profile[368]. BEDROC is defined as a normalized version of the Robust Initial Enhancement (RIE)[369] score:

$$BEDROC(\alpha) := \frac{RIE(\alpha) - RIE_{min}(\alpha)}{RIE_{max}(\alpha) - RIE_{min}(\alpha)} \text{ with } RIE(\alpha) \sim \sum_{i=1}^{n} e^{-\alpha r_i}$$

Here, the summation runs over all true actives, and r_i represents the relative rank of the i-th true active, defined as $r_i = R/N$, where R is its absolute rank and N is the total number of ligands. BEDROC emphasizes early enrichment by exponentially weighting ranks using a factor α, which serves as a hyperparameter. Larger values of α place greater importance on identifying true actives early in the ranked list. A commonly used and recommended value in the literature is $\alpha = 20$[368, 370]. I apply the BEDROC metric using the ranks of the known actives and report the BEDROC score profile across varying values of α.

Supporting Figures

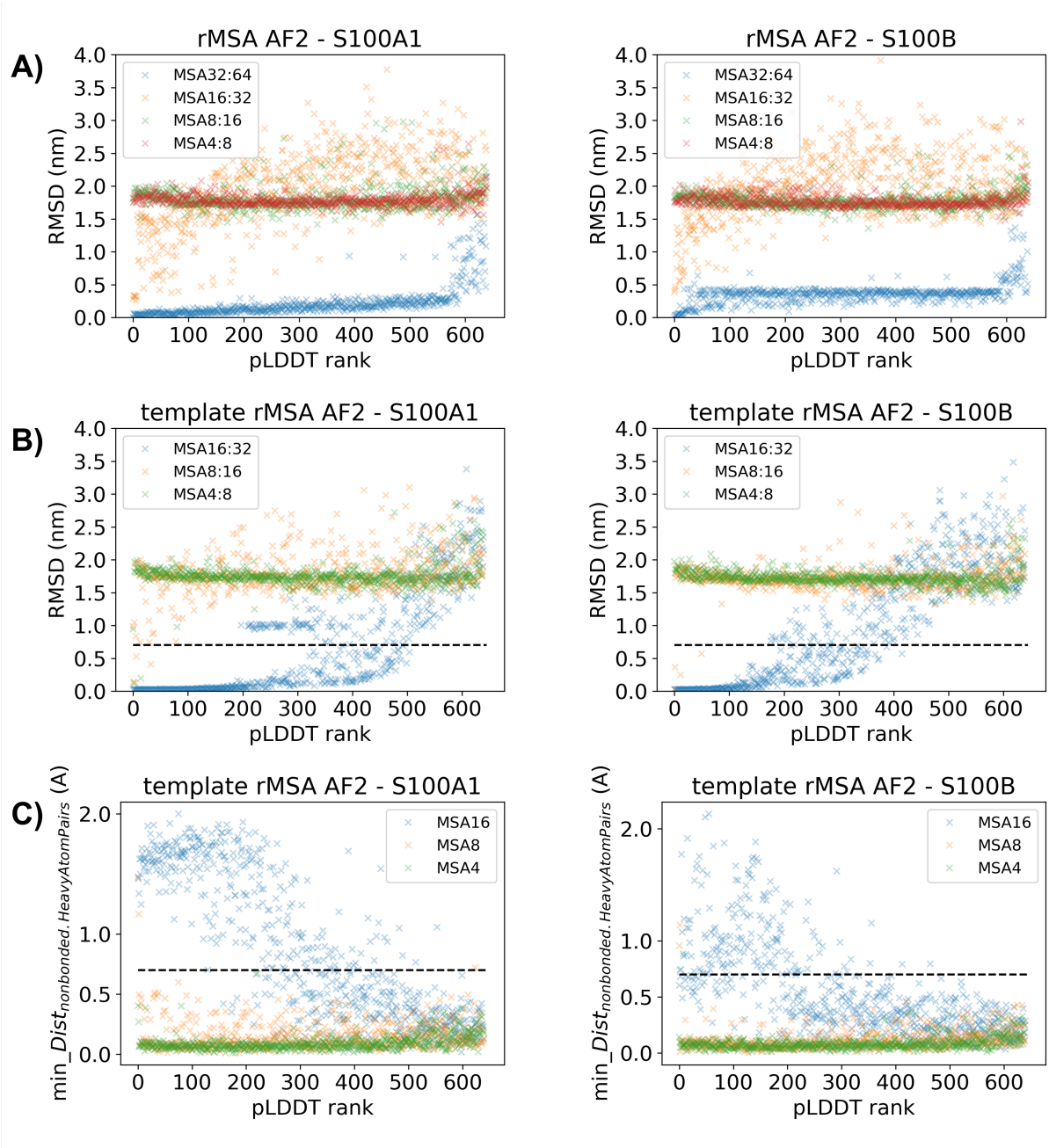


Figure F.2: Structural quality assessment of (template) rMSA AF2 ensembles for S100A1 (left) and S100B (right) dimers. A) Backbone C_α RMSD values of reduced MSA AF2 ensembles generated with varying MSA depths, evaluated against the corresponding AF2 reference structure. B) Backbone C_α RMSD values of template reduced MSA AF2 ensembles, using the corresponding AF2 reference structure as a template. A 7 Å RMSD filter (retaining structure below black dashed line) is applied. C) Minimum distances for non-bonded heavy atom pairs in each structure from the template reduced MSA AF2 ensembles. A 0.7 Å cutoff (retaining structure above black dashed line) is applied to eliminate structures with steric clashes. The x-axes correspond to the pLDDT ranking within each ensemble.

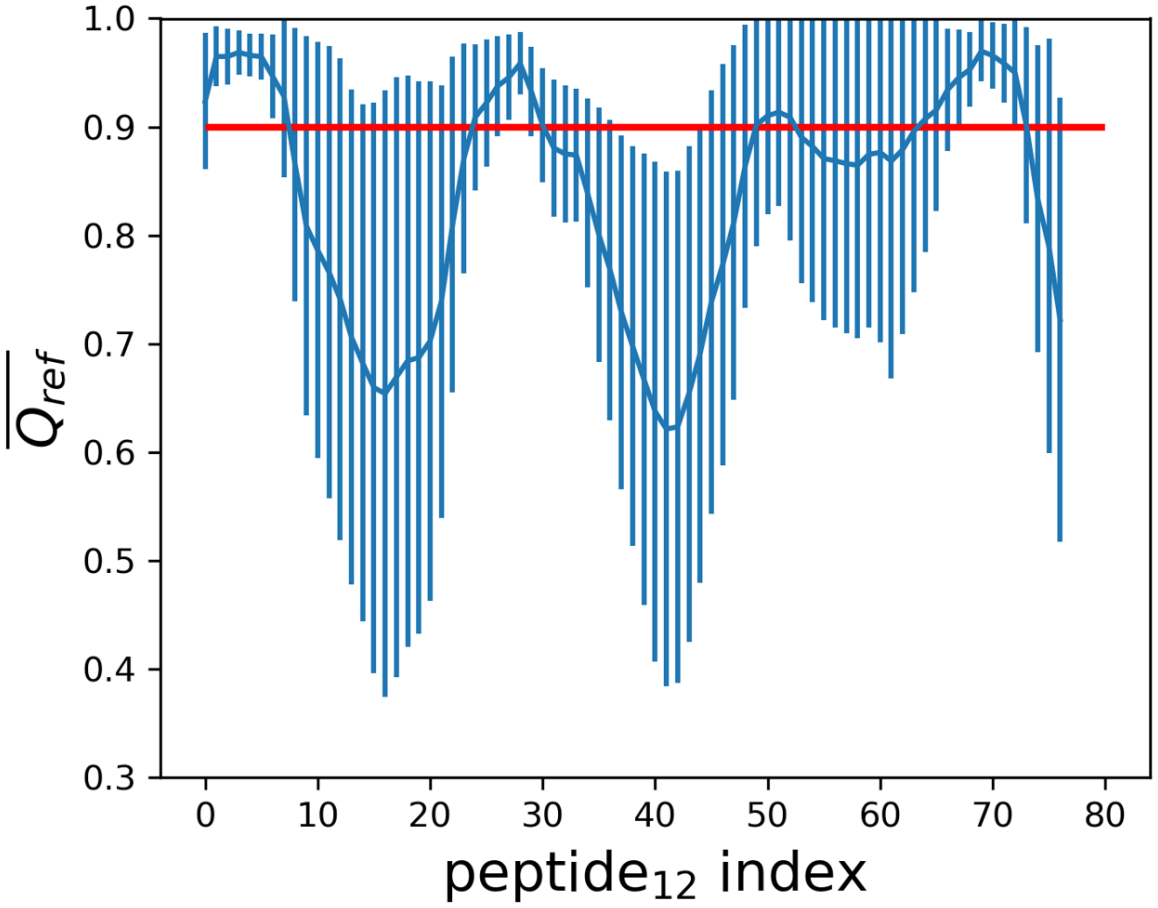


Figure F.3: Structural variability in long MD along the full amino acid sequence of S100 proteins, using a sliding window of 12 residues. The mean and variance of Q_{ref} values are represented by blue lines and error bars, respectively. The horizontal red line indicates the cutoff threshold set at 0.9.

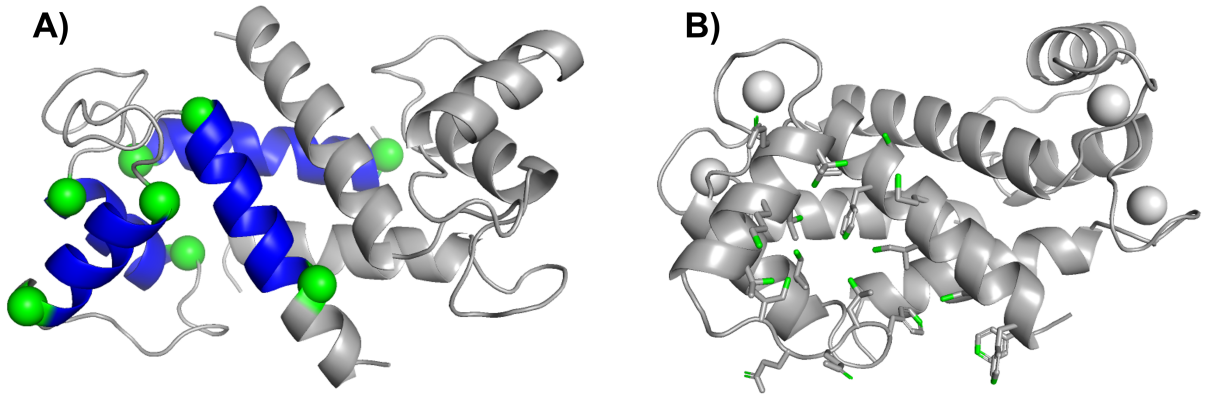


Figure F.4: A) The four rigid α-helices within a single chain of the Ca^{2+}-S100 dimer are shown in blue. The terminal C_α atoms of each helix, used as anchors for global feature construction, are highlighted as green spheres. Pairwise distances between these anchor atoms define the global features. Calcium ions are omitted for visual clarity. B) Side chains of pocket residues are displayed as sticks, with their distal atoms colored in green. Pocket features are defined by the pairwise distances between these distal atoms.

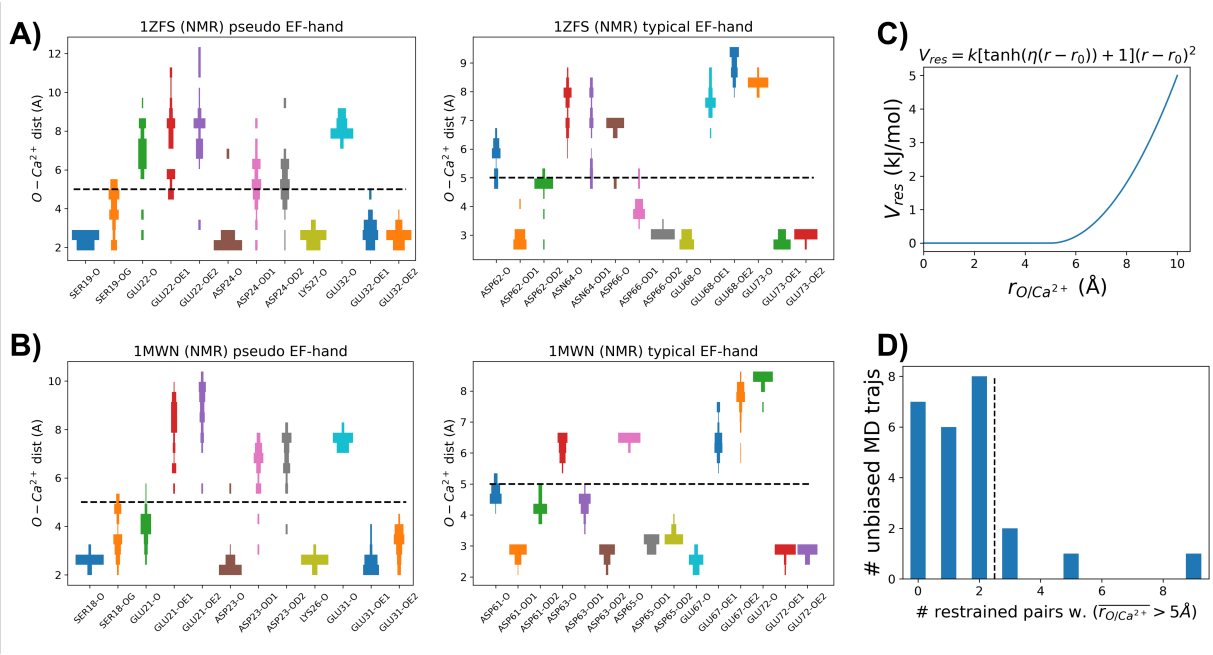


Figure F.5: A-B) Distributions of distances between each oxygen atom in EF-hand residues and the corresponding Ca^{2+} ions from NMR structures of Ca^{2+}-S100A1 (PDB: 1ZFS, subplot A) and Ca^{2+}-S100B (PDB: 1MWN, subplot B). The N-terminal pseudo EF-hands are shown on the left and the C-terminal typical EF-hands on the right of each panel. Those oxygen–Ca^{2+} pairs with distance distributions entirely below the black dashed lines (marking the 5 Å threshold) were restrained in all the MD simulations. C) The format of the restraint potential applied to selected oxygen–Ca^{2+} pairs. D) The distribution of the number of unstable oxygen-Ca^{2+} coordinate bonds (defined as average distances >5 Å, despite restrained by potential in subplot C.) across 25 unbiased MD trajectories (10 for Ca^{2+}-S100A1 and 15 for Ca^{2+}-S100B). Only trajectories with two or fewer unstable coordinate bonds (to the left of the black dashed line) were retained for subsequent SPIB analysis.

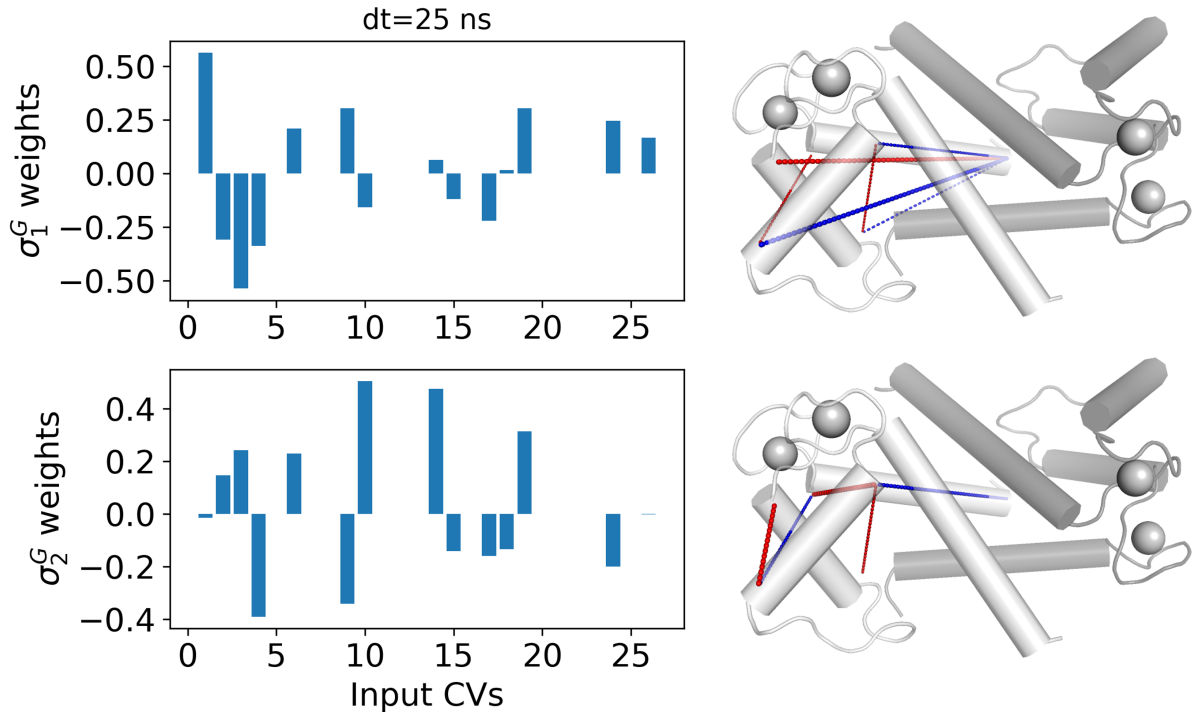


Figure F.6: Bar plots depict the linear weights assigned to each of the 28 global features in the learned global SPIB coordinates, using a time lag of 25 ns on the Ca^{2+}-S100B system. The right panel highlights the dominant features, defined as those with weights exceeding half of the maximum, contributing to each coordinate. The upper and lower panels correspond to σ_1^G and σ_2^G, respectively. The color and width of the dashed lines reflect the sign and absolute magnitude of the weights in the learned linear combinations.

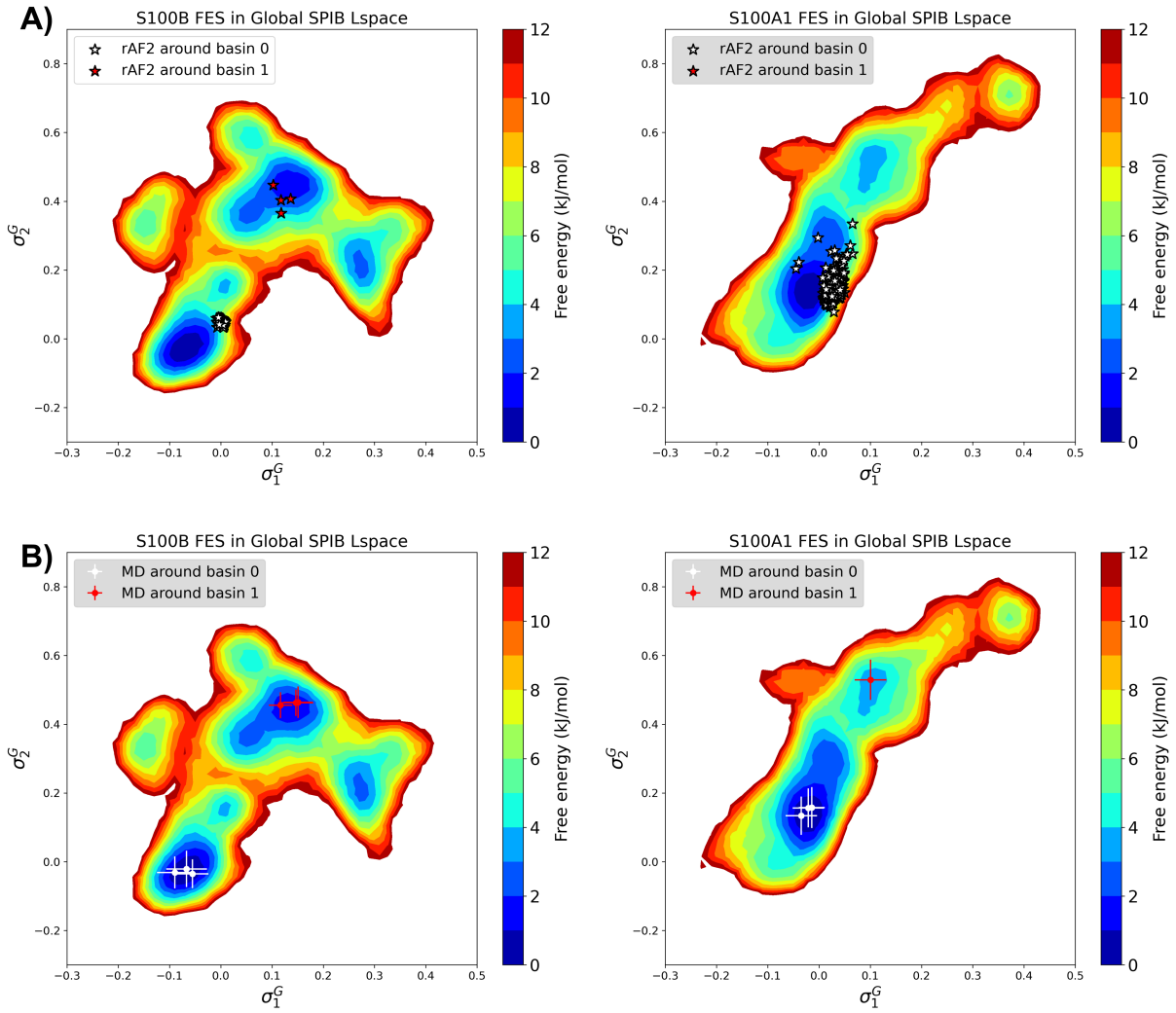


Figure F.7: A) Projection of reduced MSA AF2 (rAF2) structures which fall into the two major basins onto the global SPIB latent space, based on a free energy cutoff of 4 kJ/mol. B) The mean and variance of IB values from unbiased MD trajectories used for subsequent local/pocket SPIB analysis within each global basin. Left and right panels correspond to the Ca^{2+}-S100B and Ca^{2+}-S100A1 systems, respectively. No reduced MSA AF2 structures fall into basin 1 of the Ca^{2+}-S100A1 system, and only one unbiased trajectory met the criteria for local SPIB analysis in that basin.

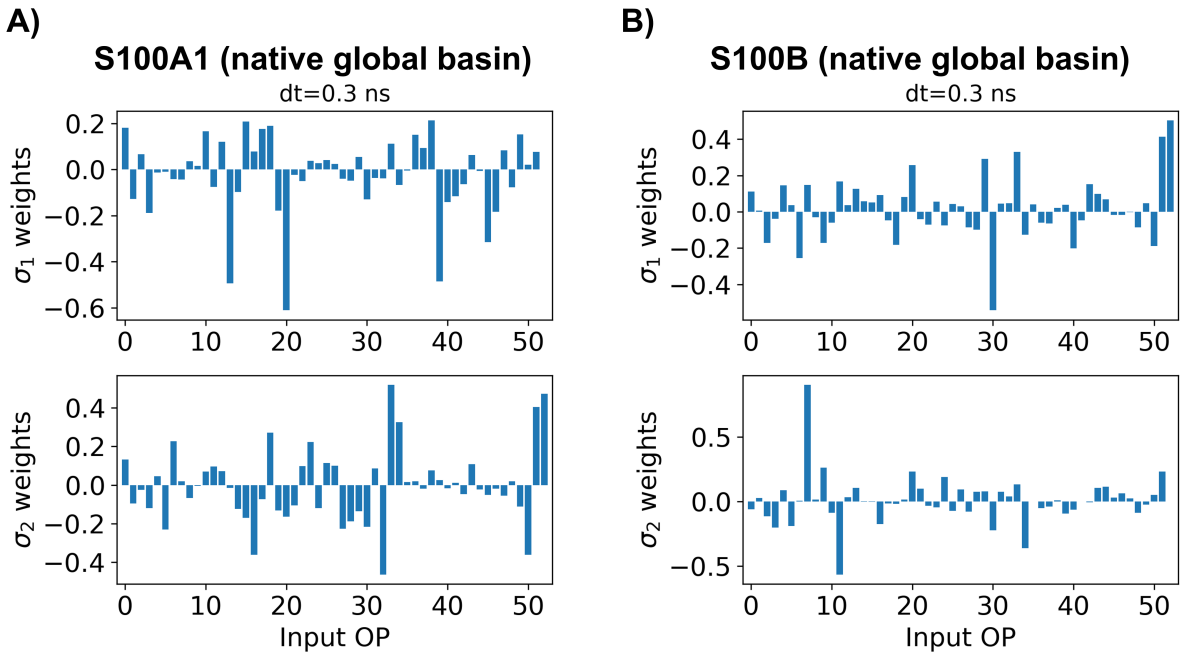


Figure F.8: Bar plots showing the linear weights assigned to each of the 53 local features in the learned local SPIB coordinates, on the Ca^{2+}-S100A1 native global basin (subplot A) and Ca^{2+}-S100B native global basin (subplot B), using a time lag of 300 ps.

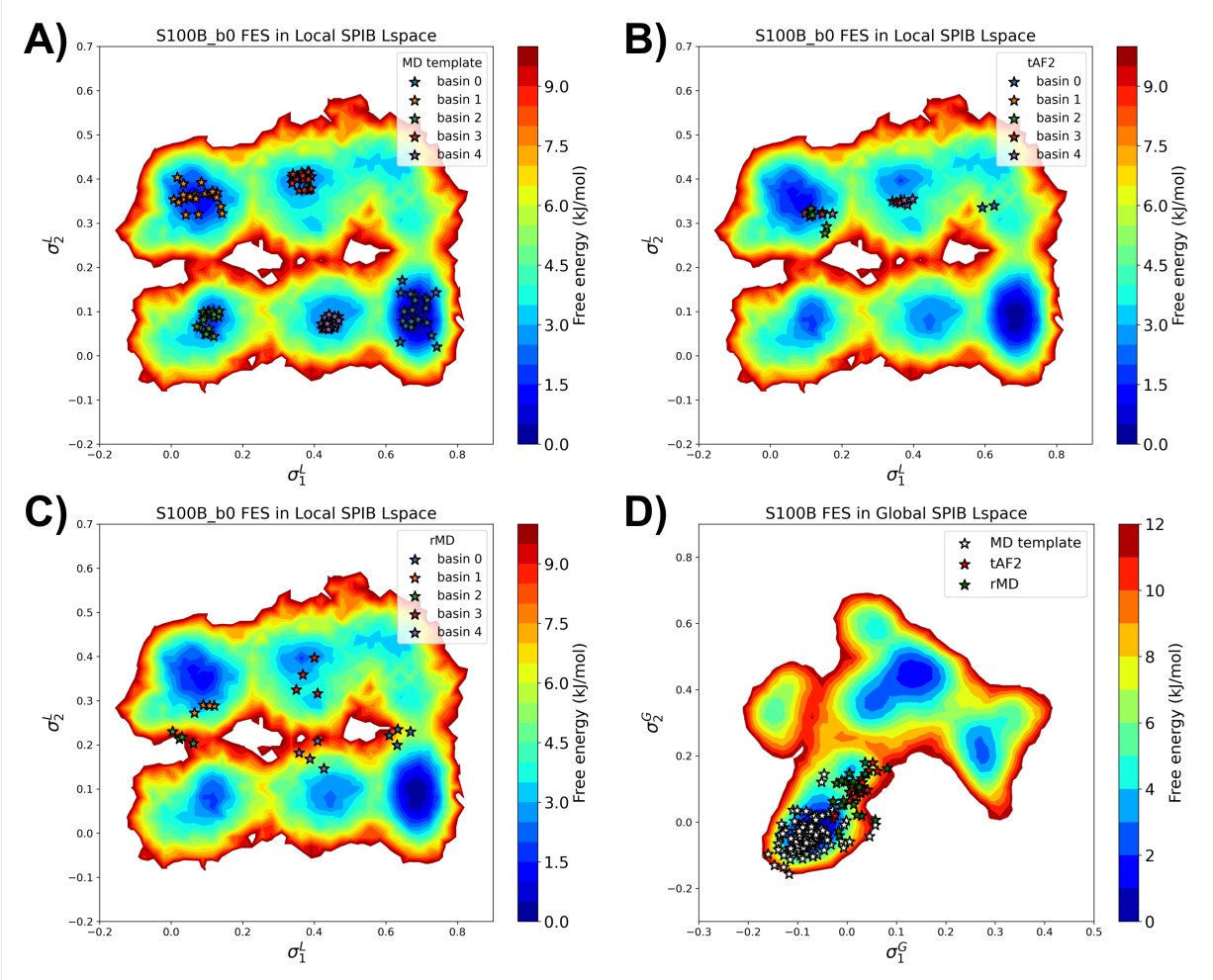


Figure F.9: A–C) Projection of structures used and generated during the AF2RAVE structure set generation pipeline on the local SPIB space corresponding to the native global basin of Ca^{2+}–S100B: A) MD frames used as templates for AF2-template (MD templates); B) the AF2-template structures (tAF2); C) final AF2RAVE structures after short, restrained MD (rMD). D) Projection of all three structure sets onto the global SPIB space for Ca^{2+}–S100B, illustrating that the global backbone configuration remains consistent throughout the structure generation process.

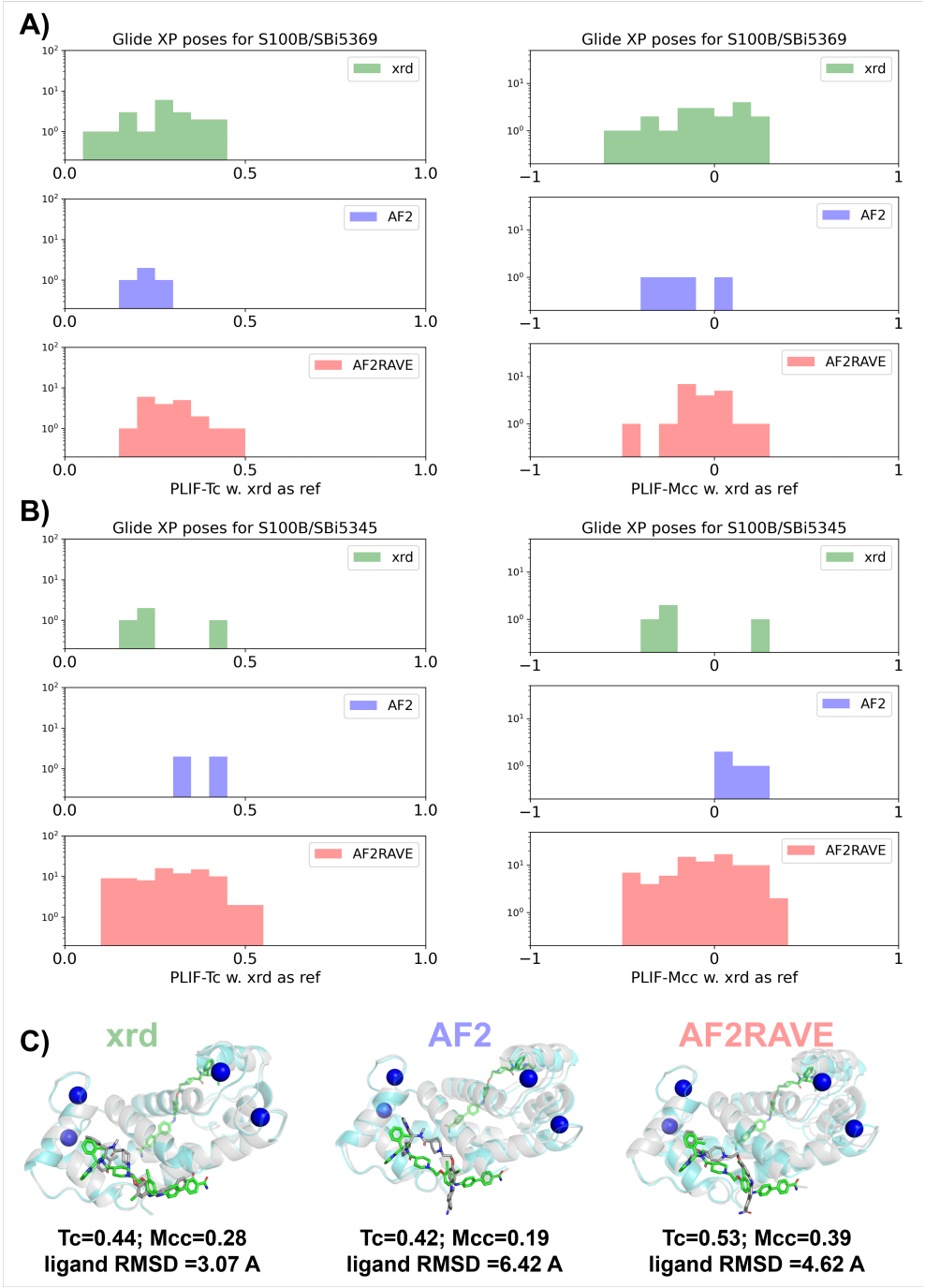


Figure F.10: A-B) Distribution of PLIF similarity coefficients, Tanimoto coefficient (Tc, left panel) and McConnaughey coefficient (Mcc, right panel), for Glide XP docking poses of SBi5369/Ca^{2+}–S100B (subplot A) and SBi5345/Ca^{2+}–S100B (subplot B), using the corresponding X-ray holo structures as references. C) SBi5345/Ca^{2+}–S100B docking poses with the highest Tc values from each structure category. Docking poses (proteins in grey, ligands as grey sticks) are superimposed onto the corresponding X-ray holo structures (proteins in cyan, ligands as green sticks) for comparison.

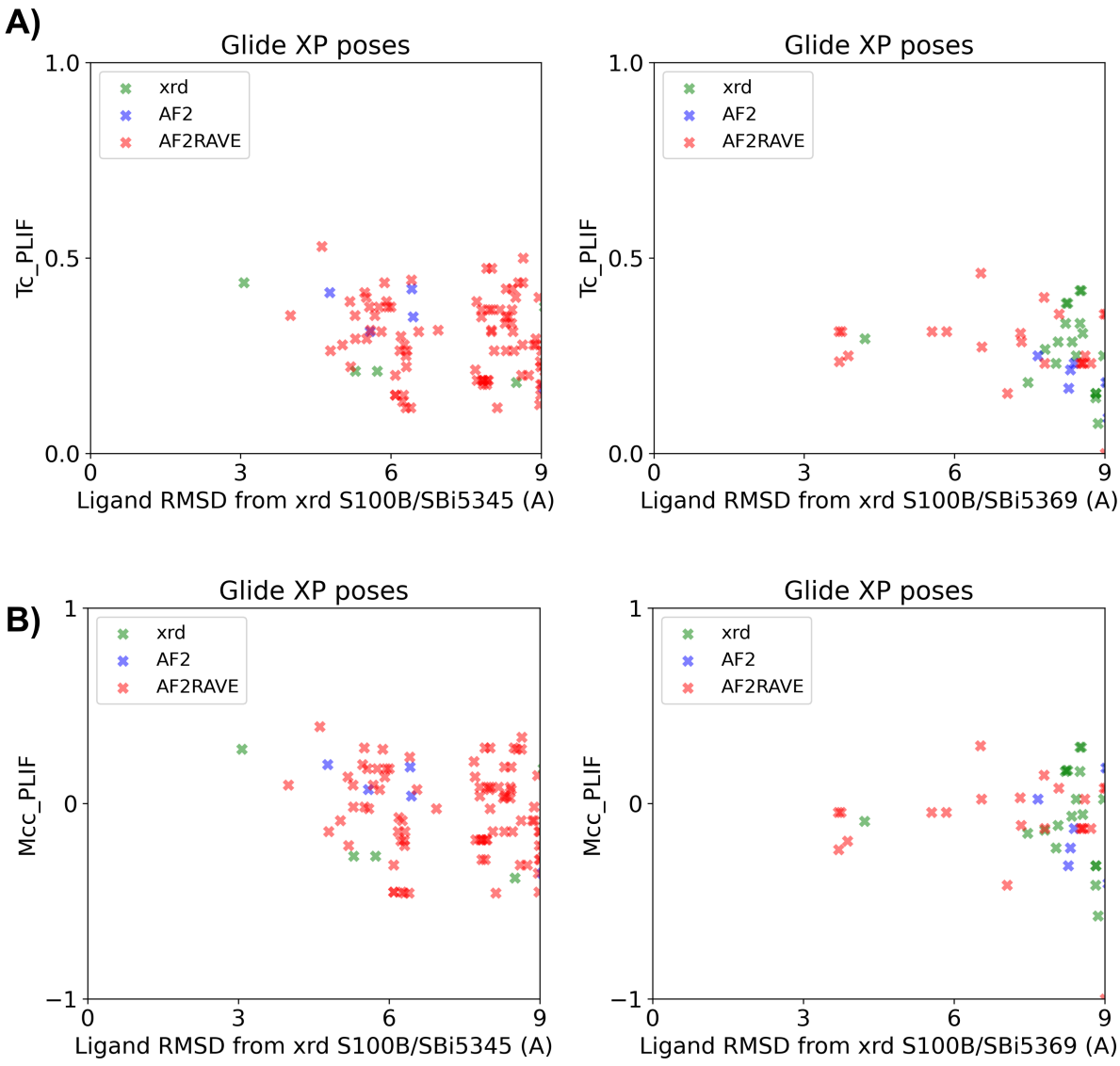


Figure F.11: Ligand RMSD values are plotted against PLIF similarity coefficients, Tanimoto coefficient (Tc, subplot A) and McConnaughey coefficient (Mcc, subplot B), for each Glide XP docking pose, using the X-ray holo structures as reference. The left and right panels represent docking poses for SBi5345/Ca^{2+}–S100B and SBi5369/Ca^{2+}–S100B, respectively. Docking poses that located in the upper left region indicate accurate atomic environment modeling of both ligand placement and protein-ligand interactions. AF2RAVE structures consistently outperform AF2 and perform comparably to experimental XRD structures.

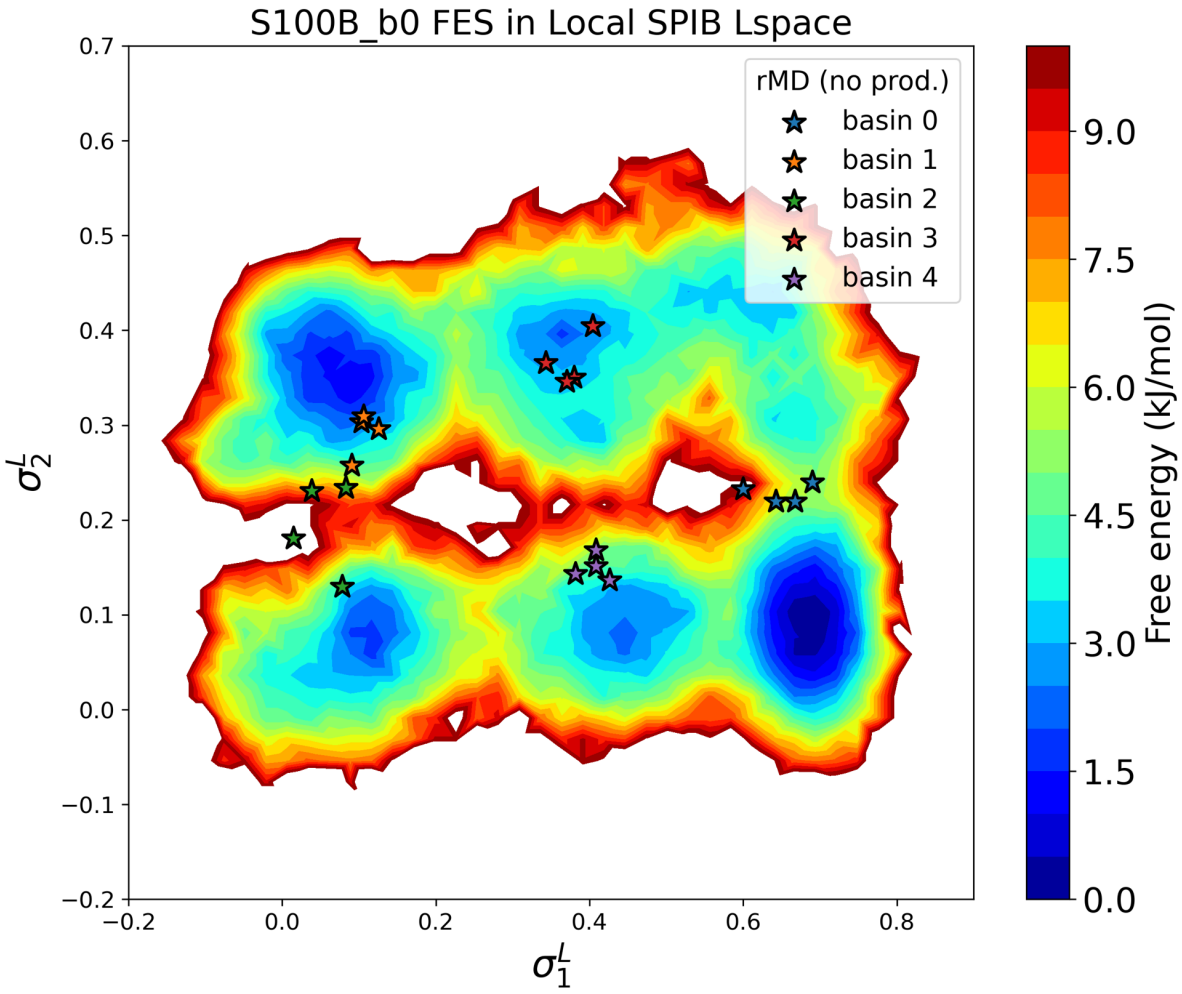


Figure F.12: The AF2RAVE structure set used for the enrichment test. We noticed equilibration trajectories without subsequent production runs in restrained MD provided slightly better coverage of the five local/pocket basins within the native global basin of Ca^{2+}–S100B. Consequently, this structure set was selected for the enrichment analysis.

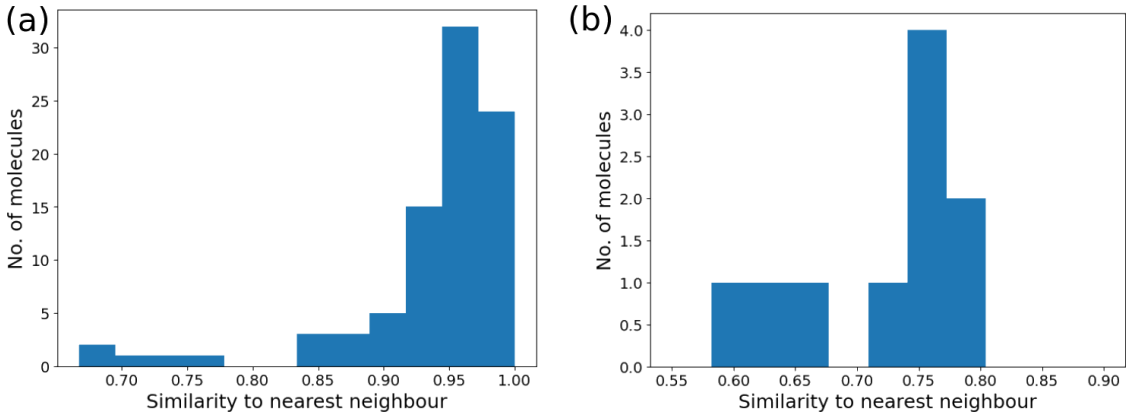


Figure F.13: Average similarity between known ligands was found to be very high. The similarity distribution (SD) of each ligand to its nearest neighbor in that sets is shown. (a) SD using all 87 ligands known to bind to S100B (b) SD using 10 selected ligands that were as distinct to each-other as possible

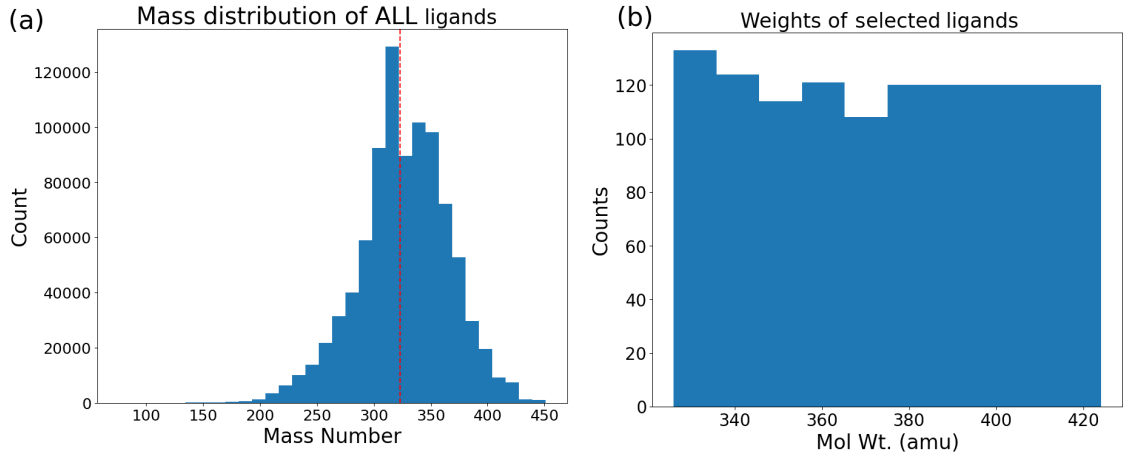


Figure F.14: The average size of ZINC ligands is much lower than our set of known binders. Molecular weight distributions (in a.m.u) for (a) 2 million random ligands from ZINC database. The red line denotes the lowest molecular weight of a known binder (325 a.m.u) (b) Selected ”heavy” subset as decoys. All decoys were selected to be heavier that the cut-off (red line in (a))

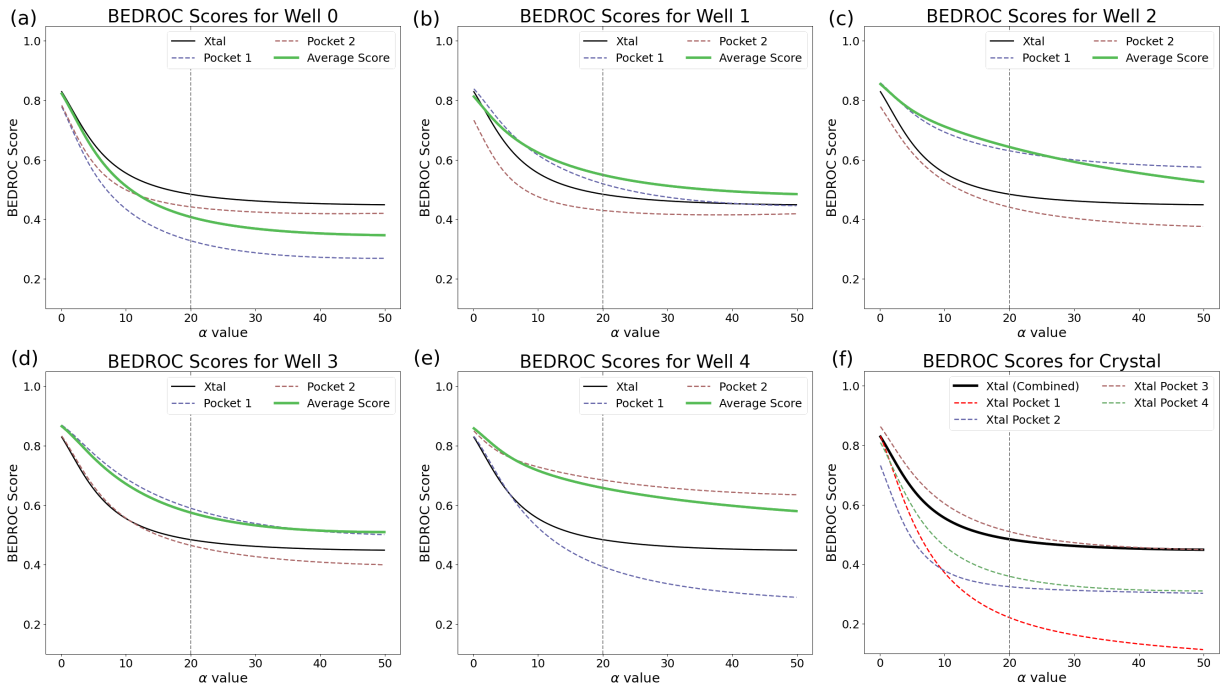


Figure F.15: Enrichment diversity from individual pockets. The average profiles (thick green) used to calculate enrichment are computed by first averaging the DOCK score from both pockets and then ranking all ligands by this average score. Profile for the crystal structure is shown for reference in each plot. (a-e) 5 basins as identified by local structure SPIB (f) 4 different pockets from 2 known crystal structures. Pocket 1 was omitted from subsequent calculations as key residue (PHE 88) was not modelled. The vertical dashed line is the value of α most commonly used in literature to compare these profiles (higher is better).

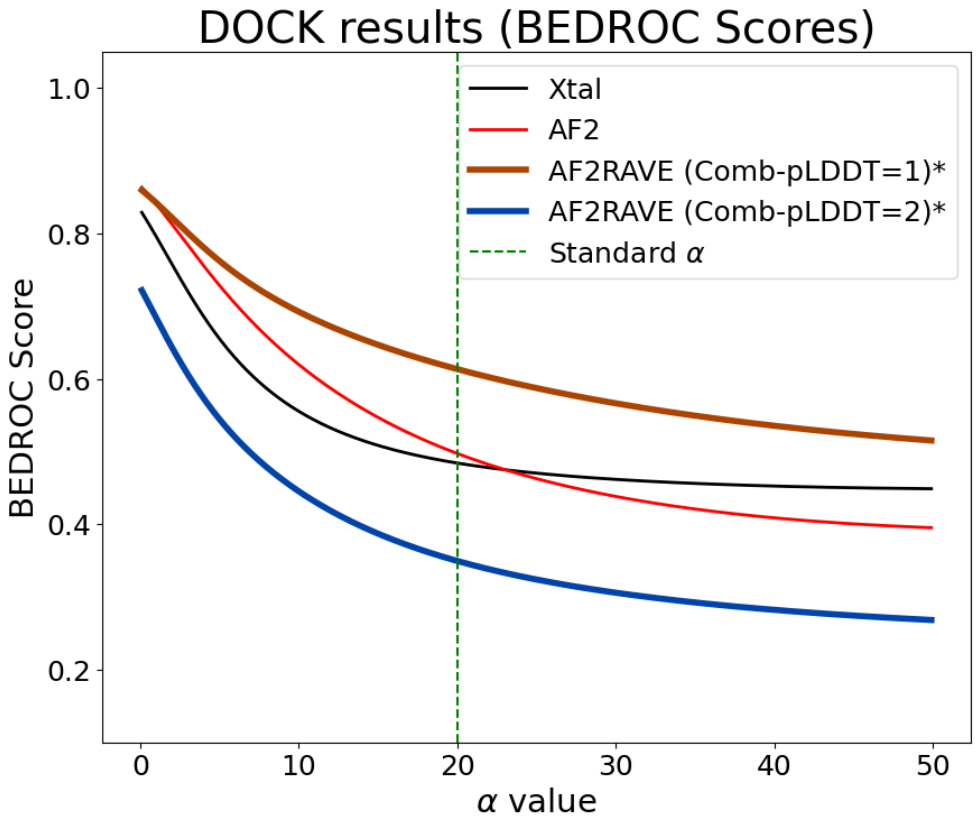


Figure F.16: Comparing enrichment profiles from AF2 structures with pLDDT rank 1 (AF2 recommended) and pLDDT rank 2 (used to improve variance). However, diversity comes at a cost of poor structure quality as lower ranked AF2 structures show extremely poor enrichment. The dashed line shows the recommended literature value of α to use for comparing these profiles (higher is better)

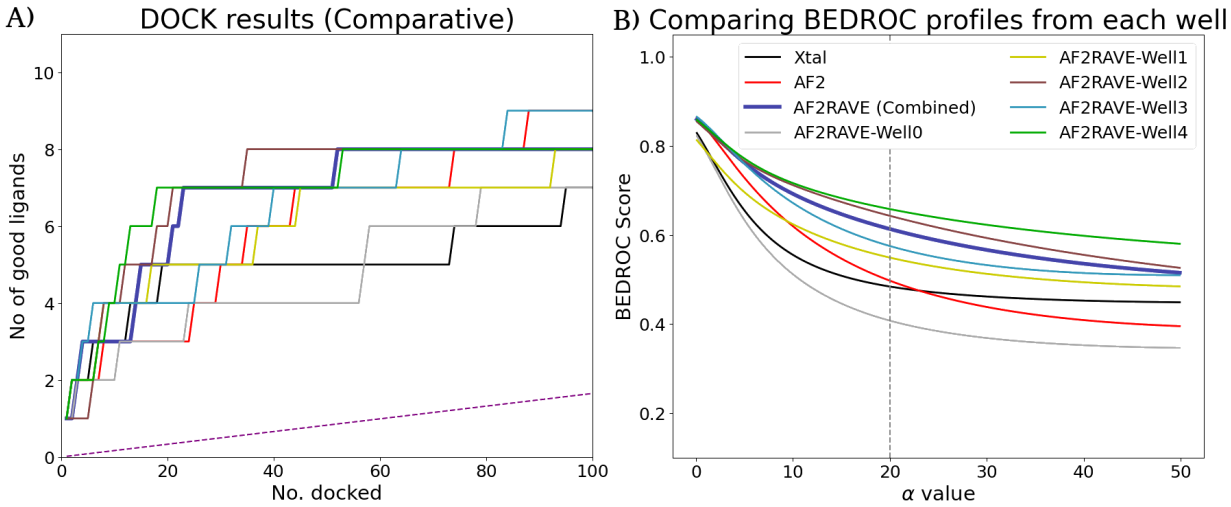


Figure F.17: Comparing the enrichment profiles from each basin. Key for both is shown in the inset of plot B. A) Direct enrichment profile showing number of true binders found when first K ligands are taken. B) BEDROC profiles constructed by changing α. Grey (bottom-most) represents Basin 0, which provides the least enrichment. Basin 4 (green) closely followed by Basin 2 (Purple) perform the best. The black profile is for the crystal structures (for reference) and the red profile is the vanilla AF2 structure. The dashed line shows the recommended literature value of α to use for comparing these profiles (higher is better)

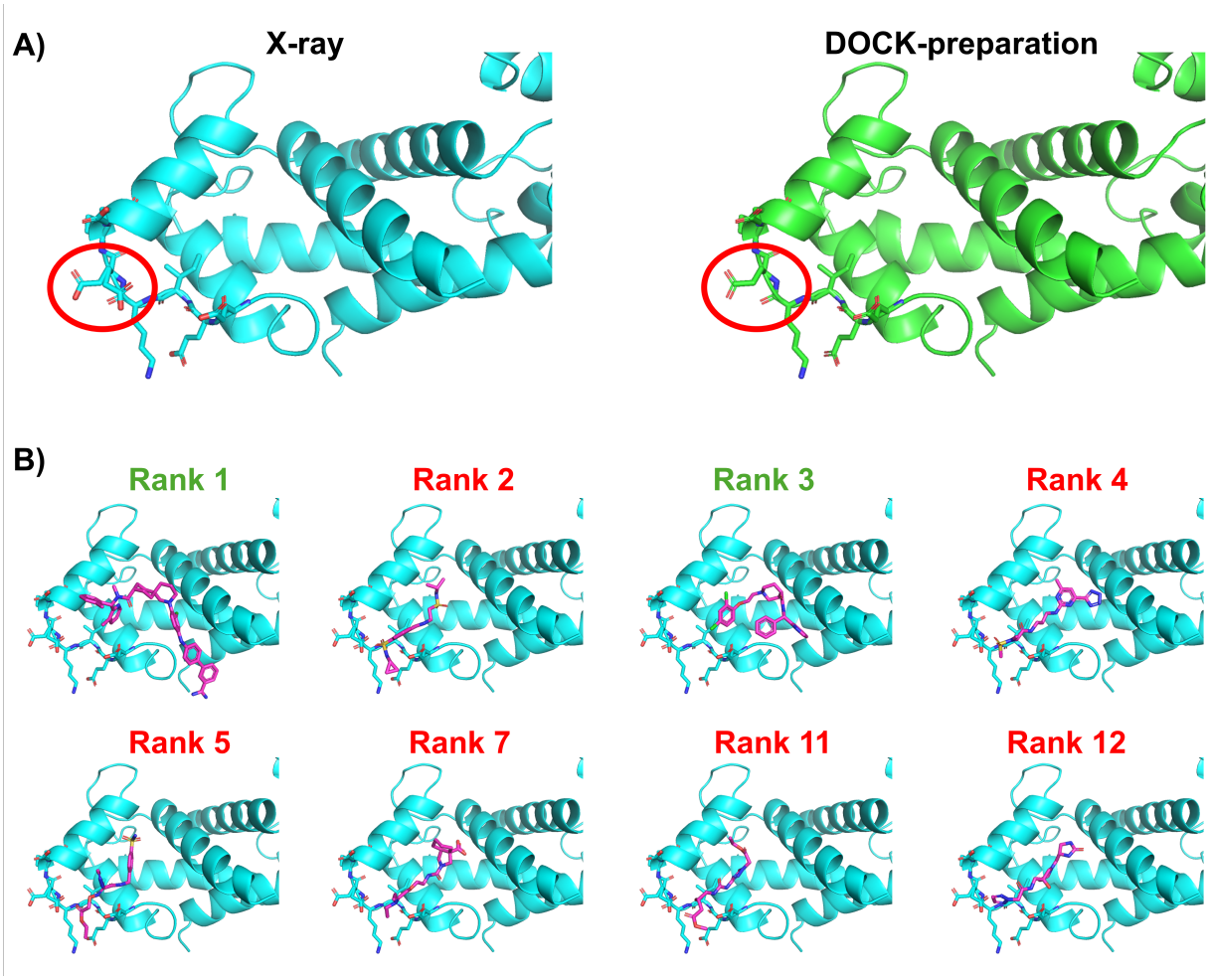


Figure F.18: A) X-ray structure of the SBi5345/Ca^{2+}–S100B complex (cyan, ligand hidden) and its corresponding protein pocket after DOCK protein preparation (green). The GLU50 residue, which exhibits two possible sidechain rotamer conformations in the X-ray structure, is highlighted with red circles. The default DOCK preparation selected the outward-facing rotamer for docking. B) Top-ranked compounds (magenta sticks) from the enrichment test using the DOCK-prepared X-ray structure of SBi5345/Ca^{2+}–S100B. Docking poses are aligned with the original X-ray structure (cyan). Ranking of active inhibitors are labeled in green, while ranking of decoys are labeled in red. Several top-ranked decoys exhibit potential steric clashes with the inward-facing GLU50 conformation, indicating that ambiguities in X-ray sidechain assignments allow these decoys to artificially achieve high docking scores, thereby compromising enrichment performance. GLU50 and neighboring residues are shown as sticks; calcium ions are omitted for visual clarity.

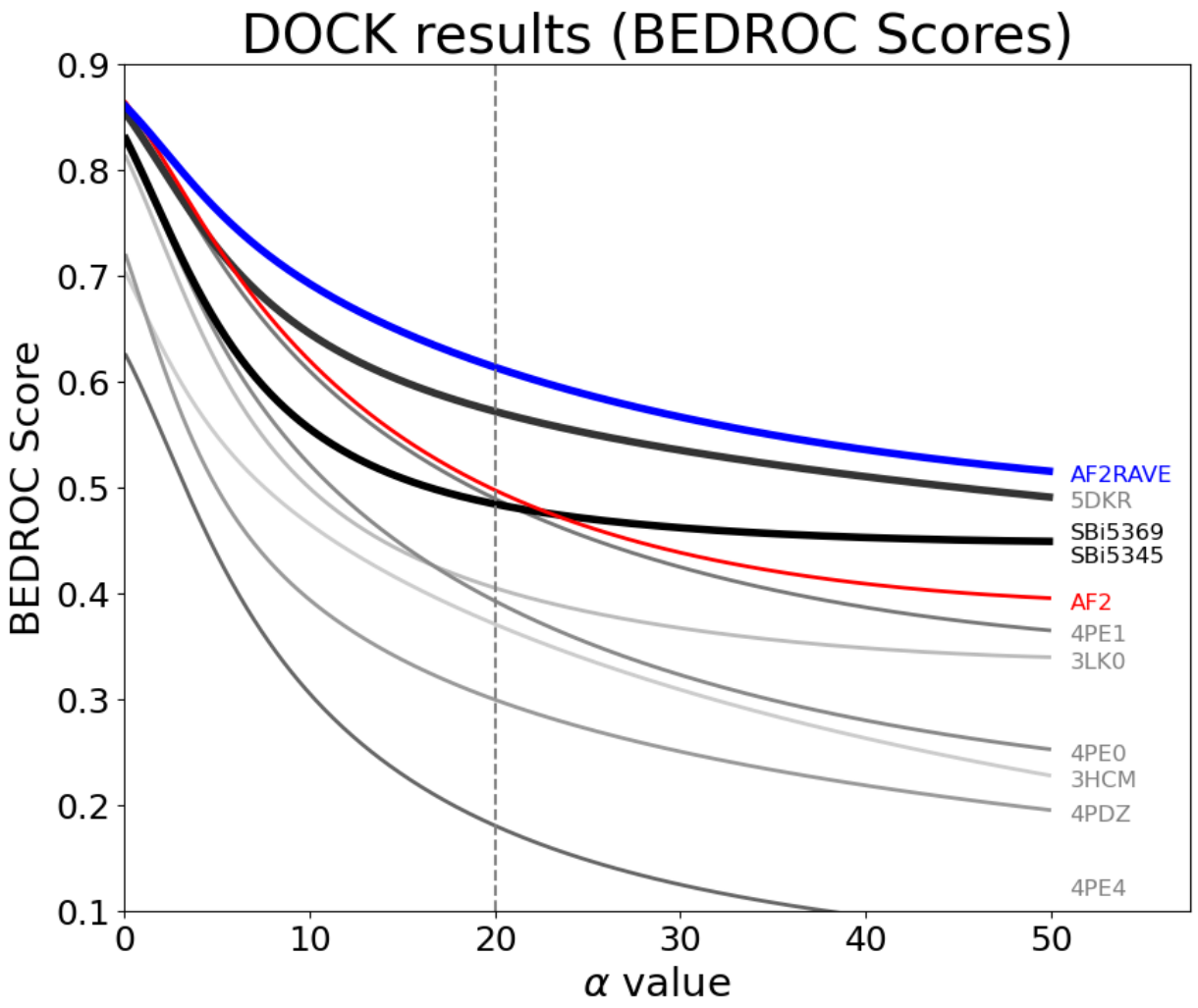


Figure F.19: Variation in BEDROC scores seen by changing the crystal structures (each bound to a different ligand). Each curve represents one crystal structures (score averaged from both pockets). Structures that were monomeric, covalently modified, or bound to large molecules and peptides were omitted from the plot. All curves are labelled at their bottom-right by the PDB ID of the crystal structure used for docking. The black profile is for the crystal structures (for reference) and the red profile is the vanilla AF2 structure. The dashed line shows the recommended literature value of α to use for comparing these profiles (higher is better)

Appendix G: Supporting Information for Chapter 7

G.1 Methods

G.1.1 Molecular dynamics setup

All molecular dynamics simulations were performed using OpenMM [341]. Initial structures were prepared using PDBFixer to replace nonstandard residues and add missing atoms. Before building the solvated system, the complex was oriented such that its principal axis was aligned along the positive z direction, and its molecular center was placed at the origin. Hydrogen atoms were then added according to the AMBER99SB-ILDN protein force field [344], and each system was solvated with TIP3P water [345] in a rectangular box of size $11 \times 11 \times 14$ nm^3. Sodium and chloride ions were added to neutralize the system. Long-range electrostatic interactions were treated using particle mesh Ewald [333], with a nonbonded cutoff of 1.2 nm and a switching distance at 1.0 nm. Bonds involving hydrogen atoms were constrained, and dynamics were propagated using the LangevinMiddleIntegrator with a 2 fs timestep at 300 K [346, 371]. During equilibration, positional restraints with a force constant of 10 kJ mol^{-1} Å^{-2} were applied to protein heavy atoms. The restrained system was equilibrated for 1 ns steps in the NVT ensemble and then for 1 ns steps in the NPT ensemble at 1 bar. Pressure was controlled using the Monte Carlo barostat implemented in OpenMM. After equilibration, the positional restraints were removed before production or binding-pose metadynamics simulations.

G.1.2 Binding-pose metadynamics

Binding-pose metadynamics simulations were performed using PLUMED patched with OpenMM [334, 372]. The biased collective variable was an interface RMSD calculated from selected Cα atoms at the TCR-pMHC interface. The reference structure for this RMSD was generated from the equilibrated complex, using interface residues from the TCRα, TCRβ, peptide, and MHC chains. Well-tempered metadynamics was performed by depositing Gaussian bias potentials along the interface RMSD coordinate [64, 70]. The Gaussian height was 1.5 kJ mol^{-1}, the Gaussian width was 0.1 Å, the bias factor was 20, and bias was deposited every 250 steps(0.5 ps). The metadynamics grid spanned 0–50 Å with 1000 bins. The time-dependent reweighting constant $c(t)$ was calculated using the PLUMED reweighting estimator and used as the binding-pose stability score.

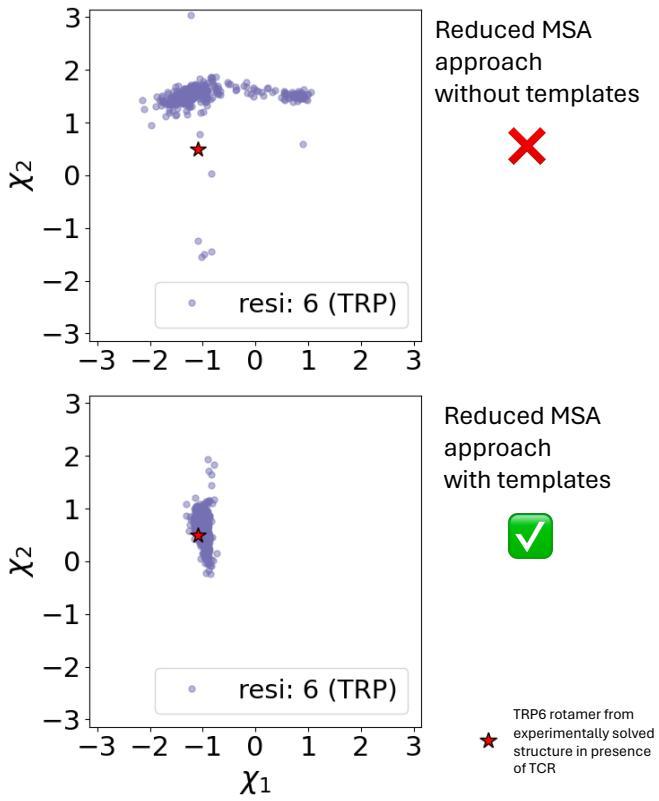


Figure G.1: Template usage case study. The χ_1-χ_2 conformational space of the peptide position-6 tryptophan residue is shown. Violet dots represent generated structures, and the red star denotes the experimental structure.

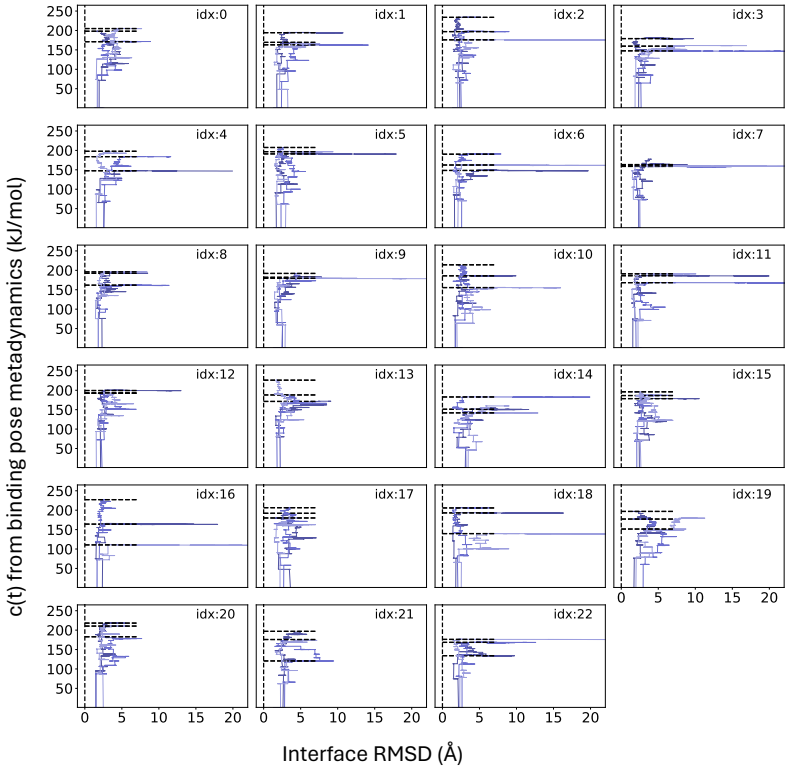


Figure G.2: Binding pose metadynamics results.

Appendix H: Supporting Information for Chapter 8

H.1 Synaptic Cleft Calculation

Measured and computed NALM6-CAR T-cell intermembrane distances from cryo-ET images are presented as follows. "Outer to Outer" refers to the distance between the outer plasma membranes of the two cells, while "Inner to Inner" refers to the distance between the two inner plasma membranes, including the phospholipid bilayer (Supplementary Fig. H.7a). The synaptic cleft was determined as the distance between "Outer to Outer" together with the proteinaceous layer (PL) of each plasma membrane.

We used the following calculations to build a dataframe of the mean and standard deviation of the measured microscopy distances and to estimate the synaptic cleft distances. The lipid bilayer was assumed to have a thickness of approximately 4 nm [373].

$$H \in \{8, 15, 47\}, \tag{H.1}$$

$$d_{\text{in-in}} : \text{distance between inner membranes of the synapse}, \tag{H.2}$$

$$d_{\text{out-out}} : \text{distance between outer membranes of the synapse}, \tag{H.3}$$

$$t : \text{estimated lipid bilayer thickness } (4 \text{ nm}), \tag{H.4}$$

$$pm_{\text{mean}} = \frac{\text{avg}\left(d^H_{\text{in-in}} - d^H_{\text{out-out}}\right)}{2}, \tag{H.5}$$

$$pm_{\text{std}} = \frac{\text{std}\left(d^H_{\text{in-in}} - d^H_{\text{out-out}}\right)}{2}, \tag{H.6}$$

$$PL(pm) = \frac{pm_{\text{mean}} - t}{2}, \tag{H.7}$$

$$\text{SynapticCleft}^H_{\text{mean}} = d^H_{\text{out-out}} + 2PL(pm). \tag{H.8}$$

To extract data from regions corresponding to the cSMAC of a synapse, we focused on CAR-T/target-cell apposed membranes with high CAR accumulation, defined by higher fluorescence intensity (Supplementary Fig. H.7c–e). Since this calculation is based on sampling a small set of regions from two-dimensional slices of a synapse, even in the presence of fluorescence, certain regions can behave as outliers (for instance, Supplementary Fig. H.7c R8 and Supplementary Fig. H.7e R6). We therefore included only the regions that were observed with higher probability in our analysis; that is, we removed outlier regions based on the interquartile range (IQR) defined from the mean synaptic cleft values from each region (Supplementary Fig. H.8a,b).

H.2 Modelling Methods

The in silico modeling of the hinges for our CAR-BOOST method was adapted from the AF2RAVE family of methods that comprise both bioinformatic and biophysical approaches [1, 94, 167, 184]. At the core of this framework is RAVE (Reweighted Autoencoded Variational Bayes for Enhanced Sampling), an AI-augmented molecular dynamics (MD) scheme to iteratively reweight and produce accurate probability distributions along any chosen low-dimensional coordinate [374]. In combination with AF2, AF2RAVE generates an ensemble of protein structures weighted by their thermodynamic stability, as represented through Boltzmann weights, offering a more realistic view of protein flexibility and function than a static picture from AF2 [94]. In our past works, we have demonstrated the performance of AF2RAVE on disordered regions and its ability to overcome the MSA bias from AF2 [94, 114]. These results led us to use the AF2RAVE protocol for the CAR CD8α-derived disordered hinge region. To obtain an ensemble of possible structures, we used AF2 [1] and AF-multimer [184] from ColabFold [11]. These were used for structural ensemble generation of CAR CD8α-derived hinges, ordered domains of the TCR IS, Ig domains of the target receptors, and the scFv domain of CAR. We modified ColabFold to implement the reduced MSA approach (rMSA AF2) [91] and to add a post hoc geometry filter to account for a membrane plane. While such an approach can generate an ensemble of structures, it is not clear which of the structures are biophysically meaningful. To solve this, we used enhanced sampling, specifically well-tempered metadynamics, to reweight the z_{e2e} distance distribution [64]. The final synaptic cleft calculations for a CAR against a specific receptor and the complementary synapse cSMAC proteins were determined under an independent distribution assumption. This assumption allows us to significantly reduce the needed computational power and transfer the CAR CD8α-derived hinge distribution to other target receptors, thus avoiding additional simulation runs.

H.2.1 rMSA AF2/AF-multimer

Reduced Multiple Sequence Alignment AlphaFold2 (rMSA AF2) is a variant of AlphaFold2 (AF2) where the amount of coevolutionary information that goes into the AF2 network for a given protein sequence is reduced [91]. This modification allows AF2 to generate conformationally diverse structures. Modified ColabFold was used to generate all the structures/ensembles in this chapter with additional templates if required. The MSA parameter information, the number of samples generated, and the usage of templates are all shown in the respective tables in the Supporting Information (Supplementary Fig. H.3b, Supplementary Fig. H.4a, and Supplementary Fig. H.9a). For CAR hinges and disordered linker regions of the target protein, MSA parameters of 8:16 and 16:32, respectively, were used, with each setting generating 640 structures.

We added an additional post hoc geometry filter that calculates a plane based on the alpha helix of the transmembrane domain to incorporate information about the membrane plane into the distribution (Supplementary Fig. H.3g). Since AF2 is trained to predict a global structure, most disordered regions predicted by AF2 represent structures with a lower radius of gyration. When a disordered-region sequence contains two cysteines, AF2, regardless of the MSA depth, places these cysteines together, thus skewing the distribution. To address this AF2 bias, we mutated the membrane-distal cysteine in the CD8α-derived hinges to its sterically equivalent serine (Supplementary Fig. H.5b–d). We then used the resulting conformations to design our biasing

coordinate for well-tempered metadynamics. This is justified because the membrane-proximal cysteine is present in a sterically improbable position for a disulfide bond, and experiments have shown that these cysteines are possibly involved in the formation of an interprotein disulfide bond rather than an intradisulfide bond [375].

H.2.2 Molecular dynamics

Structures obtained from rMSA AF2 without the cysteine mutation were clustered by uniformly binning them in z_{e2e} space. Structures from different bin centers were randomly sampled and placed in a POPC bilayer membrane using the CHARMM-GUI web tool [350, 351, 376]. We used the OpenMM Python package [341] to perform MD simulations under the CHARMM36m force field [349]. LangevinMiddleIntegrator was used as the simulation integrator; it uses a BAOAB-type approach and handles the temperature [346, 371]. Equilibration runs were performed for 300 ps under the Constant Volume method of MonteCarloMembraneBarostat in OpenMM with restraints on the head groups of POPC and the heavy atoms of the protein chain. Further rounds of equilibration were performed under the Z Free method of the barostat with different scales of the restraints until they were completely removed. In each case, an equilibration run of 300 ps was performed. Overall, equilibration steps accounted for approximately 1.5 ns of simulation time. The production run was performed under the well-tempered metadynamics scheme at different simulation times for different CAR hinges. The time series and other details can be found in Supplementary Fig. H.6b–k and Supplementary Fig. H.10a,d.

H.2.3 Enhanced sampling

CD8α-derived hinges are intrinsically disordered regions, making it challenging to sample all possible conformations with their corresponding thermodynamic weights [377]. To this end, we used well-tempered metadynamics to overcome the sampling problem [64, 70]. Well-tempered metadynamics (WT-metaD) is a class of enhanced sampling that deposits history-dependent penalties to discourage the system from visiting previously visited states. WT-metaD incorporates history-dependent penalties as Gaussian weights in the potential energy of the system. This scheme requires a biasing coordinate along which the Gaussian weights can be deposited. To obtain meaningful sampling from WT-metaD, the biasing coordinate must effectively represent the mechanism that is being studied. While z_{e2e} is an intuitive coordinate for this purpose, depositing bias directly along a distance coordinate can create artifacts. We therefore used a nonlinear function of the z_{e2e} distance, specifically a sigmoid function (Supplementary Fig. H.6a). We parametrized the sigmoid function from the rMSA AF2 conformations. We performed a linear regression on the maximum z_{e2e} values of the rMSA AF2 output for different CAR hinges. This regression provided a guess for where the hinge N-terminal endpoint is most likely present in z_{e2e} space (Supplementary Fig. H.6a). Thus, we centered our sigmoid for a CAR with a specific number of hinge residues around its respective linear regression output with an offset. The sigmoid slope was chosen so that the linear phase of the sigmoid starts at the membrane in z_{e2e} space (Supplementary Fig. H.6a). The parameters used for the sigmoid are given in Supplementary Fig. H.6a. Further, we defined lower walls at 1.5 nm from the membrane to avoid collapse of the hinge domain onto the membrane. We used the open-source, community-developed PLUMED package [334], version 2.7.3 [372], patched with OpenMM to perform WT-metaD. The initial

height of the Gaussian weight was 1.5 kJ mol^{-1}. The Gaussian weight was deposited every 5 ps. This procedure allowed us to enhance the sampling, as seen from the transitions between extended and collapsed states in the time series presented in Supplementary Fig. H.6b–l and Supplementary Fig. H.10a,d.

H.2.4 Defining lower and upper bounds

For the lower-bound calculation, we used AF-multimer to generate multiple conformations of the protein pairs present in the synapse. We removed sequences corresponding to the disordered extracellular or hinge regions of the cSMAC protein pairs. We verified whether the binding poses of these complexes are similar to those in the Protein Data Bank. Notably, CD28 exhibited a longer hinge region than CD2 and the TCR, resulting in z_{e2e} distributions positioned near or below 10 nm. Consequently, the lower bound was calculated using z_{e2e} values from CD2-CD58 and TCR-MHC complexes. The mean value across these distributions was approximately 12.5 nm.

For the upper-bound calculation, we employed rMSA AF2 to analyze the CD45RO sequence. Using our geometry filter, we calculated the z_{e2e} distribution for the structured D1-D4 domains, yielding an approximated average value of 14.5 nm. Additional sampling from molecular dynamics on these systems can yield more robust upper and lower bounds. However, given the presence of these complexes in AF2's training set, additional simulations in this case would probably contribute changes on a 0.1 nm scale. Thus, to verify the bounds in an alternative approach, we directly computed the synapse distribution of these complexes rather than performing additional simulations. These two calculations enabled us to assess whether the complexes reside within the defined upper and lower limits.

H.2.5 TCR IS calculation using the CAR-BOOST approach

The z_{e2e} distance distributions for the TCR-pMHC complex within the immunological synapse were calculated using the algorithm described in Fig. 8.3c. We used rMSA AF2 as the distribution generator and calculated z_{e2e} distances for each receptor, as shown in Supplementary Fig. H.4. For the TCR beta chain, we used the OPM server to calculate the orientation of the transmembrane helix, thereby reorienting the normal to the membrane plane [378]. In cases of multiple binding partners, such as the TCR alpha and beta chains interacting with the MHC class I alpha chain, the synapse probability was averaged across the pairwise interactions. Additionally, the mean z_{e2e} distribution from all the synapse protein pairs was determined by averaging the probability in each bin, as illustrated in Supplementary Fig. H.4b.

H.2.6 Free energy calculation

The free energy was calculated by reweighting the enhanced MD sampled distribution using the reweighting procedure presented by Tiwary et al. [6]. The bias corresponding to samples where the N-terminal endpoint of the CAR hinge was interacting with the membrane was neglected to avoid bias from such states. The free energy calculated by this procedure is shown in Supplementary Fig. H.6b–j. The kernel density estimator parameters used for the free energy calculations are presented in Supplementary Fig. H.9c.

H.2.7 Intermembrane distance and synaptic cleft calculation

The probability distribution $P_{z^{\mathrm{CAR}}}(z)$, where z^{CAR} is the membrane-perpendicular z_{e2e} distance, was calculated using the AF2RAVE protocol shown in Fig. 8.3e. The target protein's $P_{z^{\mathrm{Target}}}(z)$ was calculated by obtaining the z_{e2e} distribution for the disordered linker region using rMSA AF2 and adding an offset from the ordered target domain (Supplementary Fig. H.9b). Similarly, the scFv domain offset is added to $P_{z^{\mathrm{CAR}}}(z)$ (Supplementary Fig. H.9b). Finally, by assuming that these two distributions are independent, we can calculate the intermembrane distance distribution $P_{z^{\mathrm{IM}}}(z)$, where $z^{\mathrm{IM}} = z^{\mathrm{CAR}} + z^{\mathrm{Target}}$, by convolving them as follows:

$$P_{z^{\mathrm{IM}}}(z) = \int_{-\infty}^{\infty} P_{z^{\mathrm{CAR}}}(z') P_{z^{\mathrm{Target}}}(z' - z)\, dz'. \tag{H.9}$$

H.2.8 ϕ value calculation

To obtain an ensemble parameter that quantifies the overlap of the synaptic cleft probability with the functional synapse gap, we performed the following calculations. The probability of a state s_i is related to the free energy via $P(s_i) \sim e^{-\beta G(s_i)}$, where β is inversely related to temperature scaled by the Boltzmann constant (k_B). This implies that the relative probability of two states, in this case two regions in the synaptic cleft, is related to their free energy difference (ΔG) as follows:

$$\Delta G_{O_1 \to S} = -\frac{1}{\beta} \ln \left(\frac{\int_S P(z)\, dz}{\int_{O_1} P(z)\, dz} \right). \tag{H.10}$$

Thus, the ϕ values are calculated as

$$\phi = \beta \left[\Delta G_{O_1 \to S} + \Delta G_{O_2 \to S} \right]. \tag{H.11}$$

H.2.9 Error calculation for the ϕ values

The error calculation for the ϕ values was performed using the following error propagation formula:

$$\Delta\phi = \pm \sqrt{\left(\frac{\Delta P_{O_1}}{\overline{P}_{O_1}} \right)^2 + \left(\frac{\Delta P_{O_2}}{\overline{P}_{O_2}} \right)^2 + 4 \left(\frac{\Delta P_S}{\overline{P}_S} \right)^2}. \tag{H.12}$$

The individual relative uncertainties were calculated using the standard error from the simulation replicas with respect to their average ($\overline{P}_{O_1}$, $\overline{P}_{O_2}$, and $\overline{P}_S$). The coefficient of 4 in the case of region S is due to the double counting of P_S in the ϕ value.

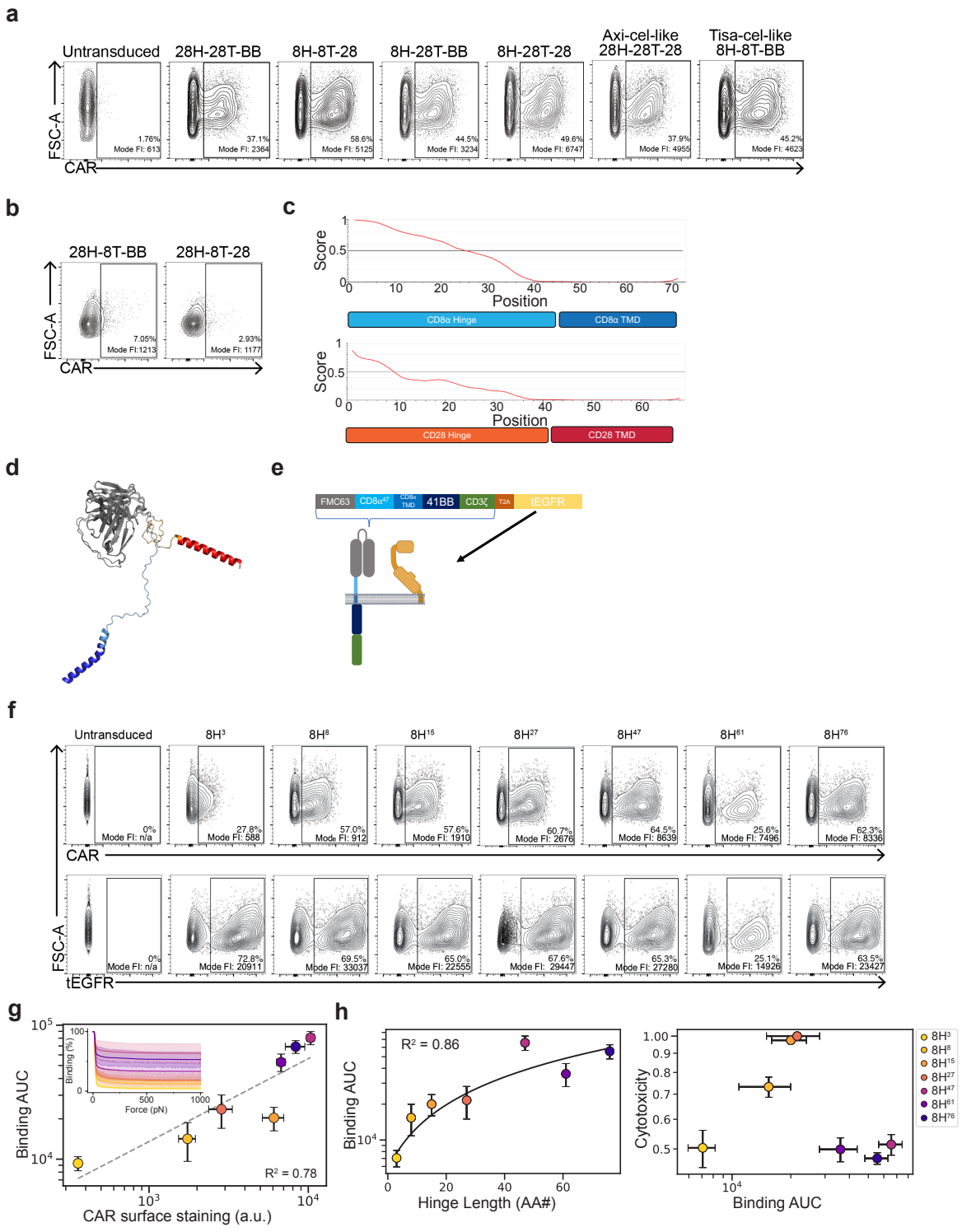

a
Untransduced
28H-28T-BB
8H-8T-28
8H-28T-BB
8H-28T-28
Axi-cel-like 28H-28T-28
Tisa-cel-like 8H-8T-BB
FSC-A
CAR
1.76% Mode FI: 613
37.1% Mode FI: 2364
58.6% Mode FI: 5125
44.5% Mode FI: 3234
49.6% Mode FI: 6747
37.9% Mode FI: 4955
45.2% Mode FI: 4623
b
28H-8T-BB
28H-8T-28
7.05% Mode FI:1213
2.93% Mode FI: 1177
c
Score
Position
CD8α Hinge
CD8α TMD
CD28 Hinge
CD28 TMD
d
e
FMC63
CD8α47
CD8α TMD
41BB
CD3ζ
T2A
tEGFR
f
Untransduced
8H3
8H8
8H15
8H27
8H47
8H61
8H76
CAR
FSC-A
tEGFR
0% Mode FI: n/a
27.8% Mode FI: 588
57.0% Mode FI: 912
57.6% Mode FI: 1910
60.7% Mode FI: 2676
64.5% Mode FI: 8639
25.6% Mode FI: 7496
62.3% Mode FI: 8336
0% Mode FI: n/a
72.8% Mode FI: 20911
69.5% Mode FI: 33037
65.0% Mode FI: 22555
67.6% Mode FI: 29447
65.3% Mode FI: 27280
25.1% Mode FI: 14926
63.5% Mode FI: 23427
g
Binding AUC
Binding (%)
Force (pN)
R² = 0.78
CAR surface staining (a.u.)
h
R² = 0.86
Binding AUC
Hinge Length (AA#)
Cytotoxicity
Binding AUC
8H3
8H8
8H15
8H27
8H47
8H61
8H76

Figure H.1: Cell surface expression and functional evaluation of modular CD19 CAR constructs. **a,** CD19 CARs were designed with different hinge (28H or 8H), transmembrane (28T or 8T), and costimulatory (28 or BB) domains, and their expression on the cell surface of transduced T-cells was evaluated using an anti-Whitlow linker antibody. Representative flow plots ($n = 4$ independent donors), percentages of CAR$^+$ T-cells, and modal fluorescence intensities (FIs) of CAR$^+$ T-cells are shown. **b,** Flow cytometry plots of T-cells transduced with CD19 CAR constructs harboring a CD28 hinge domain (28H) combined with a CD8α transmembrane domain (8T, representative of $n = 4$ individual T-cell donors). **c,** IUPred3 Disorder Profiles for the CD8α (top) and CD28 (bottom) CAR hinges combined with their transmembrane domains. **d,** Predicted structures for the tisa-cel-like (blue) and axi-cel-like (orange-red) CAR constructs, using AlphaFold 2 (AF2). Structures were aligned according to their identical scFv domains. **e,** Schematic of a portion of the CD19 CAR expression plasmid, incorporating sequences coding for a truncated EGFR (tEGFR) downstream of the T2A peptide (top), and expression of the two expressed proteins on a transduced cell (bottom). **f,** Representative flow cytometry plots of CAR (top) and tEGFR (bottom) surface staining on T-cells transduced with CD19 CARs harboring CD8α N-terminal truncated and extended hinges using anti-Whitlow linker (CAR) and anti-EGFR antibodies, respectively. **g,** Line plot showing the percentage of binding between T-cells transduced with the indicated CD19 CAR and CD19$^{\text{high}}$ NALM6 leukemia cells ($\sim 7 \times 10^4$ CD19 molecules/cell) as a function of applied force, measured using the Lumicks zMovi® microfluidics platform ($n = 3 - 4$ individual donors). Scatter plot showing CAR surface expression (modal fluorescence intensity) in relation to the cell-to-cell binding area under the curve (AUC), pooled across $n = 3 - 4$ individual donors. The coefficient of determination, R^2, from a linear regression fit to the log-transformed data is shown. **h,** Scatter plot of CAR binding AUC on CD19$^{\text{high}}$ NALM6 leukemia cells ($\sim 7 \times 10^4$ CD19 molecules/cell) as a function of the number of residues in the CD8α hinge. R^2 from a weighted linear regression of binding AUC versus hinge length is shown; the y-axis is log-scaled for visualization ($n = 3 - 4$ individual donors). **i,** Scatter plot showing max-scaled cytotoxicity as a function of CAR binding AUC on matched CD19$^{\text{low}}$ NALM6 leukemia cells ($\sim 1 \times 10^3$ CD19 molecules/cell; $n = 3 - 4$ individual donors). AA#, number of residues.

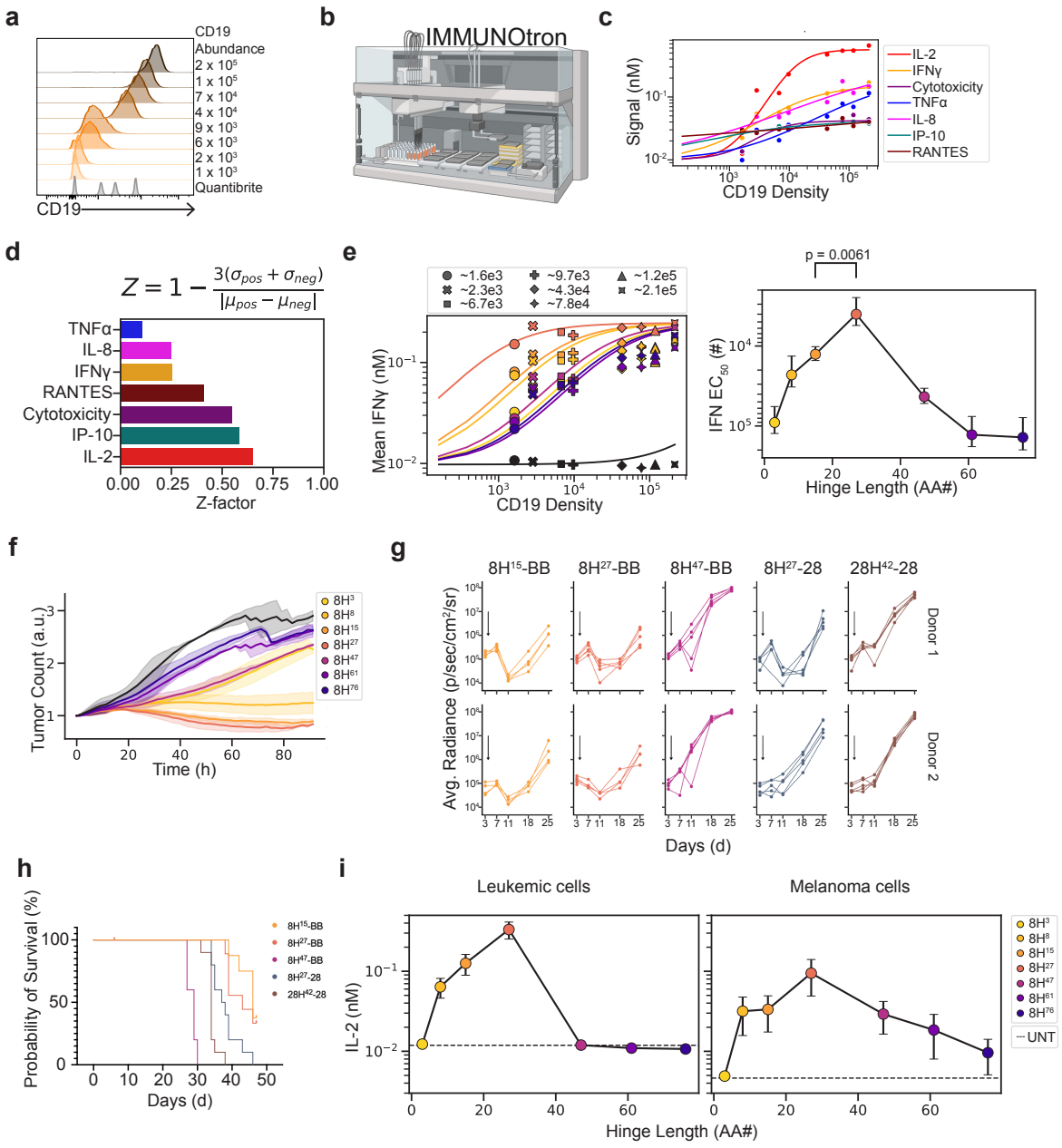

a
CD19
Abundance
2 x 10^5
1 x 10^5
7 x 10^4
4 x 10^4
9 x 10^3
6 x 10^3
2 x 10^3
1 x 10^3
Quantibrite
CD19
b
IMMUNOtron
c
Signal (nM)
CD19 Density
IL-2
IFNγ
Cytotoxicity
TNFα
IL-8
IP-10
RANTES
d
$Z = 1 - \frac{3(\sigma_{pos} + \sigma_{neg})}{|\mu_{pos} - \mu_{neg}|}$
TNFα
IL-8
IFNγ
RANTES
Cytotoxicity
IP-10
IL-2
Z-factor
e
~1.6e3
~2.3e3
~6.7e3
~9.7e3
~4.3e4
~7.8e4
~1.2e5
~2.1e5
Mean IFNγ (nM)
CD19 Density
p = 0.0061
IFN EC50 (#)
Hinge Length (AA#)
f
Tumor Count (a.u.)
Time (h)
8H3
8H8
8H15
8H27
8H47
8H61
8H76
g
8H15-BB
8H27-BB
8H47-BB
8H27-28
28H42-28
Donor 1
Donor 2
Avg. Radiance (p/sec/cm2/sr)
Days (d)
h
Probability of Survival (%)
8H15-BB
8H27-BB
8H47-BB
8H27-28
28H42-28
Days (d)
i
Leukemic cells
Melanoma cells
IL-2 (nM)
Hinge Length (AA#)
UNT

Figure H.2: Cell surface expression and functional evaluation of hinge-modified CD19 CAR T-cells. **a,** Distribution of CD19 surface expression across a set of NALM6 clones[213], measured by anti-CD19 antibody staining. Representative flow cytometry plots are shown. **b,** Schematic representation of the IMMUNOtron robotic platform for the kinetic evaluation of cytokine secretion[224]. **c,** Dose-response curves as a function of CD19 antigen density showing functional outputs (cytokine secretion, chemokine secretion, or cytotoxicity) of tisa-cel-like CAR T-cells co-cultured with NALM6 clones. **d,** Z-factor values for all functional outputs, calculated using the indicated equation (top). Quantification of assayed parameters in tisa-cel-like CAR T-cells (panel c) demonstrates that IL-2 exhibits the highest Z factor (i.e., best dynamic range). **e,** IFNγ dose-response curves for the indicated CD19 CAR T-cells following 24 h co-culture with leukemic cells expressing graded levels of CD19, as indicated in the legend ($n = 4$ individual donors, left). EC_{50} values derived from these curves are plotted as a function of CD8α CAR hinge length (number of residues; right). Error bars represent the standard error of EC_{50} values calculated for each donor (right). **f,** Normalized tumor cell counts following co-culture of CD19 CAR T-cells bearing CD8α hinges of varying length with NALM6 cells expressing $\sim 1 \times 10^3$ CD19 molecules per cell ($n = 2$–4 individual donors). **g,** Line plot showing average in vivo tumor radiance over time in individual mice treated with the indicated CD19 CAR T-cells. The arrow denotes the time of CAR T-cell infusion (day +4). **h,** Kaplan-Meier survival curves for NSG mice engrafted with the $GFP^+Luciferase^+$NALM6 clone expressing $\sim 6 \times 10^3$ CD19 molecules per cell and treated with the indicated CAR T-cells derived from two independent donors. **i,** IL-2 production by CD19 CAR T-cells bearing CD8α hinges of varying length following co-culture with NALM6 leukemic cells expressing $\sim 1 \times 10^3$ CD19 molecules per cell ($n = 4$ individual donors; left) or CD19-expressing SK23-mel-NYESO1OE melanoma cells ($\sim 6 \times 10^3$ CD19 molecules/cell; $n = 2$ individual donors; right). The dashed line indicates IL-2 production by untransduced T-cells. AA#, number of residues.

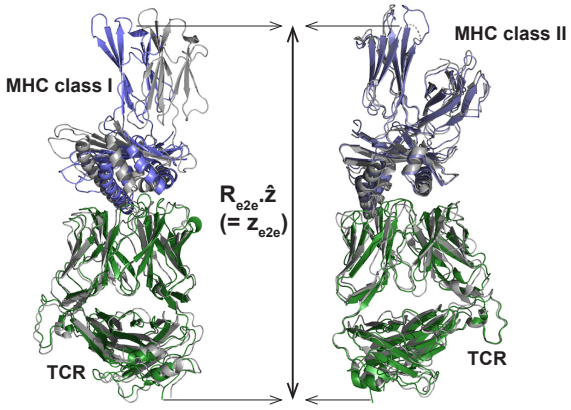
a

Synapse Proteins	Chains	Uniprot ID	Sequence	
			Extracellular ordered	Hinge/disordered + Transmembrane
TCR (PDB: 8ES8)	Chain α	P01848	34-227	227-278
	Chain β	P01850	21-260	260-319
MHC – Class I (PDB: 8ES8)	Chain α	P04439	25-298	298-332
MHC – Class II (PDB:7T2C)	DP α 1 chain	P20036	29-209	209-260
	DP β 1 chain	P04440	30-217	217-258
Costimulatory Protein Receptors				
CD2	Chain A	P06729	25-203	203-235
CD58	Chain A	P19256	29-201	201-238
CD28	Chain A	P10747	19-134	134-155
CD80	Chain A	P33681	35-232	232-263
CD86	Chain A	P42081	24-235	235-268
Out of Synapse Proteins				
CD45 RO	Chain A	P08575	195-600	
CD11a	Chain A	P20701	26-1094	-
CD54	Chain A	P05362	28-503	-

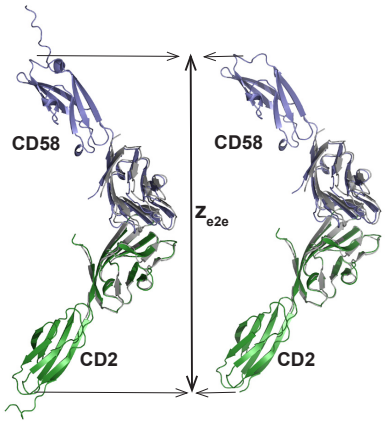
b

Protein	MSA Parameters (MSA_cluster : extra_MSA)	Number of samples	Templates
TCR : MHC I	128 : 256	160	PDB ID: 8ES8
	64 : 128	160	PDB ID: 8ES8
TCR : MHC II	512 : 1204	5	PDB100 dataset
Costimulatory Protein Receptors			
CD2 : CD58	128 : 256	160	None
	64 : 128	160	None
CD28 : CD80	128 : 256	160	None
	64 : 128	160	None
CD28 : CD86	128 : 256	160	None
	64 : 128	160	None
Out of Synapse Proteins			
CD45 (D1 –D4)	16:32	160	Full MSA AF2
CD45RO	32:64	320	Full MSA AF2
CD11a : CD54	512 : 1204	5	None

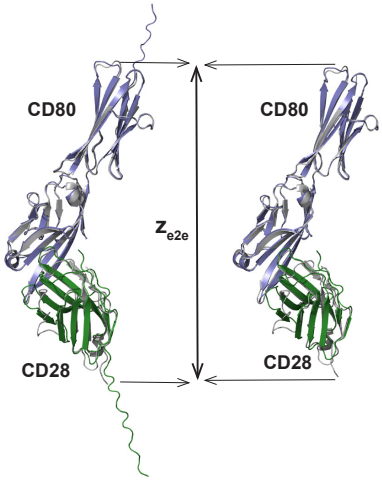
c



d



e



f

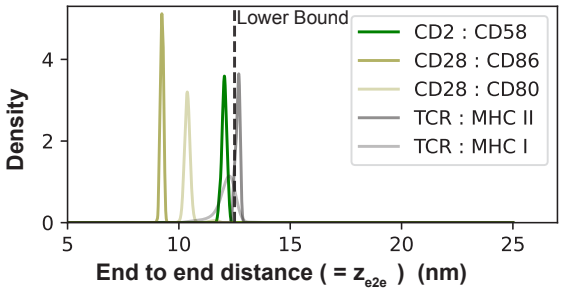


Figure H.3: Defining lower and upper geometric bounds of a functional immunological synapse. **a,** UniProt identifiers and domain segmentation for proteins present at the immune synapse are presented. **b,** Overview of the computational methods used to generate structural models. **c–e,** Structural predictions for TCR-MHC class I/II, CD2-CD58, and CD28-CD80 interaction pairs and their alignments with PDB 8ES8, PDB 7T2C, PDB 1AQ9, and PDB 1I8L (CD80 bound to CTLA-4), respectively, are shown in grey. The highest-ranked structures from the methods in Fig. H.3b are used here for visualization. **f,** Distributions of the end-to-end distances (z_{e2e}) for the ordered domains of CD2-CD58, CD28-CD80/86, and TCR-MHC class I and class II complexes. The black line denotes the lower geometric bound applied in this study ($z_{\mathrm{e2e}} = 12.5$ nm).

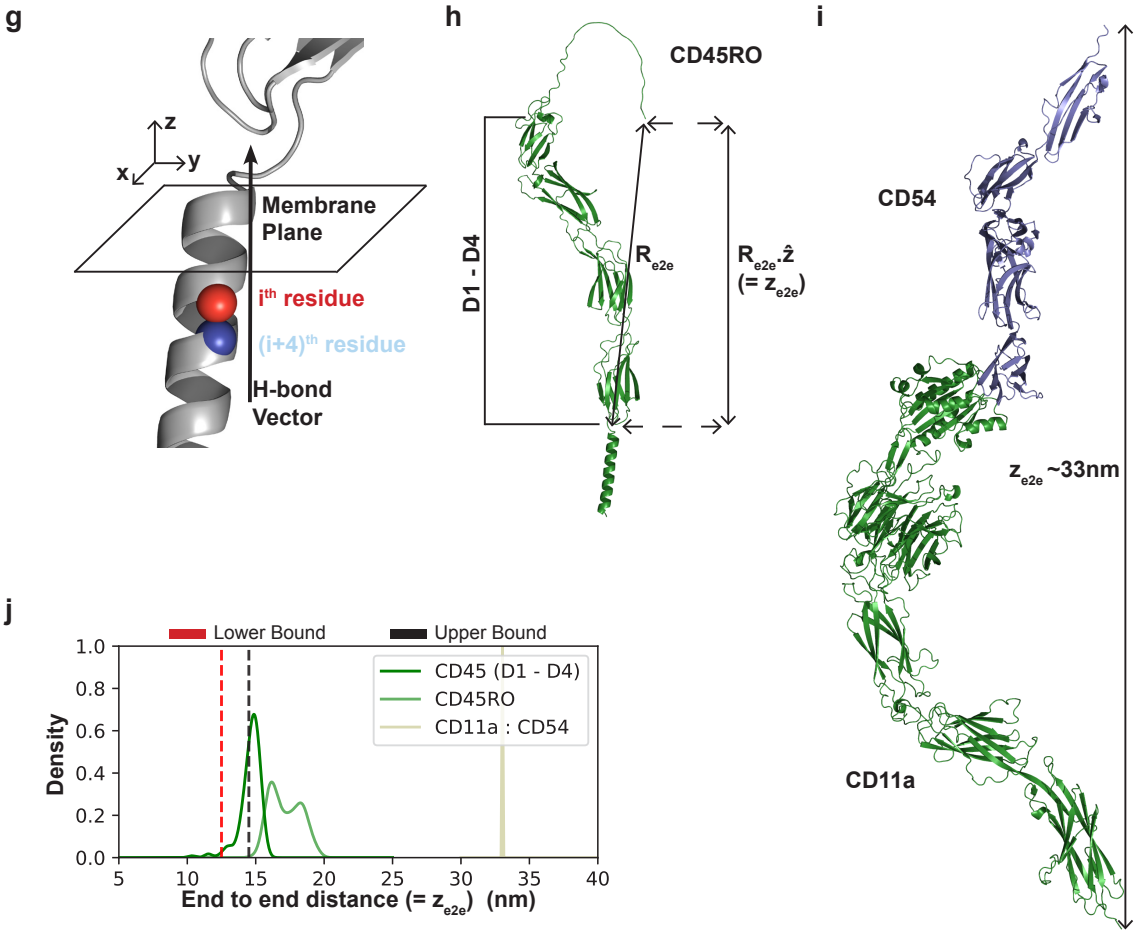
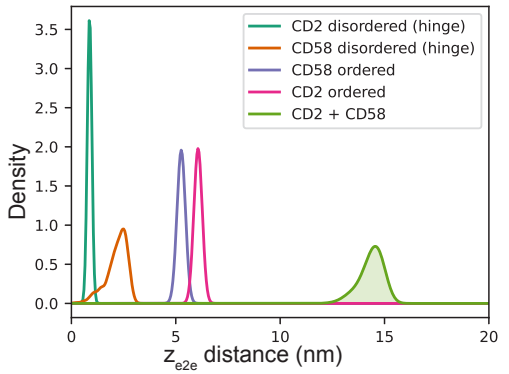
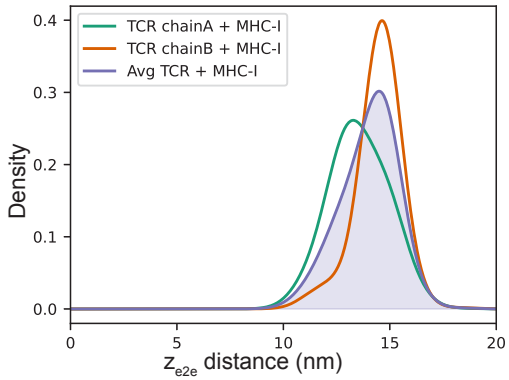

g
z
y
x
Membrane
Plane
i^th residue
(i+4)^th residue
H-bond
Vector
h
CD45RO
D1 - D4
R_e2e
R_e2e.ẑ
(= z_e2e)
i
CD54
z_e2e ~33nm
CD11a
j
Lower Bound
Upper Bound
CD45 (D1 - D4)
CD45RO
CD11a : CD54
Density
1.0
0.8
0.6
0.4
0.2
0.0
5
10
15
20
25
30
35
40
End to end distance (= z_e2e) (nm)

Figure H.3: Defining lower and upper geometric bounds of a functional immunological synapse. **g,** Geometric filtering strategy used for CD45 end-to-end distance calculations. The membrane normal vector is inferred from hydrogen bond patterns within the transmembrane α-helix, defining a boundary that excludes unphysical conformations arising from the absence of an explicit membrane in AF2 and enabling accurate calculation of the z_{e2e} distance. **h,** Structural model of CD45RO and definition of the CD45[D1-D4] domains. **i,** Structure of the large CD11a-CD54 (LFA-1/ICAM-1) protein complex, which is spatially segregated to the peripheral supramolecular activation cluster (pSMAC), outside the central synapse (cSMAC). **j,** Distributions of z_{e2e} distances for the ordered domains of CD45RO, CD45[D1-D4], and CD11a-CD54 complexes. The black line denotes the upper bound applied in this study ($z_{\mathrm{e2e}} < 14.5$ nm), and the red line denotes the lower bound ($z_{\mathrm{e2e}} > 12.5$ nm). Together, these bounds define the geometric constraints used to filter for structurally compatible protein complexes within a functional immunological synapse.

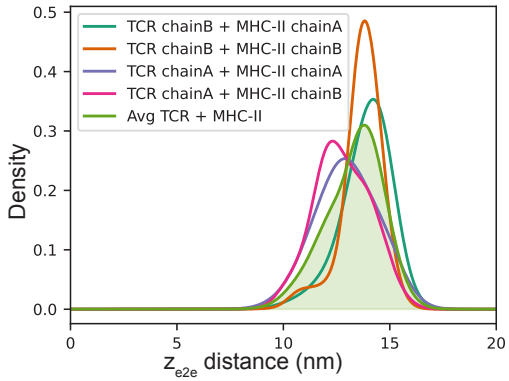
a



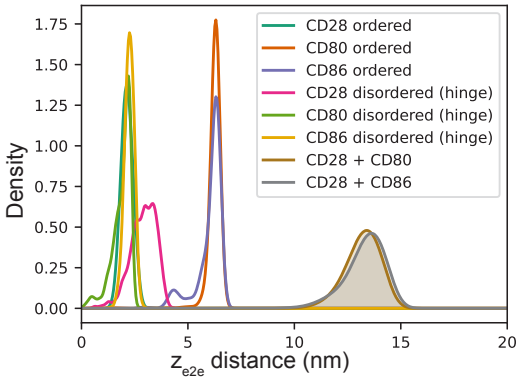
	Distribution Generator	MSA Parameters
CD2 - ordered	rMSA AF2	16:32
CD2 - disordered	rMSA AF2	16:32, 8:16
CD58 - ordered	rMSA AF2	16:32, 8:16
CD58 - disordered	rMSA AF2	64:128, 32:64



	Dist. Gen.	MSA Parameters
TCR α - ordered	rMSA AF2	16:32, 8:16
TCR α - disordered	rMSA AF2	16:32, 8:16
TCR β - ordered	rMSA AF2	16:32, 8:16
TCR β - disordered	rMSA AF2	16:32, 8:16
MHC-I - ordered	rMSA AF2	16:32, 8:16
MHC-I - disordered	rMSA AF2	16:32, 8:16

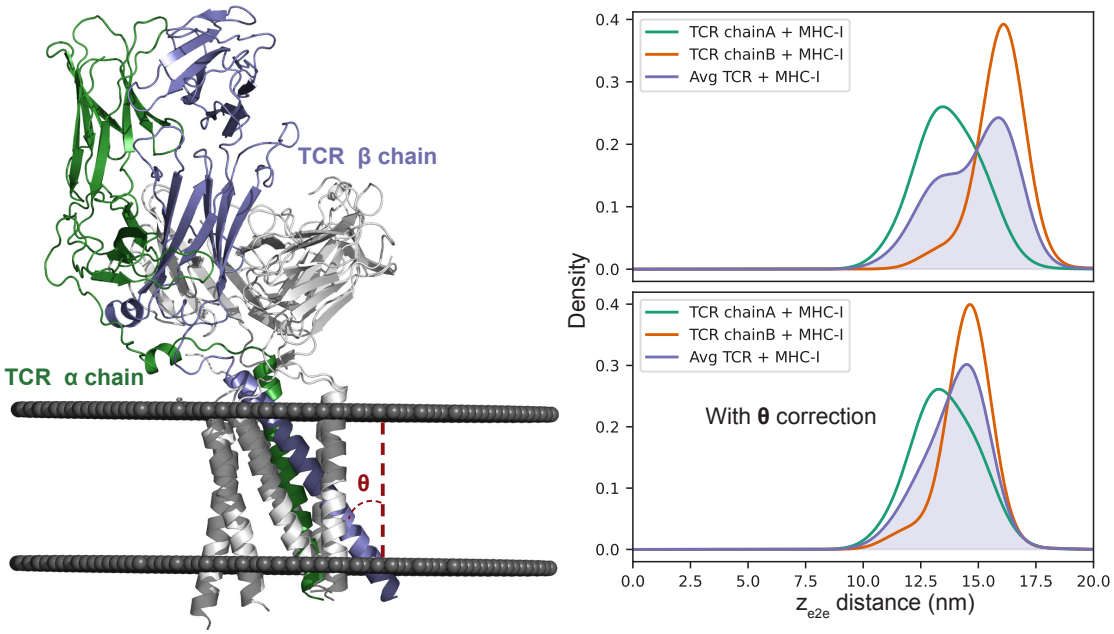


	Dist. Gen.	MSA Parameters
MHC-II chain A - ordered	rMSA AF2	16:32, 8:16
MHC-II chain A - disordered	rMSA AF2	16:32, 8:16
MHC-II chain B - ordered	rMSA AF2	16:32, 8:16
MHC-II chain B - disordered	rMSA AF2	16:32, 8:16



	Dist. Gen.	MSA Parameters
CD28 - ordered	rMSA AF2	16:32, 8:16
CD28 - disordered	rMSA AF2	16:32, 8:16
CD86 - ordered	rMSA AF2	16:32, 8:16
CD86 - disordered	rMSA AF2	16:32, 8:16
CD80 - ordered	rMSA AF2	16:32, 8:16
CD80 - disordered	rMSA AF2	16:32, 8:16

b


TCR β chain
TCR α chain
θ
TCR chainA + MHC-I
TCR chainB + MHC-I
Avg TCR + MHC-I
Density
With θ correction
z_{e2e} distance (nm)

c

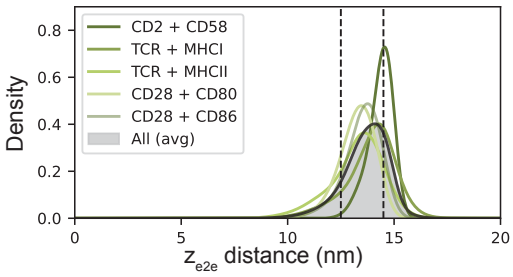

CD2 + CD58
TCR + MHCI
TCR + MHCII
CD28 + CD80
CD28 + CD86
All (avg)
Density
z_{e2e} distance (nm)

Figure H.4: Validation of fluctuation independence and synaptic bounds definitions using CAR-BOOST-like modeling of the canonical T-cell synapse.

a, Predicted distributions of synaptic z-distance for individual proteins included in the immunological synapse, assuming statistical independence of fluctuations across constituent domains. Tables show the methods used to generate distributions for each corresponding protein complex. For TCR-MHC I and TCR-MHC II complexes, plots illustrate how the computational model (Fig. 8.3c) accommodates multi-chain assemblies. The model incorporates the geometric filter defined in Fig. H.3g for the calculation of end-to-end (z_{e2e}) distance density distributions.**b,** Left: The orientation of the TCR complex is determined using the OPM server on PDB 6JXR[378]. The TCR β chain forms a θ angle with respect to the normal of the membrane plane. Thus, assuming the transmembrane helical axis is perpendicular to the membrane plane may lead to an overestimation of z_{e2e}. Right: Distribution of intermembrane distances. The top plot illustrates the effect of overestimation, which is corrected in the bottom plot. **c,** Density plot of the averaged z-distance distributions across all synapse-included proteins, providing qualitative validation of the lower and upper synaptic cleft distance bounds applied in this study.

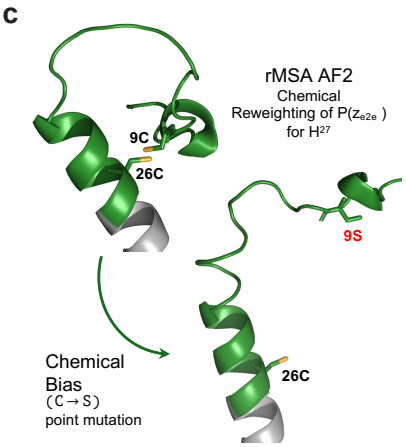
a

Hinge	#AA (Hinge)	Dist. Gen. Method	CD8α - derived Hinge Sequence
H^{3}	3	rMSA AF2	**SG**D
H^{5}	5	rMSA AF2	**SG**ACD
H^{8}	8	rMSA AF2	**SG**LDFACD
H^{15}	15	AF2RAVE	**SG**GAVHTRGDLFACD
H^{27}	27	AF2RAVE	**SG**SLRPEACRPAAGGAVHTRGLDFACD
H^{33}	33	AF2RAVE	**SG**IASQPLSLRPEACRPAAGGAVHTRGLDFACD
H^{40}	40	AF2RAVE	**SG**PPTPAPTIASQPLSLRPEACRPAAGGAVHTRGLDFACD
H^{47}	47	AF2RAVE	**SG**TTTPAPRPPTPAPTIASQPLSLRPEACRPAAGGAVHTRGLDFACD
H^{51}	51	AF2RAVE	**SGGGGS**TTTPAPRPPTPAPTIASQPLSLRPEACRPAAGGAVHTRGLDFACD
H^{56}	56	AF2RAVE	**SGGGGSGGGGS**TTTPAPRPPTPAPTIASQPLSLRPEACRPAAGGAVHTRGLDFACD
H^{61}	61	AF2RAVE	**SGGGGSGGGGSGGGGS**TTTPAPRPPTPAPTIASQPLSLRPEACRPAAGGAVHTRGLDFACD
H^{66}	66	AF2RAVE	**SGGGGSGGGGSGGGGSGGGGS**TTTPAPRPPTPAPTIASQPLSLRPEACRPAAGGAVHTRGLDFACD
H^{71}	71	AF2RAVE	**SGGGGSGGGGSGGGGSGGGGSGGGGS**TTTPAPRPPTPAPTIASQPLSLRPEACRPAAGGAVHTRGLDFACD
H^{76}	76	AF2RAVE	**SGGGGSGGGGSGGGGSGGGGSGGGGSGGGGS**TTTPAPRPPTPAPTIASQPLSLRPEACRPAAGGAVHTRGLDFACD
H^{91}	91	AF2RAVE	**SGGGGSGGGGSGGGGSGGGGSGGGGSGGGGSGGGGSGGGGSGGGGS**TTTPAPRPPTPAPTIASQPLSLRPEACRPAAGGAVHTRGLDFACD

b

#AA (Hinge)	Chemical Bias Description (Point Mutation)
27	C9S
33	C15S
40	C22S
47	C29S
51	C33S
56	C38S
61	C43S
76	C58S

c



d

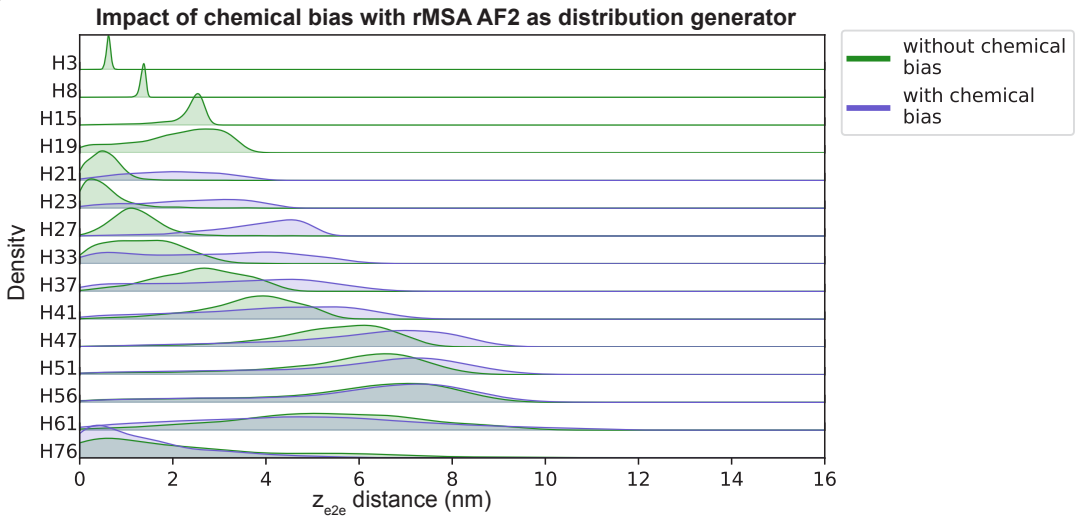
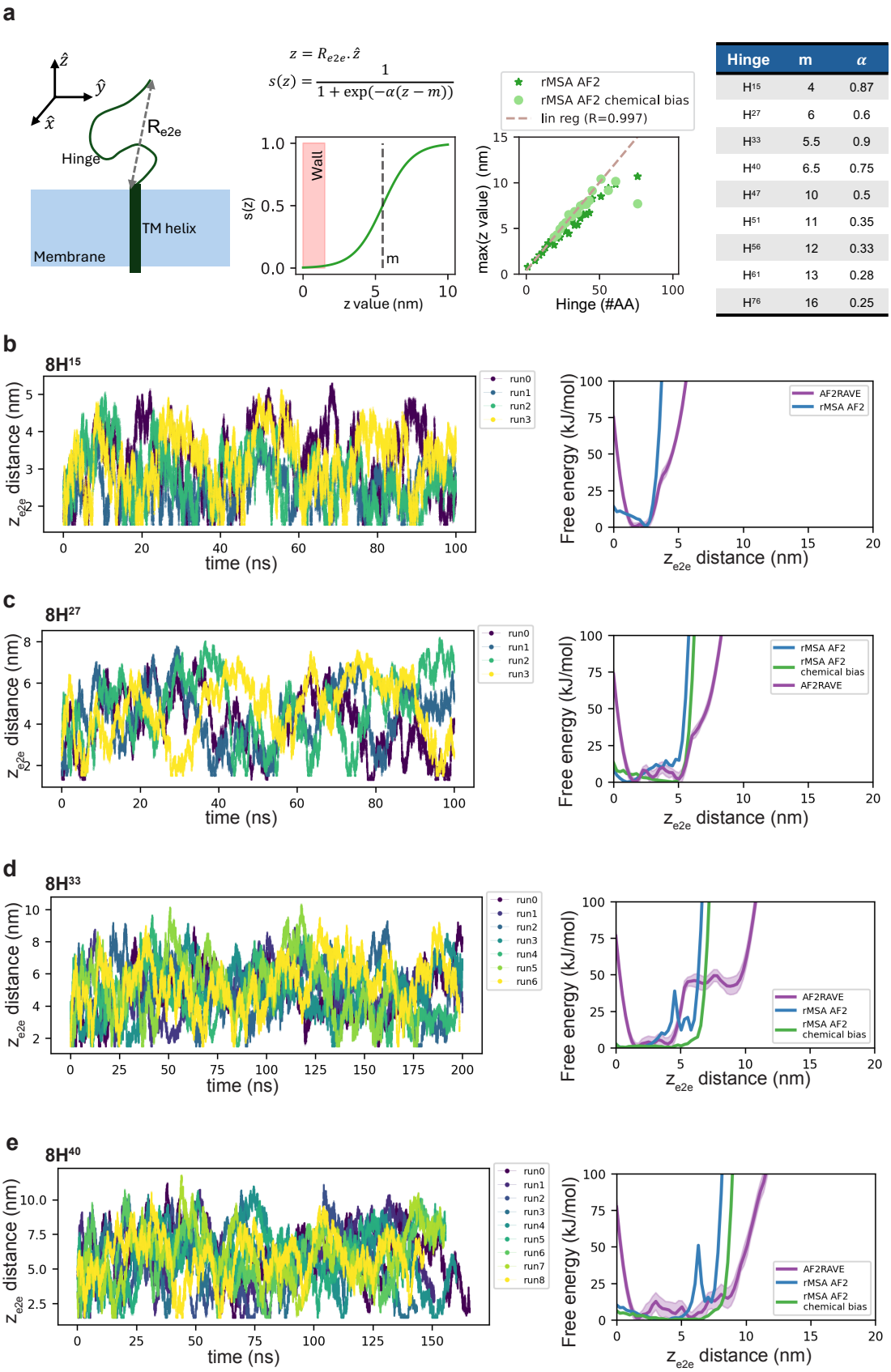
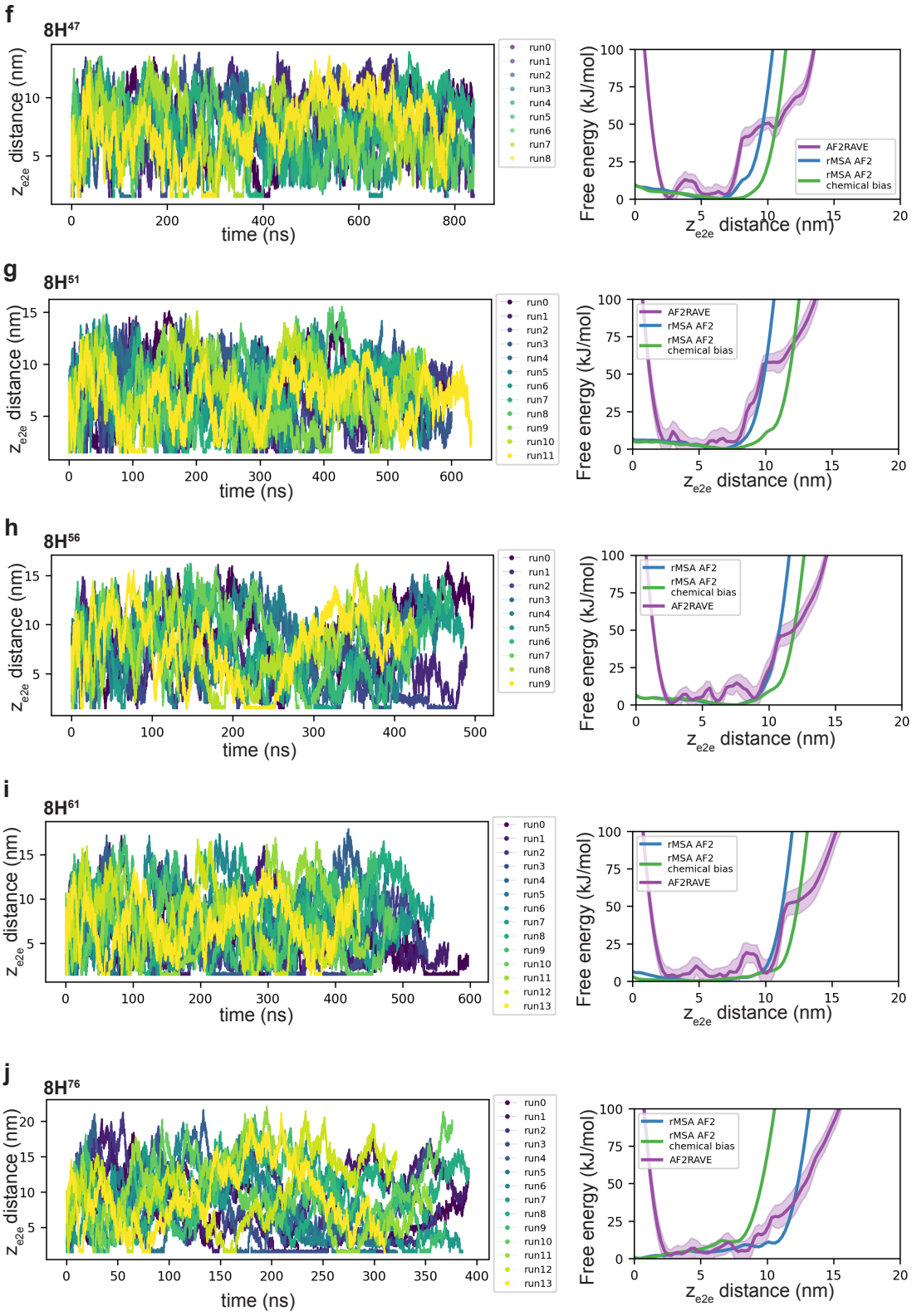


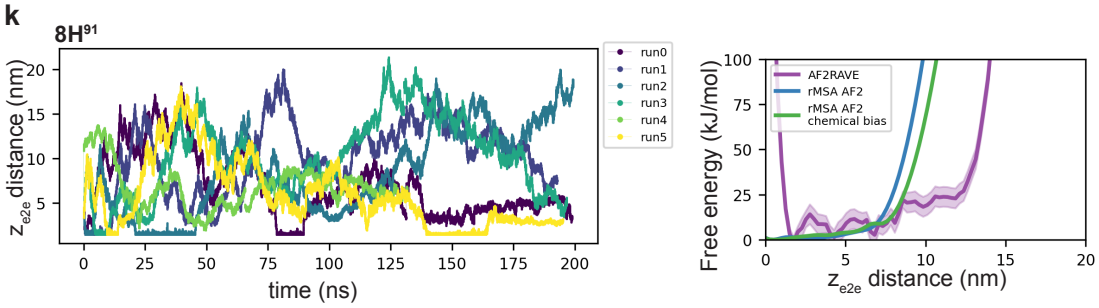
Figure H.5: Reduced-MSA AF2 performance on CD8α-derived hinge sequences.
a, CD8α-derived hinge sequences used in this study, together with the corresponding methods used to generate their conformational distributions, are shown. SG-based linker regions are indicated in bold. **b,** The incorporation of chemical bias on AF2 model outputs was evaluated as a function of point mutations. **c,** Schematic illustrating the structural effects of chemical bias (point mutation) on AF2 model outputs is shown. **d,** Illustration of AF2 bias toward disulfide bond formation within intrinsically disordered protein regions is shown. Application of chemical bias results in a shift in the predicted distance distribution (blue). The panel also defines the reweighting strategy used to account for chemical bias. The model incorporated the geometric filter defined in Fig. H.3g for calculation of the end-to-end (z_{e2e}) distance density distributions.



Hinge	m	α
H15	4	0.87
H27	6	0.6
H33	5.5	0.9
H40	6.5	0.75
H47	10	0.5
H51	11	0.35
H56	12	0.33
H61	13	0.28
H76	16	0.25


f
8H47
g
8H51
h
8H56
i
8H61
j
8H76
time (ns)
z_e2e distance (nm)
Free energy (kJ/mol)
z_e2e distance (nm)
run0
run1
run2
run3
run4
run5
run6
run7
run8
run9
run10
run11
run12
run13
AF2RAVE
rMSA AF2
rMSA AF2 chemical bias

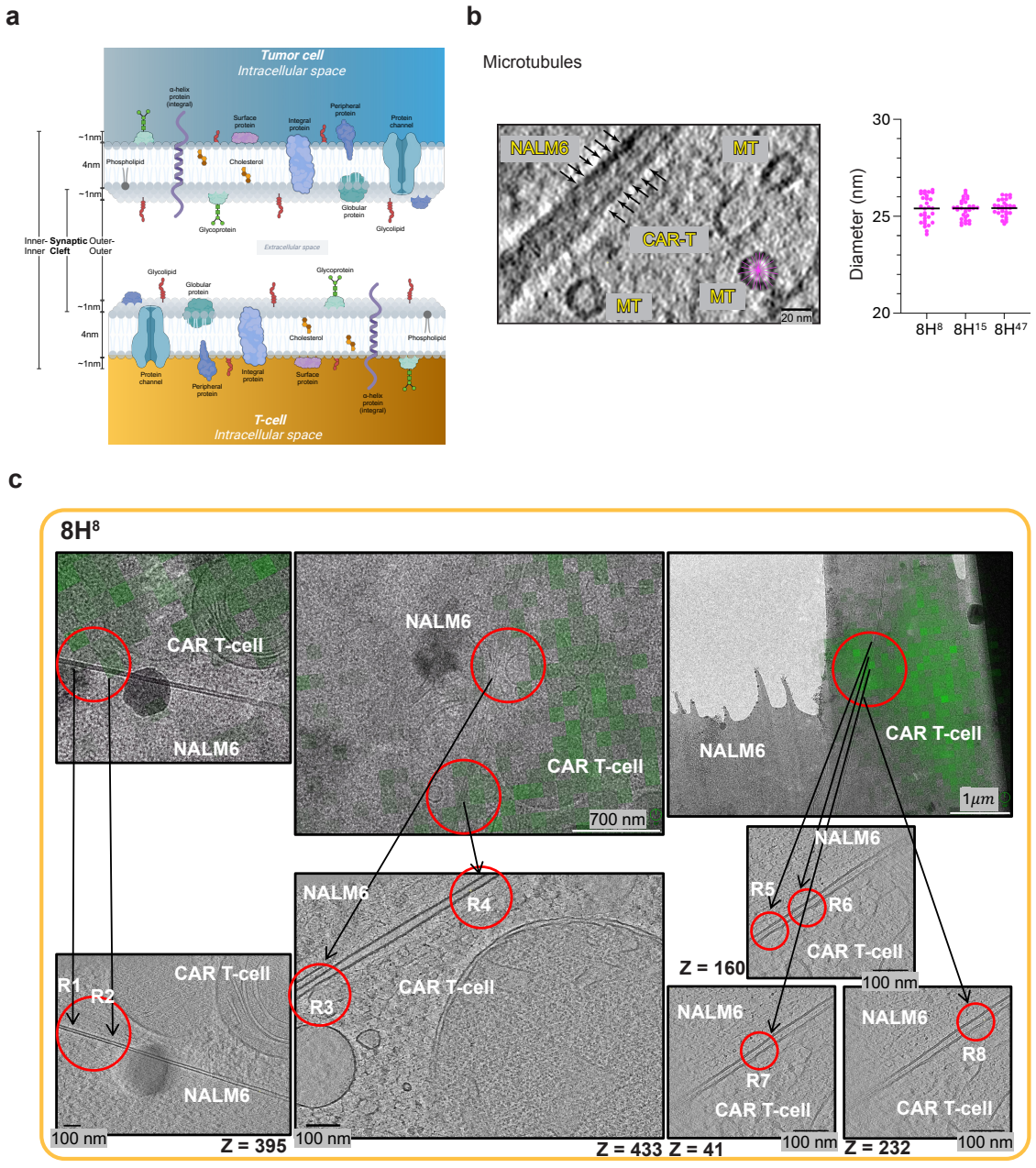
k



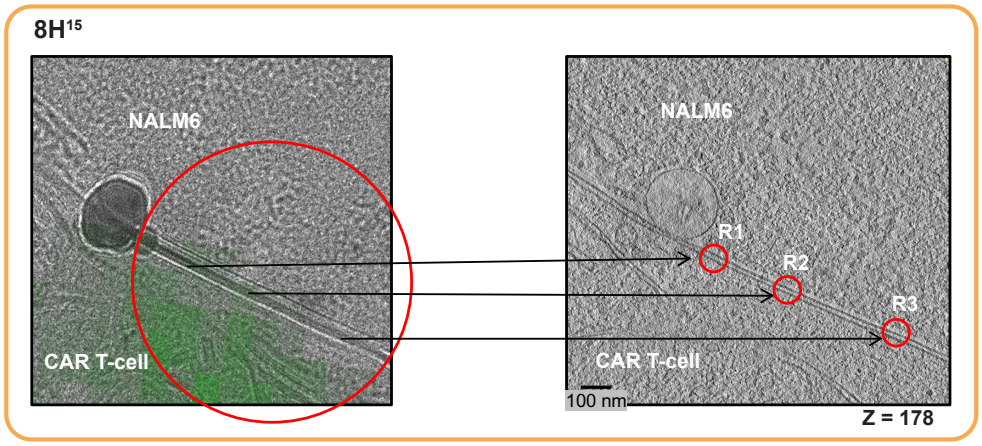
l

CAR Hinge	Number of runs (#)	Total simulation time (ns)	Avg. number of transitions per run (#)	Runs with >2 transitions (#)
8H^{15}	4	0.4 x 1e3	10	4
8H^{27}	4	0.4 x 1e3	7	4
8H^{33}	7	1.4 x 1e3	10	7
8H^{40}	9	1.4 x 1e3	8	9
8H^{47}	9	7.4 x 1e3	7	9
8H^{51}	12	6.8 x 1e3	7	12
8H^{56}	10	4.6 x 1e3	6	10
8H^{61}	14	7.0 x 1e3	5	12
8H^{76}	14	5.0 x 1e3	5	12
8H^{91}	6	1.1 x 1e3	3	4
28H^{42}	3	0.4 x 1e3	3	3
IgG4 H^{15}	11	2.2 x 1e3	18	11

Figure H.6: AlphaFold2-RAVE as a distribution generator for CD8α-derived hinge sequences.
a, Schematic representation of the simulation box, including definition of the biasing coordinate, $s(z)$, used for well-tempered metadynamics simulations. The table (right) lists simulation parameters for each CD8α-derived hinge. **b-k,** Enhanced sampling time series and corresponding free-energy profiles calculated from all replicates of 8H15, 8H27, 8H33, 8H40, 8H47, 8H51, 8H56, 8H61, 8H76, and 8H91 hinge variants. Time series traces demonstrate reversible transitions between extended and collapsed conformations, reflected as back-and-forth motion in end-to-end (z_{e2e}) distance. **l,** Summary table assessing the quality of well-tempered metadynamics simulations, including total simulation time, average number of transitions between extended and collapsed conformations per simulation run, and the number of trajectories exhibiting more than two such transitions.


a
Tumor cell
Intracellular space
α-helix protein (integral)
Surface protein
Integral protein
Peripheral protein
Protein channel
~1nm
4nm
~1nm
Phospholipid
Cholesterol
Glycoprotein
Globular protein
Glycolipid
Extracellular space
Inner-Inner
Synaptic Cleft
Outer-Outer
Glycolipid
Globular protein
Glycoprotein
~1nm
4nm
~1nm
Cholesterol
Phospholipid
Protein channel
Peripheral protein
Integral protein
Surface protein
α-helix protein (integral)
T-cell
Intracellular space
b
Microtubules
NALM6
MT
CAR-T
MT
MT
20 nm
Diameter (nm)
30
25
20
8H8
8H15
8H47
c
8H8
CAR T-cell
NALM6
NALM6
CAR T-cell
700 nm
NALM6
CAR T-cell
1μm
NALM6
R5
R6
CAR T-cell
Z = 160
100 nm
NALM6
R4
CAR T-cell
R3
CAR T-cell
R1
R2
NALM6
100 nm
Z = 395
100 nm
Z = 433
Z = 41
NALM6
R7
CAR T-cell
100 nm
NALM6
R8
CAR T-cell
Z = 232
100 nm

d


8H15
NALM6
CAR T-cell
NALM6
R1
R2
R3
CAR T-cell
100 nm
Z = 178

e

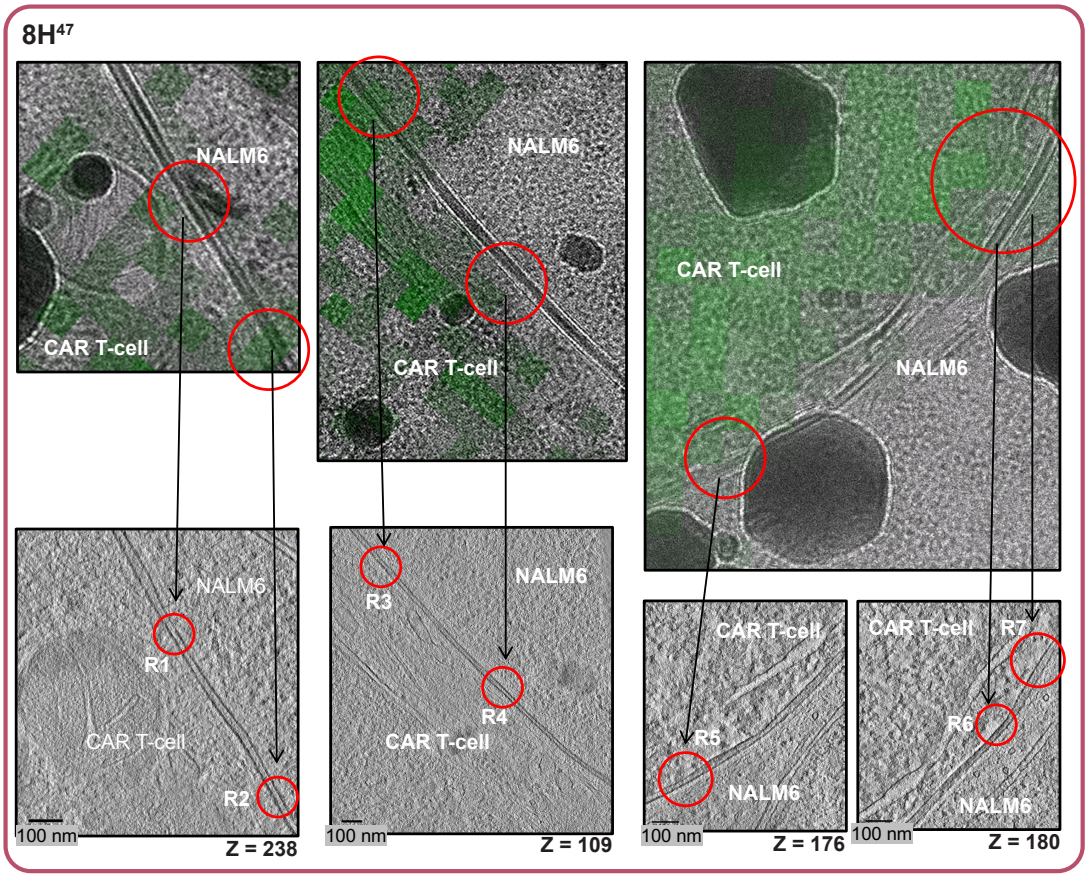
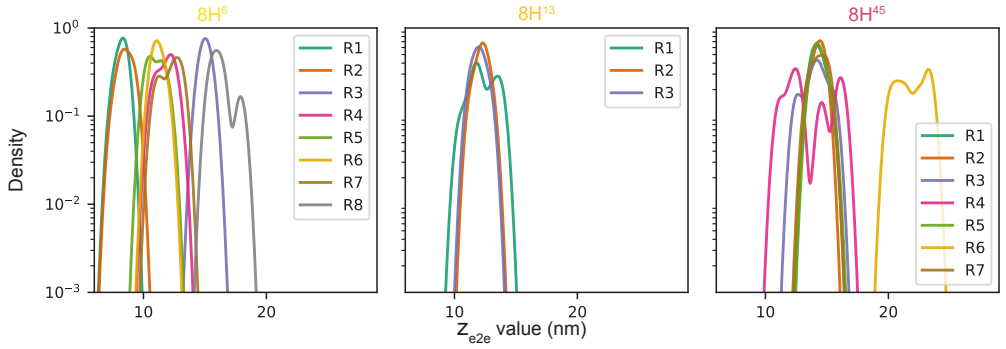
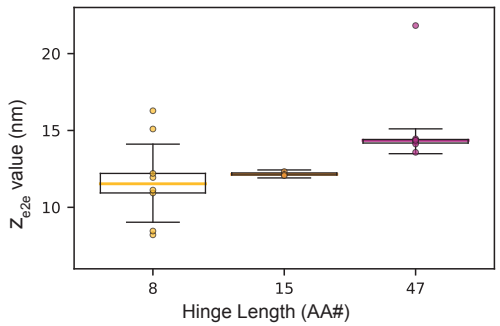

8H47
NALM6
CAR T-cell
NALM6
CAR T-cell
CAR T-cell
NALM6
NALM6
R1
CAR T-cell
R2
100 nm
Z = 238
NALM6
R3
R4
CAR T-cell
100 nm
Z = 109
CAR T-cell
R5
NALM6
100 nm
Z = 176
CAR T-cell
R7
R6
NALM6
100 nm
Z = 180

Figure H.7: Correlated fluorescence light microscopy and cryo-ET images of CAR T-cell synapses. **a,** Schematic illustration of the phospholipid bilayers and proteinaceous content at the interface between a T-cell and a tumor cell. **b,** Region from a reconstructed cryo-ET tomographic slice showing microtubules within a CAR T-cell expressing an 8H8 CD19 CAR construct. Distances were measured across the microtubule (MT) diameter, as indicated by pink lines (scale bar, 20 nm, lower left). Quantification of MT diameter measurements in T-cells expressing 8H8, 8H15, and 8H47 CD19 CAR constructs is shown; averaged values are represented by black lines. **c,** Top: Overlay of integrated fluorescence light microscope (iFLM) and cryo-ET images of a T-cell expressing a CAR with an 8H8 hinge in close proximity to a $CD19^+$ NALM6 leukemic cell (red circles). Bottom: Corresponding regions from which intermembrane distances were sampled (red circles). **d,** Left: Overlay of iFLM and cryo-ET images of a T-cell expressing a CAR with an 8H15 hinge in close proximity to a $CD19^+$ NALM6 leukemic cell (large red circle). Right: Regions from which intermembrane distances were sampled (small red circles). **e,** Top: Overlay of iFLM and cryo-ET images of a T-cell expressing a CAR with an 8H47 hinge in close proximity to a $CD19^+$ NALM6 leukemic cell (red circles). Bottom: Regions used for intermembrane distance sampling (red circles).

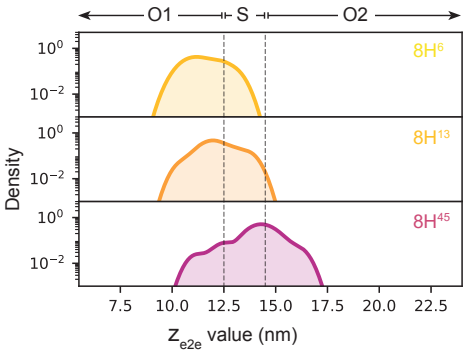
a



b



c



d

CAR	Outer to Outer	Inner to Inner
8H[8]	9.54 ± 0.82	21.60 ± 0.76
8H[15]	10.23 ± 0.83	22.21 ± 0.67
8H[47]	12.16 ± 0.95	24.21 ± 0.87

CAR_1 (q)	CAR_2 (p)	KL-Divergence ($D_{KL}(p\|\|q)$)
8H[8]	8H[15]	**18.464**
8H[15]	8H[47]	**214.612**
8H[8]	8H[47]	**286.663**

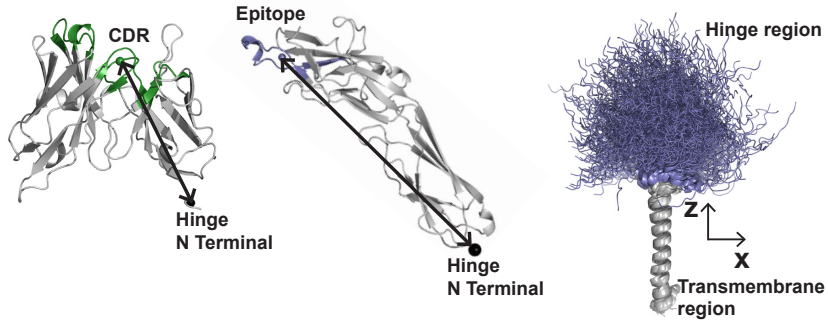
Figure H.8: Synaptic cleft distance calculation from cryo-ET images.

a, Kernel density estimation (KDE) performed on each sampled region shown in Fig. H.7. The presence of a lower peak at a distance of 8.5 nm (Fig. H.7c, R1 and R2) was observed exclusively for the 8H8 CAR, but not for the 8H15 or 8H47 CARs, suggesting additional kinetic segregation of costimulatory proteins in the context of an extremely short linker. These data highlight the requirement for a linker length that permits synaptic cleft dimensions that exceed the lower geometric bound necessary for CAR activation. **b,** Interquartile range (IQR) plots of mean synaptic cleft distances for each sampled region, used to identify and exclude outlier regions. **c,** Synaptic cleft distance calculations were performed on regions retained after IQR filtering in panel b. **d,** Summary table reporting outer-to-outer and inner-to-inner membrane distances (mean $\pm$ standard deviation), as defined in the Methods, for each CAR construct (left). Corresponding Kullback-Leibler divergence values comparing distance distributions are shown (right).

a

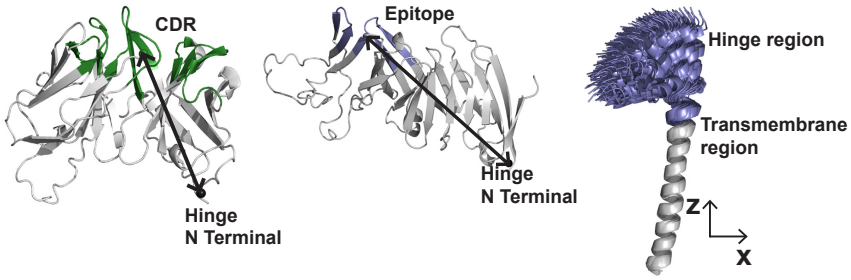
CD33

scFv	AF2
Ig domain	AF2
Hinge region (Target)	rMSA AF2



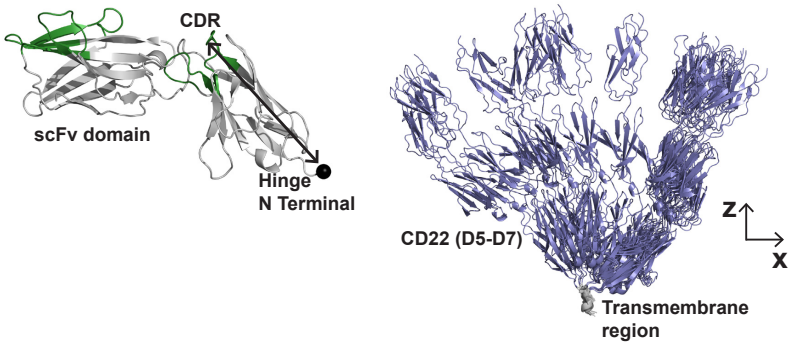
CD19

scFv	AF2
Ig domain	AF2
Hinge region (Target)	rMSA AF2



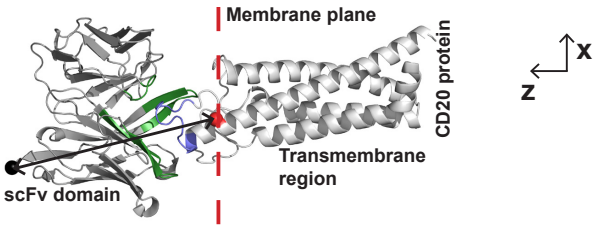
CD22

scFv	AF2
Ig domain	rMSA AF2
Hinge region (Target)	-



CD20

scFv + Ig domain	AF multimer
Hinge region (Target)	-



b

Offset values for different targets

	CD33 + scFv (nm)	CD19 + scFv (nm)	CD22 + scFv (nm)	CD20 + scFv (nm)
Offset values (nm)	11.7	8.3	4	4.9

c

KDE bandwidth values for different hinges

	H^{3}	H^{8}	H^{15}	H^{27}	H^{33}	H^{40}	H^{47}	H^{51}	H^{56}	H^{61}	H^{76}	H^{91}	$CD28H^{42}$	$IgG4H^{15}$
CD33	0.8	0.8	0.8	0.2	-	-	0.1	-	-	0.1	0.1	-	-	-
CD19	0.8	0.8	0.8	0.2	-	-	0.1	-	-	0.1	0.1	0.5	0.5	0.6
CD22	0.8	0.8	0.8	0.2	0.3	0.3	0.2	-	-	0.1	0.1	-	-	-
CD20	-	-	0.8	-	-	-	0.2	0.2	0.2	0.2	-	-	-	-

Figure H.9: Target protein modeling using CAR-BOOST.

a, Descriptions of the modeling strategies used for structure and ensemble generation of the CD33, CD19, CD22, and CD20 target antigens. CD33 and CD19 were modeled using the same CAR-BOOST strategy described in Fig. H.4 and main-text Fig. 8.3c, as both proteins contain a disordered hinge-like region coupled to an ordered Ig domain targeted by the corresponding scFv. For CD22, the three membrane-proximal ordered domains (D5-D7) were generated using reduced-MSA AF2 (rMSA-AF2) due to the absence of an intrinsically disordered region. The model includes the first six residues of the transmembrane region to enable calculation of the normal to the membrane. Ten representative structures were randomly selected from a total of 640 generated models for visualization. For CD20, where the CAR scFv targets the membrane-proximal extracellular loop, AF-multimer was used to calculate scFv-target spatial offsets. The red line indicates the membrane plane. **b,** Table listing the final offset values contributed by the ordered domains of each target antigen and the corresponding scFv domain. **c,** Table listing the bandwidth parameters used in CAR-BOOST for KDE in free-energy calculations.

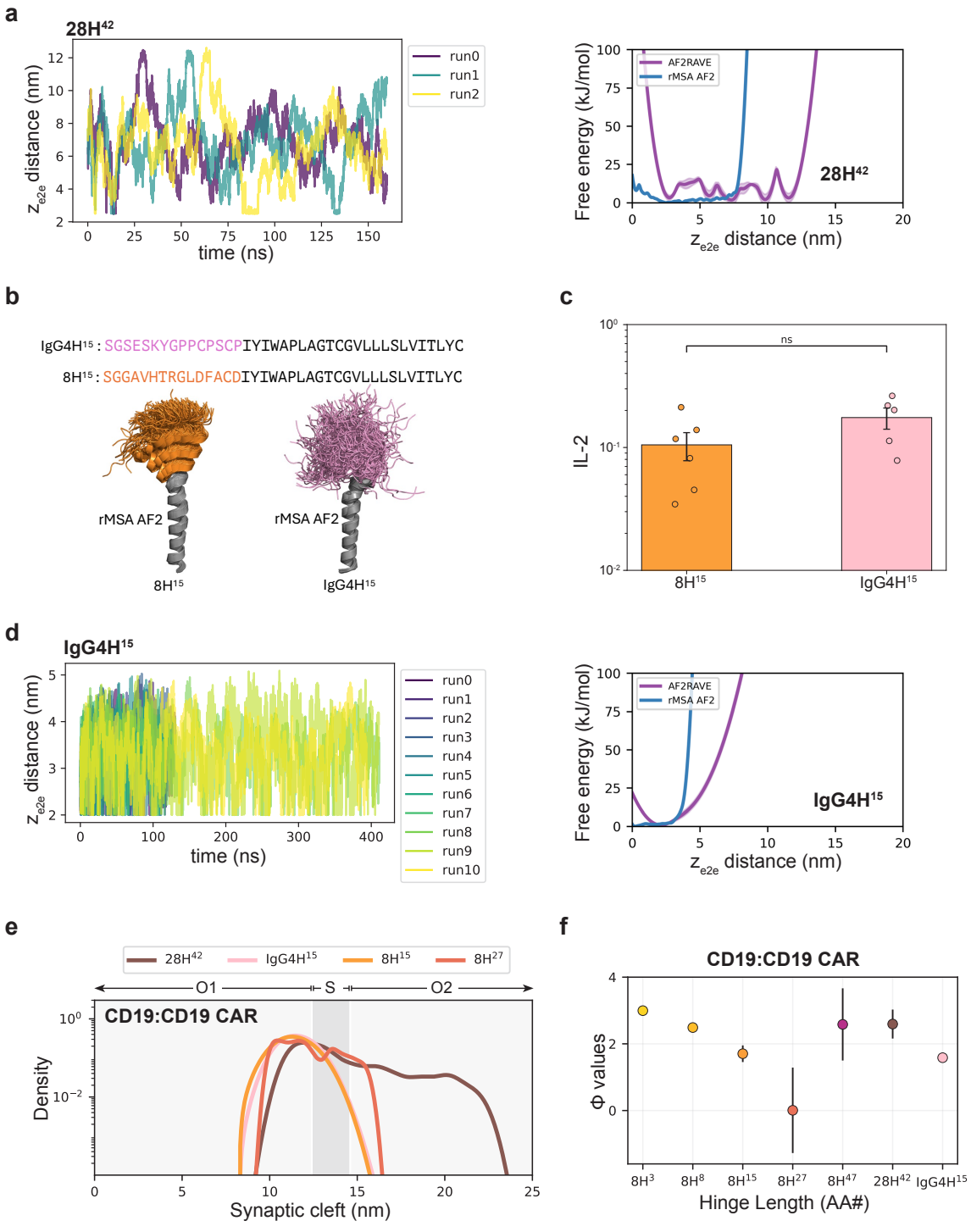

a
28H42
run0
run1
run2
z_e2e distance (nm)
time (ns)
AF2RAVE
rMSA AF2
Free energy (kJ/mol)
z_e2e distance (nm)
28H42
b
IgG4H15 : SGSESKYGPPCPSCPIYIWAPLAGTCGVLLLSLVITLYC
8H15 : SGGAVHTRGLDFACDIYIWAPLAGTCGVLLLSLVITLYC
rMSA AF2
8H15
rMSA AF2
IgG4H15
c
ns
IL-2
8H15
IgG4H15
d
IgG4H15
run0
run1
run2
run3
run4
run5
run6
run7
run8
run9
run10
z_e2e distance (nm)
time (ns)
AF2RAVE
rMSA AF2
Free energy (kJ/mol)
z_e2e distance (nm)
IgG4H15
e
28H42
IgG4H15
8H15
8H27
O1
S
O2
CD19:CD19 CAR
Density
Synaptic cleft (nm)
f
CD19:CD19 CAR
Φ values
8H3
8H8
8H15
8H27
8H47
28H42
IgG4H15
Hinge Length (AA#)

Figure H.10: Modeling non-CD8α-derived hinges for CAR targeting CD19.
a, Enhanced sampling time series and corresponding free-energy profiles calculated from all replicates of the 28H42 hinge variant. Time series traces demonstrate reversible transitions between extended and collapsed conformations, reflected as back-and-forth motion in end-to-end (z_{e2e}) distance. **b,** Sequence and structure comparison of two hinges (IgG4H15 and CD8H15) that contain the same number of residues with different amino acid composition. **c,** IL-2 production by T-cells transduced with CD19 CAR constructs harboring 15-residue CD8α (8H15) or IgG4 (IgG4H15) domains following co-culture with NALM6 cells ($\sim 1 \times 10^3$ CD19 molecules/cell). Mean IL-2 levels are presented ($n = 5$ individual donors for IgG4H15 and $n = 6$ individual donors for 8H15). **d,** Enhanced sampling time series and corresponding free-energy profiles calculated from all replicates of the IgG4 hinge variant. **e,** Density plots of the synaptic cleft formed by the CD19:CD19 CAR complex. **f,** The ϕ values calculated for the CARs targeting CD19-expressing tumor cells. A set of seven distinct hinge variants derived from CD8α, CD28, and IgG4 was used to construct the CARs described above. The ϕ values agree with the CAR T-cell response reported in Fig. 8.2c,h and in panel c.

Appendix I: Supporting Information for Chapter 9

I.1 Ensemble generating methods

I.1.1 Machine-learned Protein Structural Ensemble Generation

Here are the details of the methods used to generate structural ensembles from all the machine learning methods I utilize.

AlphaFold2[1, 11]: AlphaFold2 structures are generated at the specified MSA depth using LocalColabFold (`https://github.com/YoshitakaMo/localcolabfold`). For each sequence, I use three recycles, 64 seeds, and dropout. I raster the MSA from 8:16 until 512:1024 ("full MSA") on a base-2 logarithmic scale. The structures are ranked using the pLDDT score. The input script I use to generate AF2 structures is given in the GitHub repository affiliated with this manuscript.

AF-Cluster[95]: Structures are generated using the Colab notebook present in the GitHub repository (`https://github.com/HWaymentSteele/AF_Cluster`) with the appropriate HIV-1 protease sequence inserted instead. I use the mmseqs2 method to generate the MSA with pair_mode set to unpaired_paired. I set the number of copies equal to 1; I specify that the minimum number of samples in a cluster is 10; and I do not use any templates.

DiG[272]: I use the default parameters specified in the original manuscript and on the project's GitHub repository (`https://github.com/microsoft/Graphormer/blob/main/distributional_graphormer/protein/run_inference.py`), except that I change the input files to 1) a Python pickle file containing the full MSA and 2) the FASTA file containing the HIV-1 protease sequence of interest. As with AlphaFold2, I specify that DiG output 320 structures during the inference step.

BioEmu[244]: I run BioEmu with the default parameters specified in its GitHub repository (`https://github.com/microsoft/bioemu`), except I demand it generate 320 structures in the inferred ensemble. This number of structures is the upper bound for the number in the final ensemble, as BioEmu then proceeds to pass the structures through a filtering procedure removing physically unallowable configurations from the output.

The six HIV-1 protease sequences are taken from the following PDB identifiers: 3TTP[292], 2PC0[293], 1Q9P[379], 1EBW[380], 4Z4X[294], and 6P9A[295]. I choose to simulate the HIV-1 protease monomer because 1) it is one of the test systems in Ref. [262] and 2) it is prone to variant selection due to resistance mutations, which also have notable effects on protease's structure and dynamics.[289, 296, 381, 382, 383, 384] While functionally active only in the dimerized form, I believe that studying the monomer in isolation can still yields insights regarding the protease's functional behavior.

MD simulations of the dimers were run using the same protocol as the monomer simula-

tions. The starting structure for the WT dimer is PDB ID: 1HHP[300] and PDB ID: 3UF3[301] for the mutant. To generate AF2 equivalent structures of the dimers, I utilize AlphaFold-multimer (AFmultimer)[184]. I use the ColabFold implementation (`https://github.com/sokrypton/ColabFold`).

I.1.2 Generating Ground-Truth Conformational Ensembles Using Atomistic Molecular Dynamics

To generate a reference conformational ensemble encoding dynamics of each HIV-1 protease variant, I perform atomistic, explicit solvent, classical MD simulations using the OpenMM package[341] with the amber99sb-ildn[344] and the TIP3P water model.[345] Each simulation is started using the AF2 structure with the highest predicted confidence. The protein is solvated, ionized until neutrality, equilibrated for 300 ps in the NVT ensemble with position restraints, 300 ps in the NPT ensemble with position restraints, and 300 ps in the NPT ensemble without position restraints; all the equilibration steps are performed at a temperature of 300 K. Finally, a one microsecond production run in the NPT ensemble is performed at 300 K, which is then analyzed, including the calculation of ACFs. For the analysis at 10 and 100 ns, the first 10 and 100 ns of the simulation are extracted and analyzed, respectively. For all MD simulation runs, I use the Langevin integrator. The temperature is controlled using a Noseé-Hoover thermostat[385, 386] in all cases and, for the NPT simulations, the pressure is controlled using the Parinello-Rahman barostat.[387] Detailed scripts for reproducing the OpenMM simulations, including the energy minimization and equilibration steps, are given in the GitHub repository for this manuscript (`https://github.com/erb24/af2-dynamics`).

I.2 Relaxation Timescales as a Function of Sequence

Figure I.1 is the corollary to Figure 2 in the main text, where the timescales for each of the six sequences examined are plotted as a function of MSA depth input to AlphaFold2. In Figure I.1, I plot the timescales at four MSA depths (depths of 8, 16, 64, and 512) for each sequence separately. Plotting the timescales in this manner, it is clear how, independent of sequence, reducing the MSA depth almost universally increases the relaxation timescale of a given residue. Of course, the detailed behavior of the timescales remains a function of both primary sequence and MSA depth. As a comparison, I plot in Figure I.2 the integrated relaxation timescales for each sequence at different MD simulation lengths: 10, 100, and 1000 ns. I see no significant timescale separation as a function of sequence for any of the three simulation lengths probed. As with the MSA depth, the only consistent trend I see is that, on average, the integrated relaxation timescale increases with an increase in simulation length.

I.3 Increasing the Number of Samples Has Little Effect in Predicted Timescales

In this work, I show that changes in MSA depth alter the timescales of the samples. Here, I study the relationship between the number of samples and relaxation timescales. To this end, I examine the sequences for 4Z4X and 2PC0. I generated 5000 (5k), 1000 (1k), and 640 samples using the ColabFold implementation of AlphaFold2 at an MSA depth of 8. First, to evaluate the

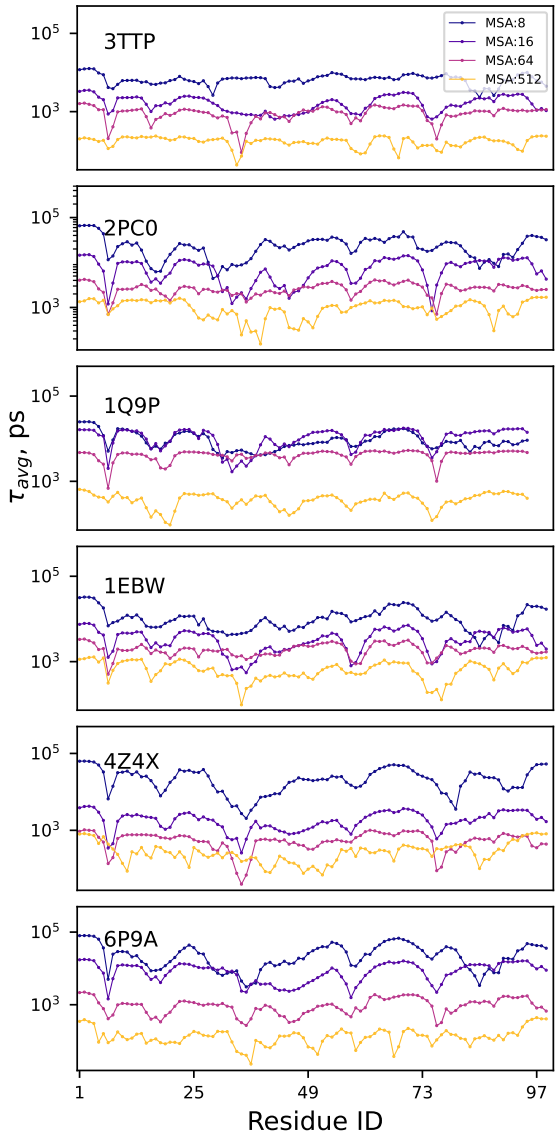


Figure I.1: Timescales at each MSA depth studied in the main text as a function of sequence. Each subpanel is labeled according to the sequence whose timescales it describes.

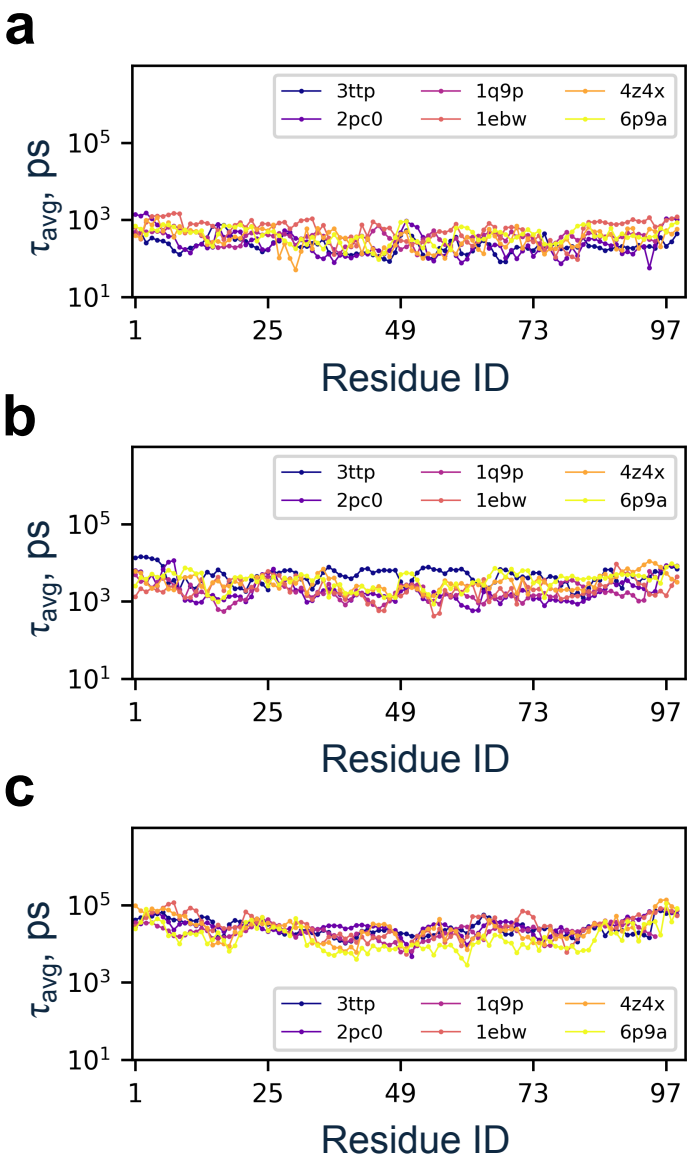


Figure I.2: Integrated relaxation timescales for each of the six HIV-1 protease variants examined in the main text as function of MD simulation length. Specifically, I probe the relaxation timescales at lengths of a) 10 ns, b) 100 ns, and c) 1000 ns.

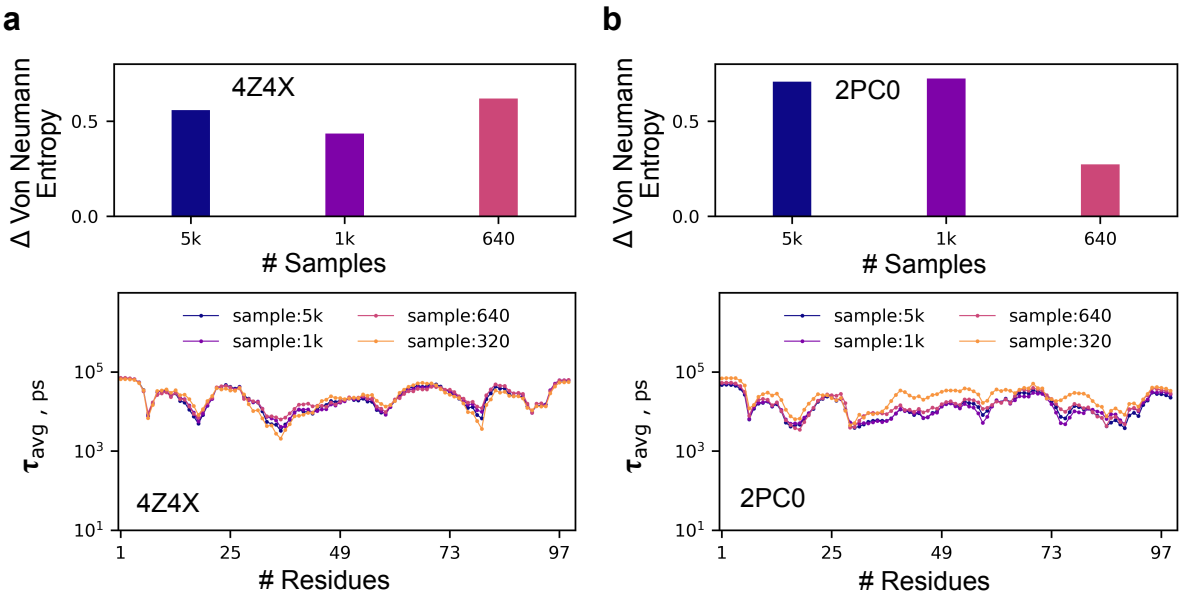


Figure I.3: Change in the information contained by the ensemble as the number of samples increases. a) (Top) shows the change in von Neumann entropy in the mean-centered covariance matrix calculated for 5000(5k), 1000(1k), and 640 samples generated independently for the sequence of 4Z4X PDB. The change is calculated with the covariance matrix calculated for 320 samples as reference. (Bottom) Displays the backbone relaxation timescales calculated using LE4PD. b) shows the result for the sequence belonging to 2PC0 PDB.

quality of the ensemble and its effect on the ensemble's mean-centered covariance matrix, I used the von Neumann entropy of the matrix. The von Neumann entropy is defined as[388]

$$s_v = -\Sigma_i e_i log(e_i),$$

where e_i is the sorted and normalized eigenvalue of the covariance matrix. Intuitively, this quantity suggests the distribution of eigenvalues, i.e., the separation of timescales. As per the construct of the entropy equation, the s_v is maximum when the eigenvalue distribution or spectrum is diffuse, i.e., when there are more states. Thus, increasing the number of samples should increase s_v if more information is being added. I see a similar trend in the top panel of Figure I.3a and b. In the top panel of Figure I.3a and b, I plotted the difference between the s_v at the mentioned sample size and s_v at 320 sample size that is being used in this work. While there is some trend seen in the case of 2PC0, overall the change is minimal.

To further analyze this effect, I applied my protocol to these samples to calculate the alpha-carbon relaxation timescales. The bottom panel of Figure I.3 qualitatively shows that indeed the order of change in timescale by increasing the number of samples is very minimal. Specifically, compared with the order of change when reducing the MSA information, it is clear that the empirically chosen sample size of 320 contains the required information for this study. However, looking forward, one has to consider more samples, or adaptive MD samples, from a representa-

tive set of these structures to apply this method to larger, more complicated systems.

I.4 LE4PD Analysis for the Slow Modes of Each HIV-1 Protease Monomer Variant

In Figure I.4 I show the local mode lengthscale (LML) or the projection of the root-mean-squared fluctuations along a single mode for the first two LE4PD modes (the ones used to construct the two-dimensional free-energy surfaces in the main text) as a function of sequence and AF2 MSA depth. For all sequences, the first LE4PD mode describes mostly fluctuations in the N-terminal tail (Figure I.4a) while the second mode features a relaxation time roughly an order of magnitude faster than the first mode, and is a much stronger function of sequence. Since the N-terminal tail must become structured in the dimer and it is where the inhibitors have been shown to bind to the monomer, I postulate that the slowest LE4PD mode describes potential binding fluctuations of the monomer. I am not aware of any biological significance of the second mode, but it tends to disagree with that extracted from the MD simulations, where a combination of C- or N-terminal and flap dynamics are described.

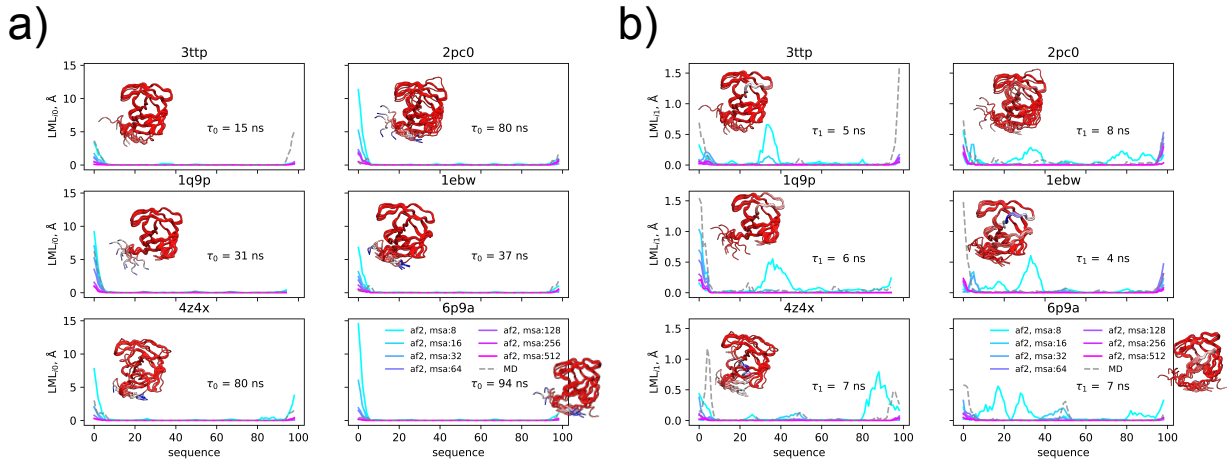


Figure I.4: Local mode lengthscale (LML) for the first a) and second b) LE4PD modes for each of the six sequence variants studied in the main text as a function of the MSA depth in AF2. Embedded ensembles show the LML projected onto the alpha-carbons for the MSA=8 ensemble for each sequence. The colorscale is red-white-blue, with red indicating no dynamics and blue indicating the most dynamic regions. The scaling is the same for each sequence for each mode, but the normalization of the scaling is set uniquely for each mode. Also given in each plot is the mode's timescale for the MSA=8 ensemble. The LML for each sequence is shown as the dashed, grey curve as a reference

I.5 Conformational Diversity Plots for all HIV-1 Protease Sequences

I display in Figure I.5 the conformational diversity plots analogous to Figure 3 in the main text for all six sequences of the HIV-1 protease protein I examine in this study. Scatter point coloring and free-energy contour lines are identical to those used in Figure 3 in the main text.

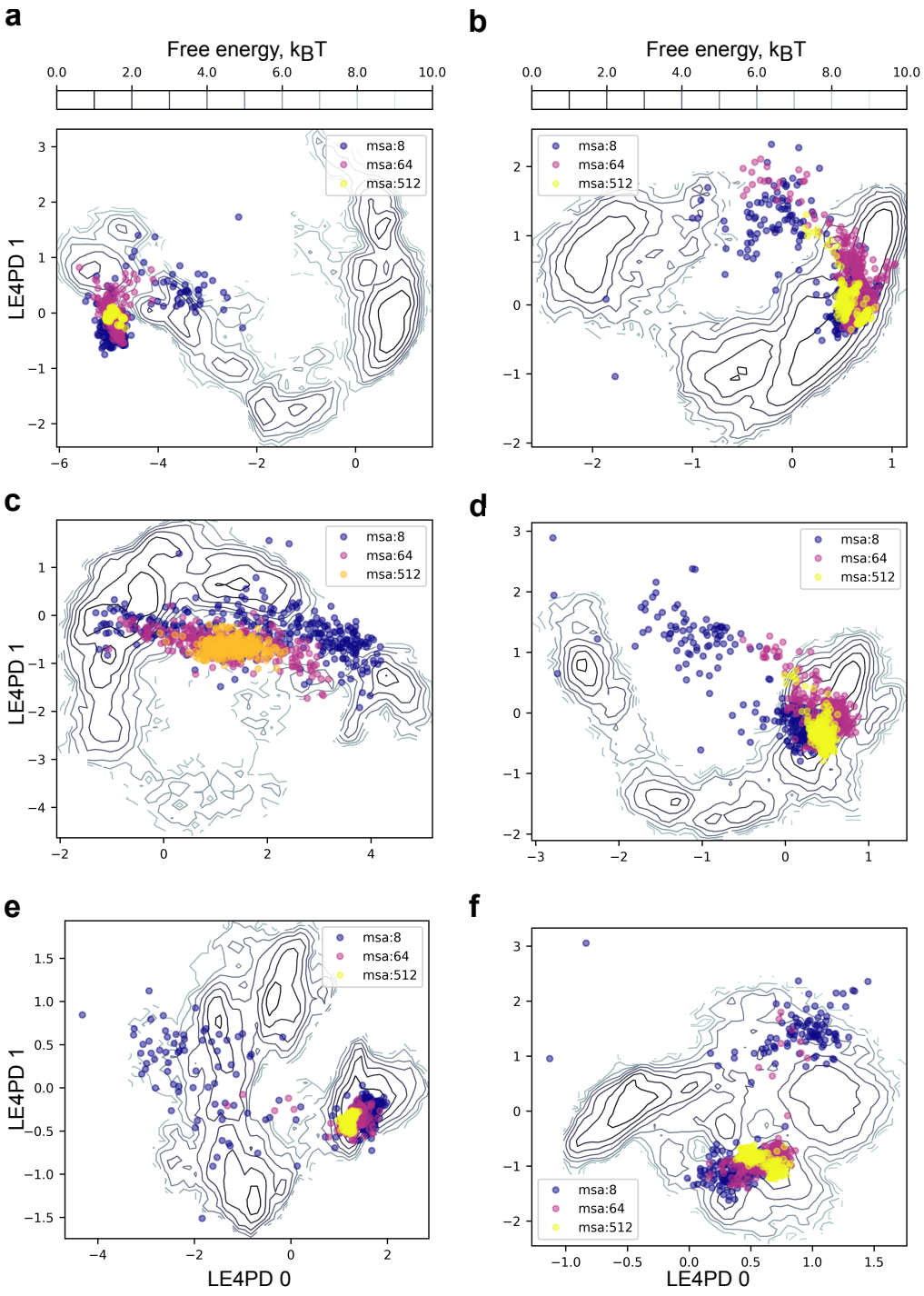


Figure I.5: Projection of AlphaFold2 structures at the indicated MSA depth onto the space spanned by the two slowest LE4PD modes, which are calculated individually for each protein from the respective one microsecond trajectory starting from the highest confidence AlphaFold two structure. Results for sequences a) 3TTP, b) 2PC0, c) 1Q9P, d) 1EBW, e) 4Z4X, and f) 6P9A.

Likewise, Figure I.6 shows the conformational diversity of the ensembles output from AlphaFold2 at MSA depths of 8 and 64 as well as AlphaFold-2 cluster, BioEmu, and DiG. This figure is analogous to Figure 6 in the main text.

For comparison to the experimental structures, I project the starting X-ray or NMR starting structures for the simulation onto the LE4PD free-energy surfaces for each monomer in Figure I.7. The results are sequence dependent: for 1Q9P, 2PC0, and 6P9A, the crystal structure is in a free-energy well. For 3TTP, 1EBW, and 4Z4X, the crystal structure lies outside a free-energy well. This result is likely because the slowest LE4PD modes describe (potential) binding fluctuations in the N-terminal tail, but the crystal structures come from dimers, where the C- and N-terminal tails are pinned in beta-sheets. However, I see that the NMR ensemble, including the starting structure for my simulations (the magenta star in the projection) overlap well with the free-energy surface from the MD simulation. Because the free monomers are unlikely to sample the exact folded conformations present in the dimer, the crystallographic structures can show poor overlap with the wells in the free-energy surface from the LE4PD method.

I.6 Dynamical Time Warping and Kolmogorov-Smirnov Distance Matrices

Figures I.8 and I.9 show the distance matrices between the brute-force calculated time correlation functions for the given HIV-1 protease sequence at the given simulation length (10, 100, or 1000 ns) and the predictions from the LE4PD method parameterized with the AlphaFold2 structural ensemble generated at the displayed MSA depth. As mentioned in the main text, for each timescale and for each sequence, that residue's time correlation function is best described by the MSA depth where the distance is minimal. Practically, this procedure amounts to taking the minimum argument across the columns of each matrix.

Figure I.8 shows the result when I use the dynamical time warping distance as the distance metric while I.9 shows the result when I convert the respective time correlation functions to its cumulative distribution functions, then calculate the distance between the simulated and analytical time correlation functions using the Kolmogorov-Smirnov test. Visual inspection of the two sets of matrices reveals that using the two different metrics gives qualitatively similar results. Thus, although I use the dynamical time warping distance to map the timescales between the simulation and the machine-learned ensembles, the Kolmogorov-Smirnov distance can be used to similar effect.

The analogous plot for the other generative machine learning methods investigated in the manuscript are given in Figure I.10 and I.11. For all distance calculations, the time correlation functions are truncated at the lagtime where the brute-force calculated time correlation function from the corresponding molecular dynamics trajectory shows a positive gradient, which indicates that noise is beginning to dominate the calculation of the correlation function since these are time correlation functions of a system decaying toward equilibrium, and they must be monotonically decreasing in the absence of statistical noise.

I.7 Timescale Correlation Plots for the Six HIV-1 Protease Sequences

Figure I.12 shows the linear fits between the MD integrated timescales from the residue-residue time correlation functions and the integrated timescales from the analytical residue-

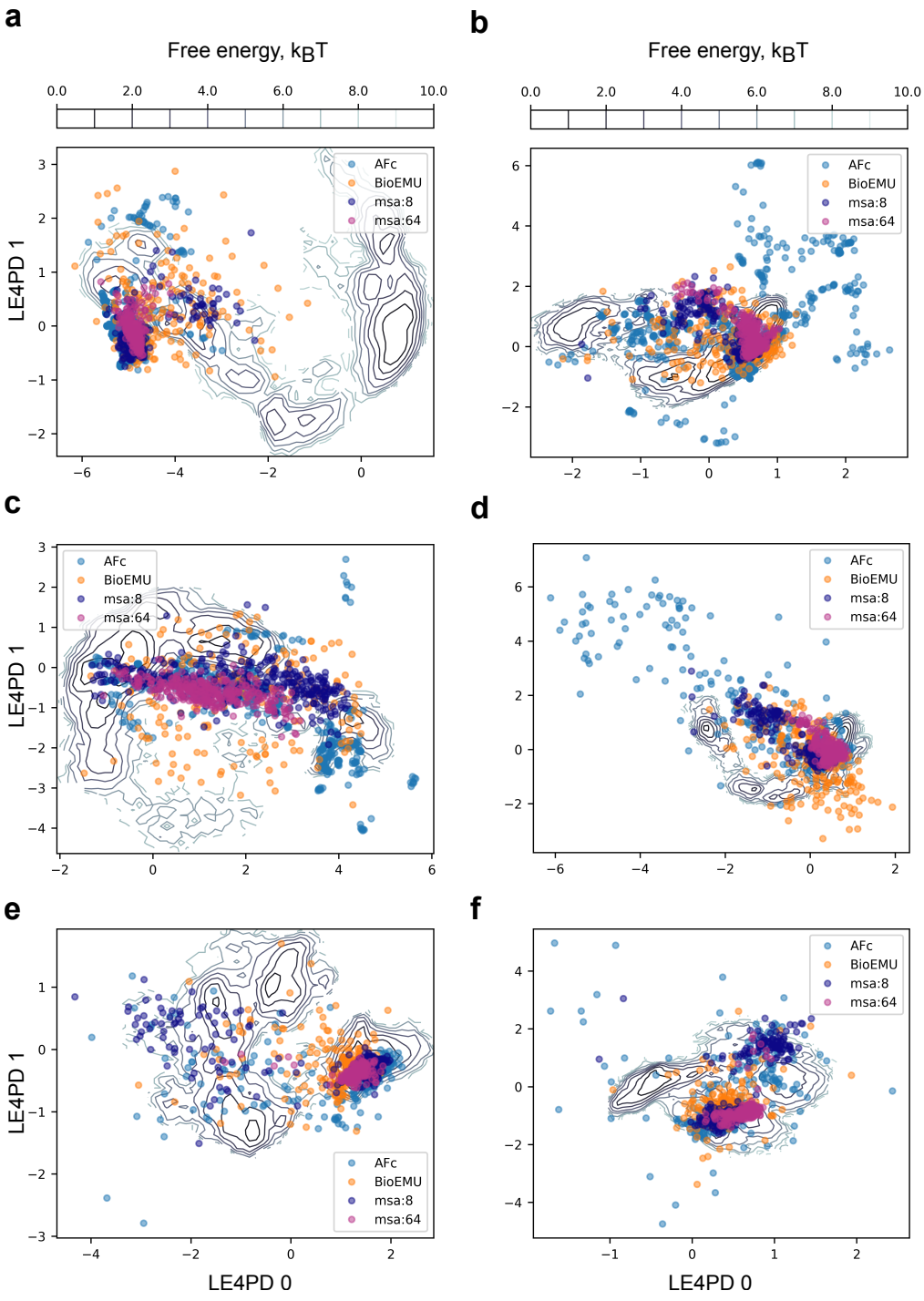


Figure I.6: Projection of AlphaFold2 structural ensembles generated at MSA depths of 8 and 64 as well as those generated from AlphaFold2-cluster, BioEmu, and DiG, onto the space spanned by the two slowest LE4PD modes, which are calculated individually for each protein from the respective one microsecond trajectory starting from the highest confidence AlphaFold2 structure. Results for sequences a) 3TTP, b) 2PC0, c) 1Q9P, d) 1EBW, e) 4Z4X, and f) 6P9A.

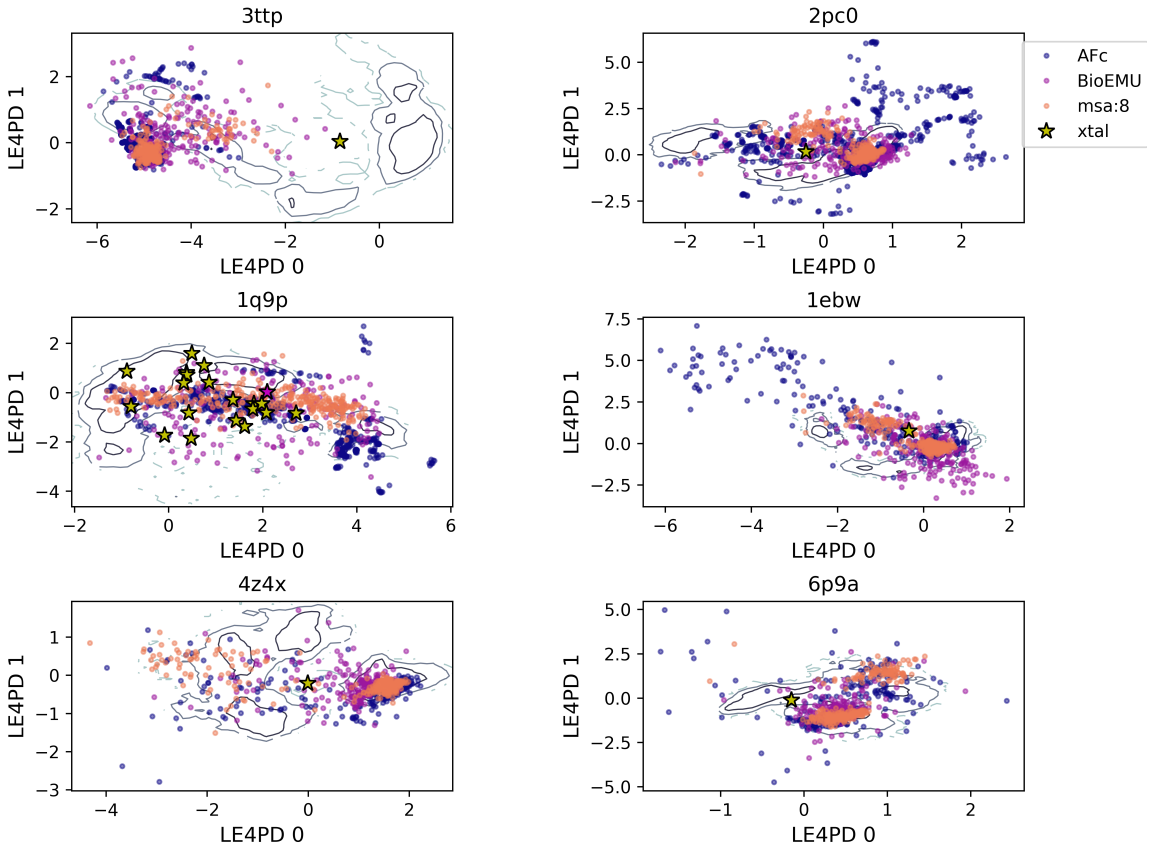


Figure I.7: Projections of samples from the indicated ML methods and MD starting structures onto the free-energy surfaces spanned by the two slowest LE4PD modes. Besides the change in the number of contour lines present in the free-energy surfaces, the X-ray structures used as the initial conformations in MD simulations are projected onto the free-energy landscapes as yellow stars, except for 1Q9P, where the starting structure harvested from the NMR ensemble is indicated with a magenta star while the remaining nineteen structures in the ensemble are indicated as yellow stars.

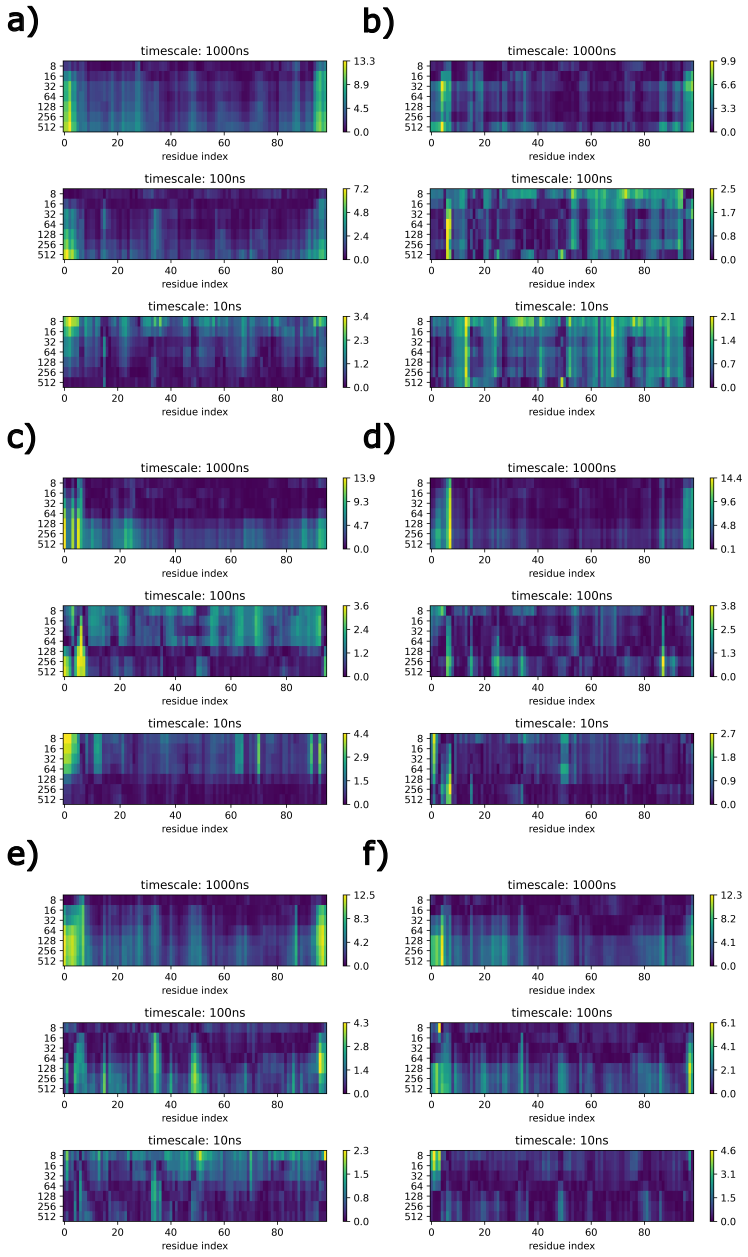


Figure I.8: Distance matrices between the brute-force calculated time correlation functions from the MD simulation and the analytical LE4PD time correlation functions parameterized using the AlphaFold2 structural ensembles generated at the indicated MSA depth for three simulation lengthscales indicated (10, 100, and 1000 ns) for each of the six HIV-1 protease sequences studied here: a) 3TTP, b) 2PC0, c) 1Q9P, d) 1EBW, e) 4Z4X, and f) 6P9A. The distance metric used is the dynamical time warping distance between the two time correlation functions.

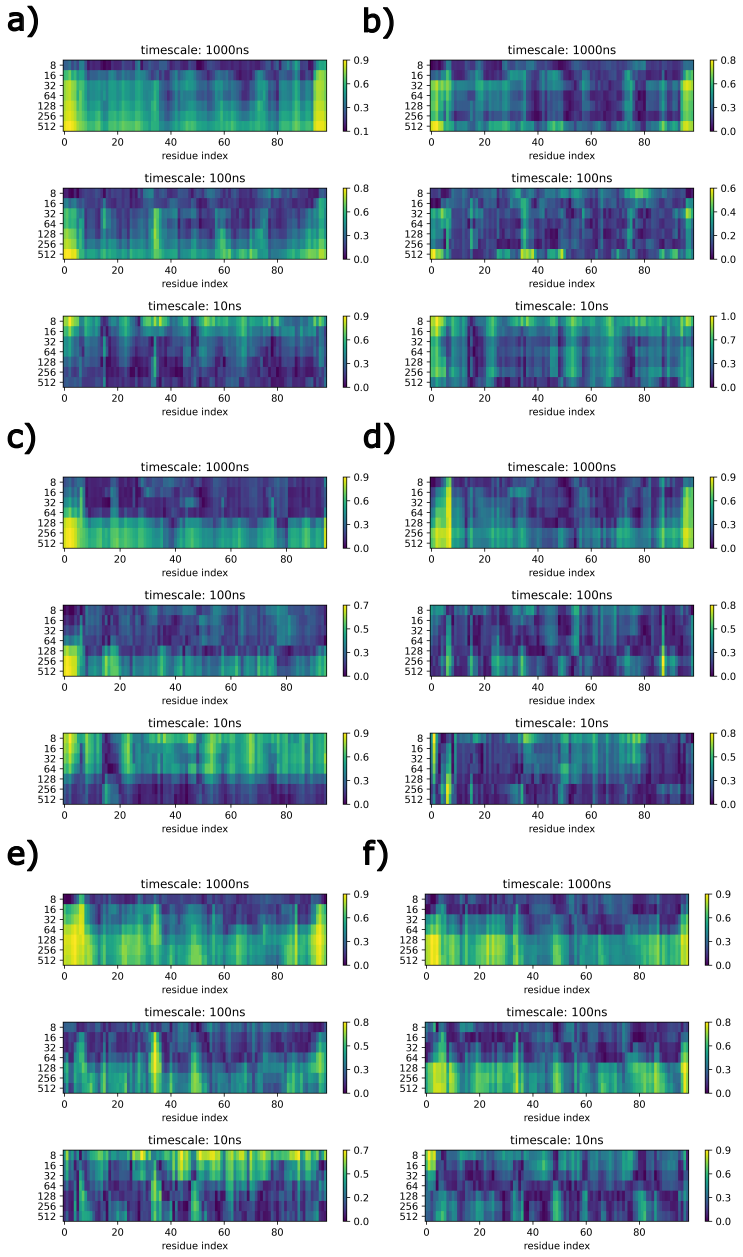


Figure I.9: Distance matrices between the brute-force calculated time correlation functions from the MD simulation and the analytical LE4PD time correlation functions parameterized using the AlphaFold2 structural ensembles generated at the indicated MSA depth for three simulation lengthscales indicated (10, 100, and 1000 ns) for each of the six HIV-1 protease sequences studied here: a) 3TTP, b) 2PC0, c) 1Q9P, d) 1EBW, e) 4Z4X, and f) 6P9A. To calculate this set of distance matrices, each time correlation function is converted to its respective cumulative distribution function. Then, the Kolmogorov-Smirnov test between the two cumulative distribution functions is calculated, and the mapping is performed to the simulated time correlation functions with the lowest value output from the Kolmogorov-Smirnov test.

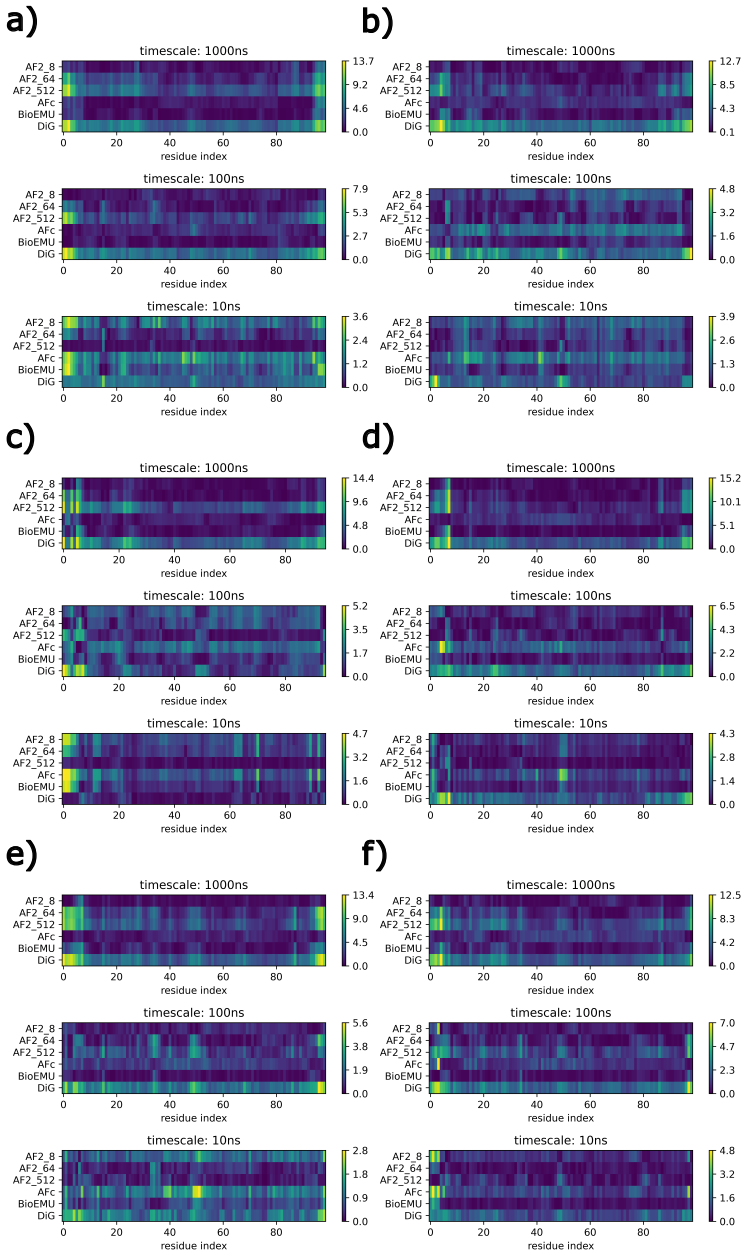


Figure I.10: Distance matrices between the brute-force calculated time correlation functions from the MD simulation and the analytical LE4PD time correlation functions parameterized using either AlphaFold2 or other machine-learned (AlphaFold2-cluster, BioEmu, or DiG) structural ensembles for three simulation lengthscales indicated (10, 100, and 1000 ns) for each of the six HIV-1 protease sequences studied here: a) 3TTP, b) 2PC0, c) 1Q9P, d) 1EBW, e) 4Z4X, and f) 6P9A. The distance metric used is the dynamical time warping distance between the two time correlation functions.

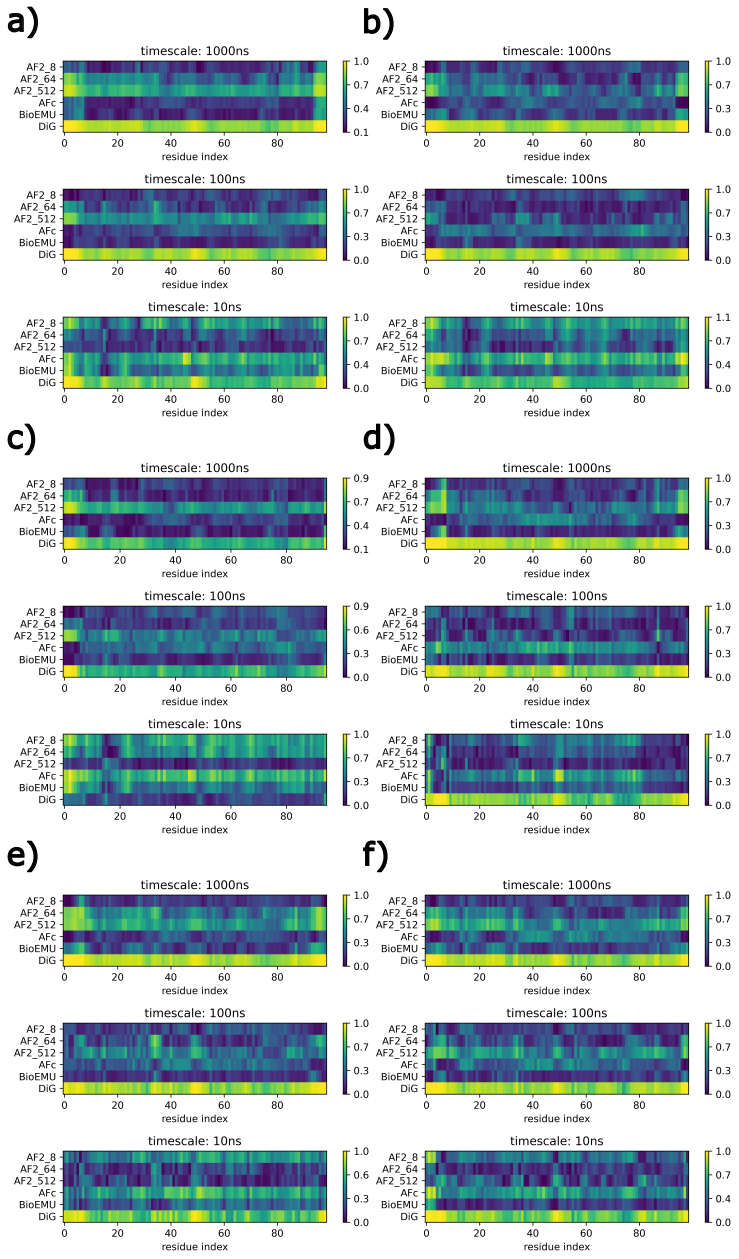


Figure I.11: Distance matrices between the brute-force calculated time correlation functions from the MD simulation and the analytical LE4PD time correlation functions parameterized using either AlphaFold2 or other machine-learned (AlphaFold2-cluster, BioEmu, or DiG) structural ensembles for three simulation lengthscales indicated (10, 100, and 1000 ns) for each of the six HIV-1 protease sequences studied here: a) 3TTP, b) 2PC0, c) 1Q9P, d) 1EBW, e) 4Z4X, and f) 6P9A. To calculate this set of distance matrices, each time correlation function is converted to its respective cumulative distribution function. Then, the Kolmogorov-Smirnov test between the two cumulative distribution functions is calculated, and the mapping is performed to the simulated time correlation functions with the lowest value output from the Kolmogorov-Smirnov test.

residue correlation functions predicted from LE4PD theory parameterized using the AlphaFold2 ensemble generated at an MSA depth of 8 for each of the six HIV-1 protease sequences studied here. The slope and intercept of each fit is given in the corresponding subplot in Figure 4c in the main text.

Figure I.13 shows the linear fits between the MD integrated timescales from the residue-residue time correlation functions and the integrated timescales from the analytical residue-residue correlation functions predicted from LE4PD theory parameterized using the either the AlphaFold2-, BioEmu-, or DiG-generated structural ensembles for each of the six HIV-1 protease sequences studied here. The slope and intercept of each fit is given in the corresponding subplot in Figure 7c in the main text.

I.8 Out-of-set-structure Analysis for All Six HIV-1 Protease Sequences

As discussed briefly in the main text, when outlying structures are removed from the AlphaFold2 ensemble, I find that the ultimate effect on the predicted dynamics and kinetics is sequence dependent. I find that, depending on the variant's particular primary sequence, there are three effects of outlying structure removal.

For the sequences encoded in PDB IDs 1EBW and 2PC0, there are a large number of outlying structures, the removal of which decreases the variance of the ML-generated ensemble. For these two sequences, removing the outlying structures from the calculation of the potential of mean force matrix in the Langevin equation results in a near universal decrease of the RMSF and integrated timescales, as shown in Figure I.14

For the sequences encoded by PDB IDs 1Q9P and 3TTP, the opposite effect is observed, where removing outlying structures generally increases both the relaxation timescales and the RMSF of the ML-generated ensemble, as shown in Figure I.15. This effect is due to the mode of the distribution of the ML-generated ensembles not overlapping with a mode of the MD distribution. As such, there are structures (especially for 1Q9P) lying near the average structure from the ML ensemble that are removed from the potential of mean force calculation, and thus increasing the variance along the slowest LE4PD mode and hence the relaxation timescales.

Finally, there may be few out-of-set structures, which is the case for the sequences encoded by PDB IDs 4Z4X and 6P9A. As shown in Figure I.16, removing the outlying structures has only a small effect on the RMSF and timescales. This result is expected since the ensemble is only slightly perturbed by the removal of a few structures.

I.9 Out-of-set-structure Analysis for BioEmu and AlphaFold2-Cluster

In Figures I.17 and I.18 I show the effect of removing out-of-set structures on the calculated RMSF and integrated timescales τ_{avg}, respectively, for the ensemble generated using BioEmu. There is little change in either the RMSF or the integrated timescales.

In Figures I.19 and I.20 I show the effect of removing out-of-set structures on the calculated RMSF and integrated timescales τ_{avg}, respectively, for the ensemble generated using AF-cluster. There are much larger changes in both RMSF and integrated timescales when out-of-set structures are removed compared to the BioEmu ensemble. However, even with the removal of the out-of-set structures, the predicted RMSF is still much larger than that calculated from the benchmark

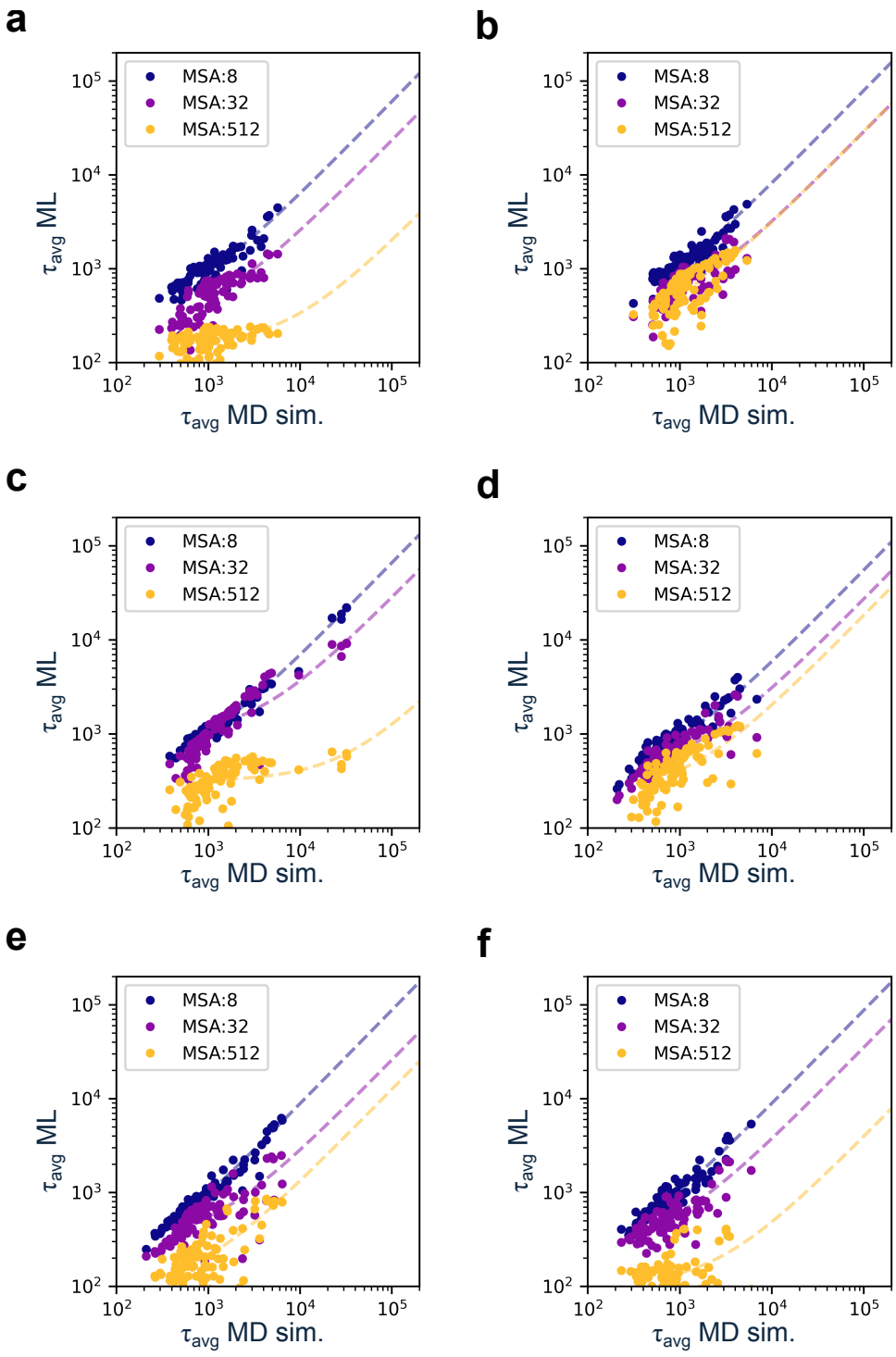


Figure I.12: Correlation between the integrated timescales from the one microsecond MD simulation starting from the most confidence structure predicted from AlphaFold2 for the given HIV-1 protease sequence and the integrated timescale from the analytical residue-residue time correlation functions predicted from the LE4PD theory parameterized using the AlphaFold2 structural ensemble generated at an MSA depth of 8. Results for sequences a) 3TTP, b) 2PC0, c) 1Q9P, d) 1EBW, e) 4Z4X, and f) 6P9A.

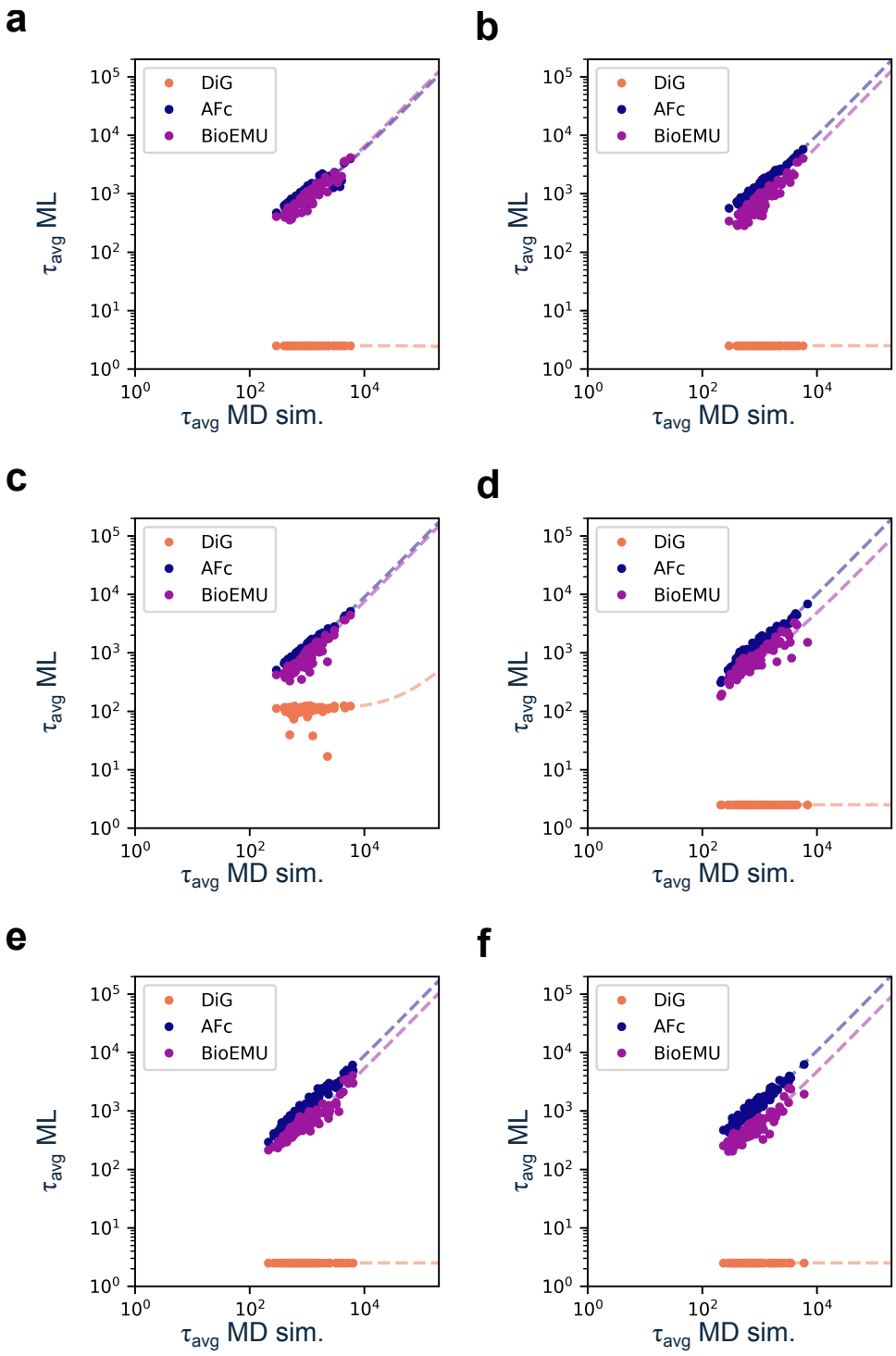


Figure I.13: Correlation between the integrated timescales from the one microsecond MD simulation starting from the most confidence structure predicted from AlphaFold2 for the given HIV-1 protease sequence and the integrated timescale from the analytical residue-residue time correlation functions predicted from the LE4PD theory parameterized using either AlphaFold2-cluster-, BioEmu-, or the DiG-generated ensembles. Results for sequences a) 3TTP, b) 2PC0, c) 1Q9P, d) 1EBW, e) 4Z4X, and f) 6P9A.

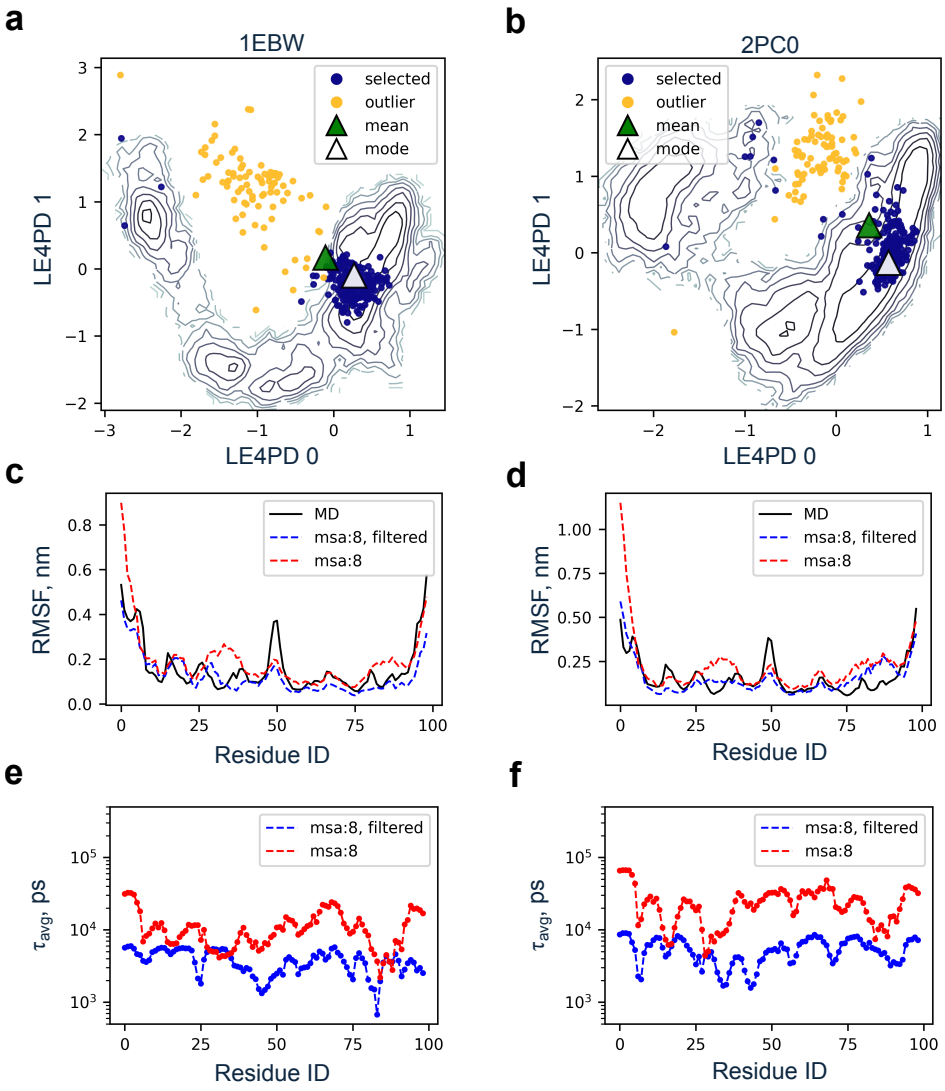


Figure I.14: Projection of structures from the AlphaFold2 ensemble with an MSA depth of 8 projected onto the free-energy surfaces spanned by the first two LE4PD modes for a) 1EBW and b) 2PC0. The mean structure of the distribution is labeled with a green triangle; the mode is labeled with a white triangle; outlying structures are colored yellow; finally, in-distribution structures (according to the microsecond MD) are colored blue. Comparison of the RMSF before (red) and after the removal of outlying structures (blue) are shown for c) 1EBW and d) 2PC0. The RMSF calculated from a microsecond MD of these two sequences is given by the solid, black curve. Finally, the integrated timescales with all structures kept in the ensemble (red) and with outlying structures removed (blue) are shown in e) for 1EBW and f) for 2PC0.

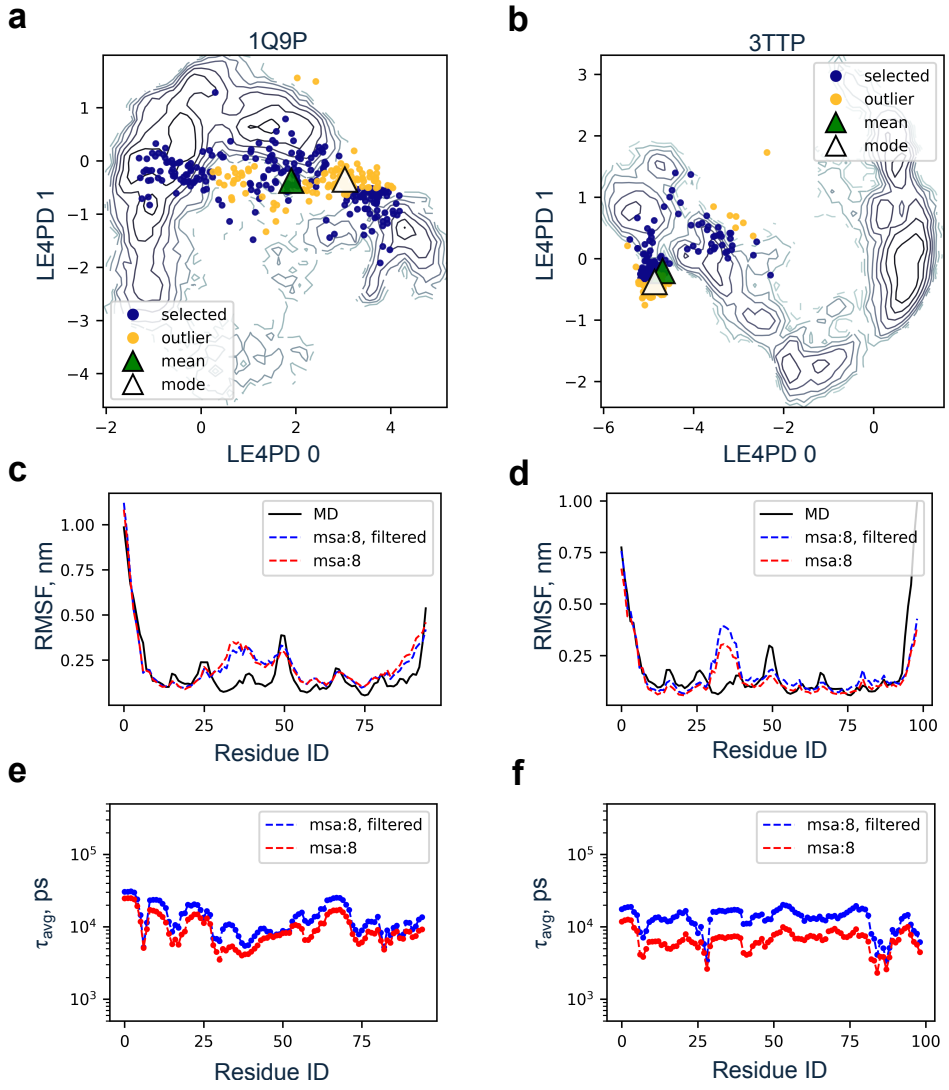


Figure I.15: Projection of structures from the AlphaFold2 ensemble with an MSA depth of 8 projected onto the free-energy surfaces spanned by the first two LE4PD modes for a) 1Q9P and b) 3TTP. The mean structure of the distribution is labeled with a green triangle; the mode is labeled with a white triangle; outlying structures are colored yellow; finally, in-distribution structures (according to the microsecond MD) are colored blue. Comparison of the RMSF before (red) and after the removal of outlying structures (blue) are shown for c) 1Q9P and d) 3TTP. The RMSF calculated from a microsecond MD of these two sequences is given by the solid, black curve. Finally, the integrated timescales with all structures kept in the ensemble (red) and with outlying structures removed (blue) are shown in e) for 1Q9P and f) for 3TTP.

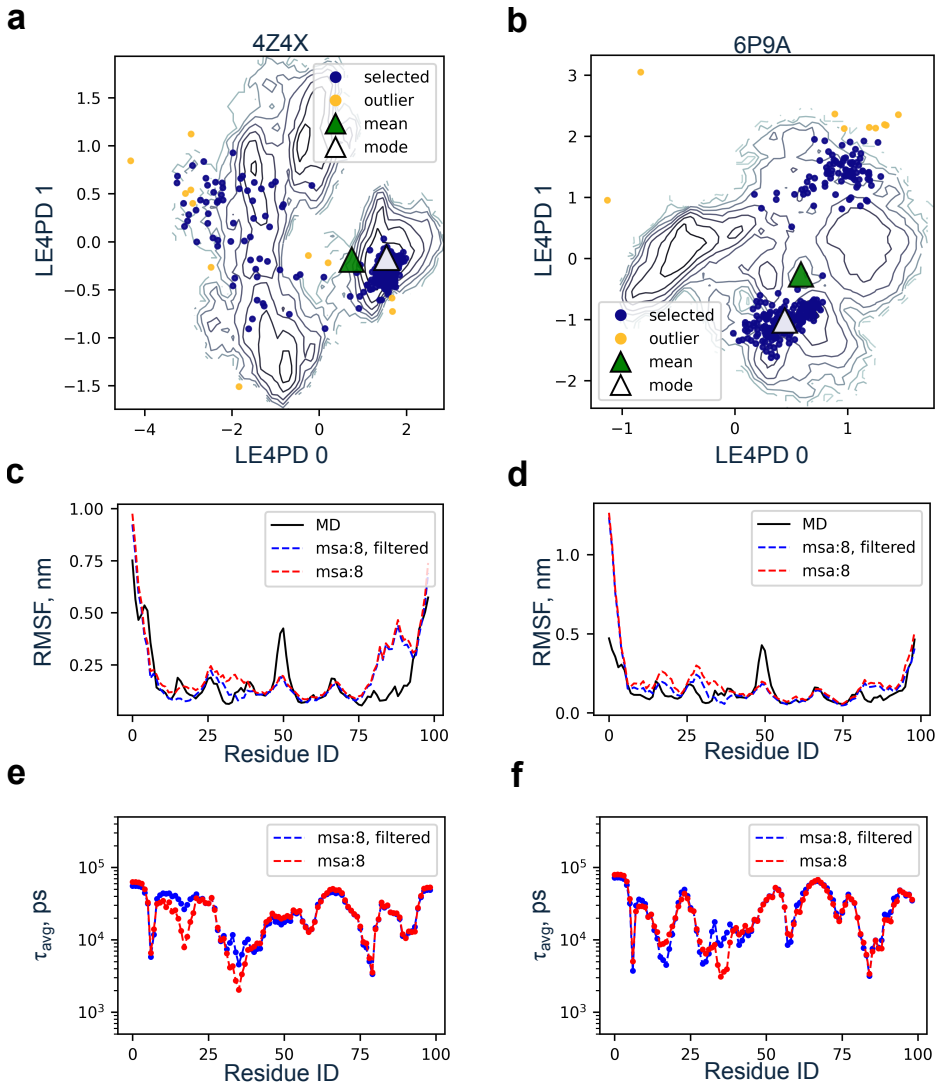


Figure I.16: Projection of structures from the AlphaFold2 ensemble with an MSA depth of 8 projected onto the free-energy surfaces spanned by the first two LE4PD modes for a) 4Z4X and b) 6P9A. The mean structure of the distribution is labeled with a green triangle; the mode is labeled with a white triangle; outlying structures are colored yellow; finally, in-distribution structures (according to the microsecond MD) are colored blue. Comparison of the RMSF before (red) and after the removal of outlying structures (blue) are shown for c) 4Z4X and d) 6P9A. The RMSF calculated from a microsecond MD of these two sequences is given by the solid, black curve. Finally, the integrated timescales with all structures kept in the ensemble (red) and with outlying structures removed (blue) are shown in e) for 4Z4X and f) for 6P9A.

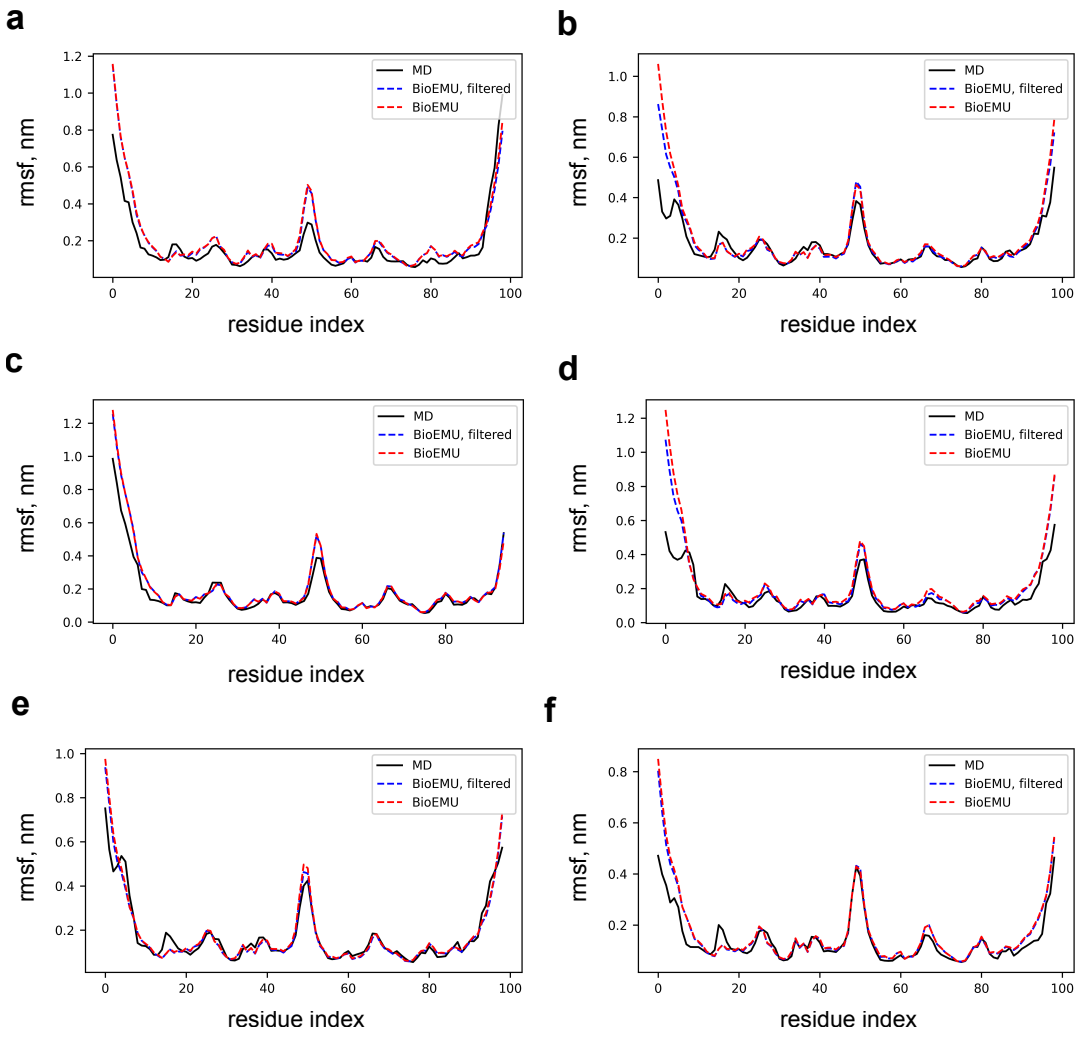


Figure I.17: RMSF predicted from the BioEmu ensemble with (red) and without (blue) out-of-set structures included in the calculation for each of six HIV-1 protease variants: a) 3TTP; b) 2PC0; c) 1Q9P; d) 1EBW; e) 4Z4X; f) 6P9A. The RMSF is benchmarked by the RMSF calculated directly from a one microsecond equilibrium MD simulation (black).

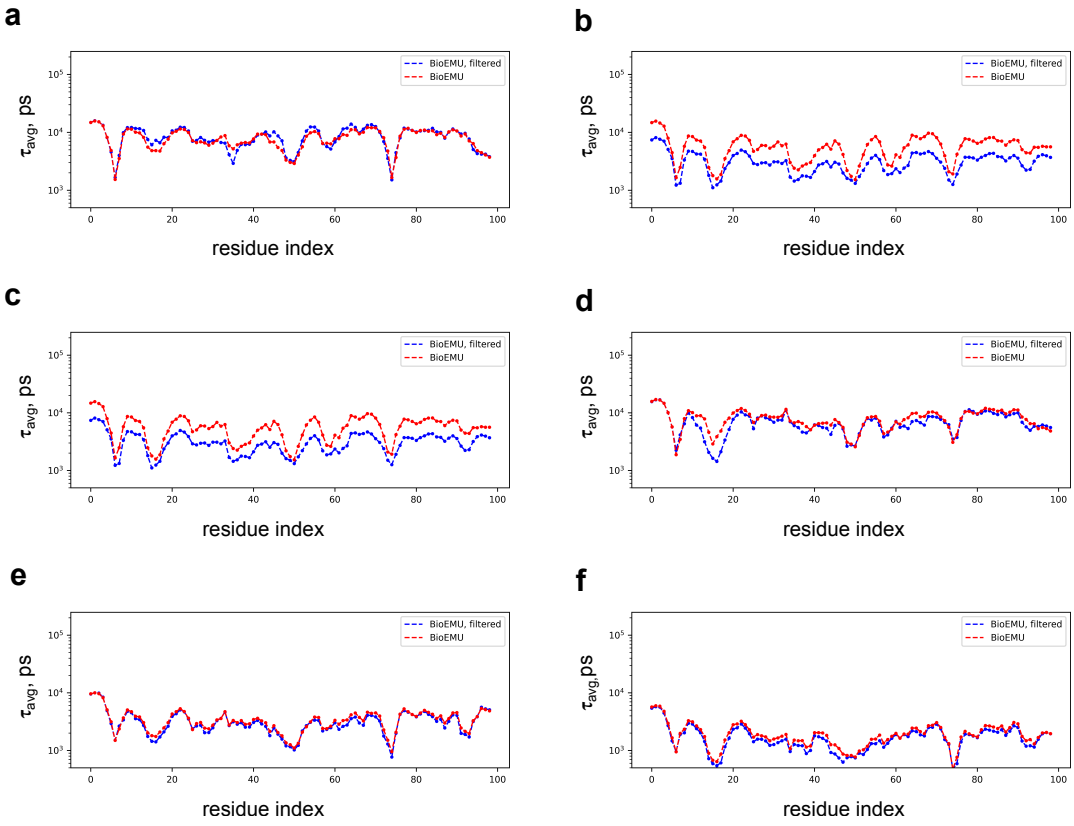


Figure I.18: Integrated timescales τ_{avg} predicted from the BioEmu ensemble with (red) and without (blue) out-of-set structures included in the calculation for each of six HIV-1 protease variants: a) 3TTP; b) 2PC0; c) 1Q9P; d) 1EBW; e) 4Z4X; f) 6P9A.

microsecond MD simulation.

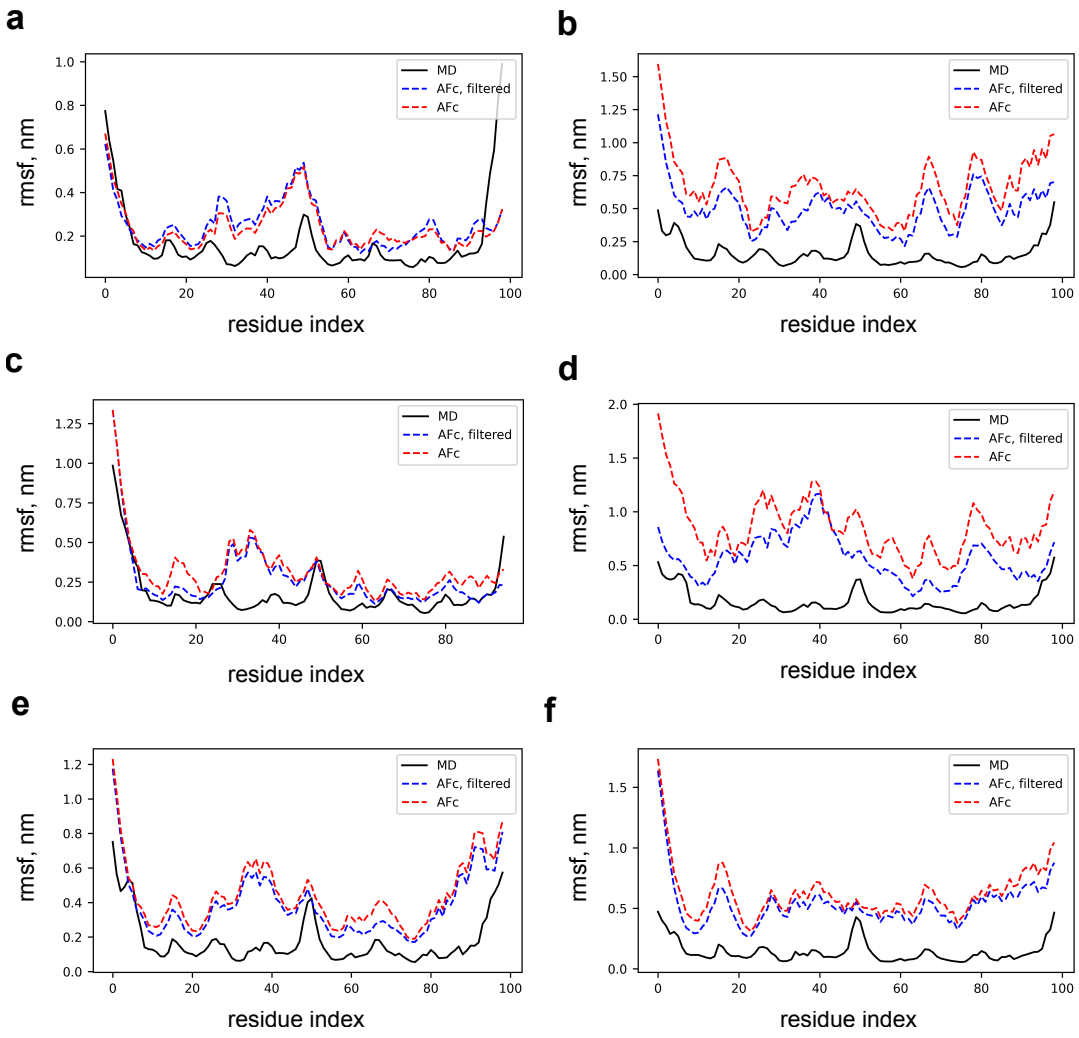


Figure I.19: RMSF predicted from the BioEmu ensemble with (red) and without (blue) out-of-set structures included in the calculation for each of six HIV-1 protease variants: a) 3TTP; b) 2PC0; c) 1Q9P; d) 1EBW; e) 4Z4X; f) 6P9A. The RMSF is benchmarked by the RMSF calculated directly from a one microsecond equilibrium MD simulation (black).

I.10 Analysis for a Wild-type Dimer and a Mutant Dimer

Since the functional form of the HIV-1 protease is dimeric, I present here a brief analysis of a wild-type, dimeric HIV-1 protease (PDB ID: 1HHP[300]) and a mutant (PDB ID: 3UF3[301]). I perform one microsecond MD simulations using the same parameterization as for the monomer variants analyzed in the main text. After performing the simulation, I analyze the distance between the tips of the flaps (measured by the distance between the I50 and I149 residues), the distance from this isoleucine residue at the tip of the flap to the active site aspartate on the same

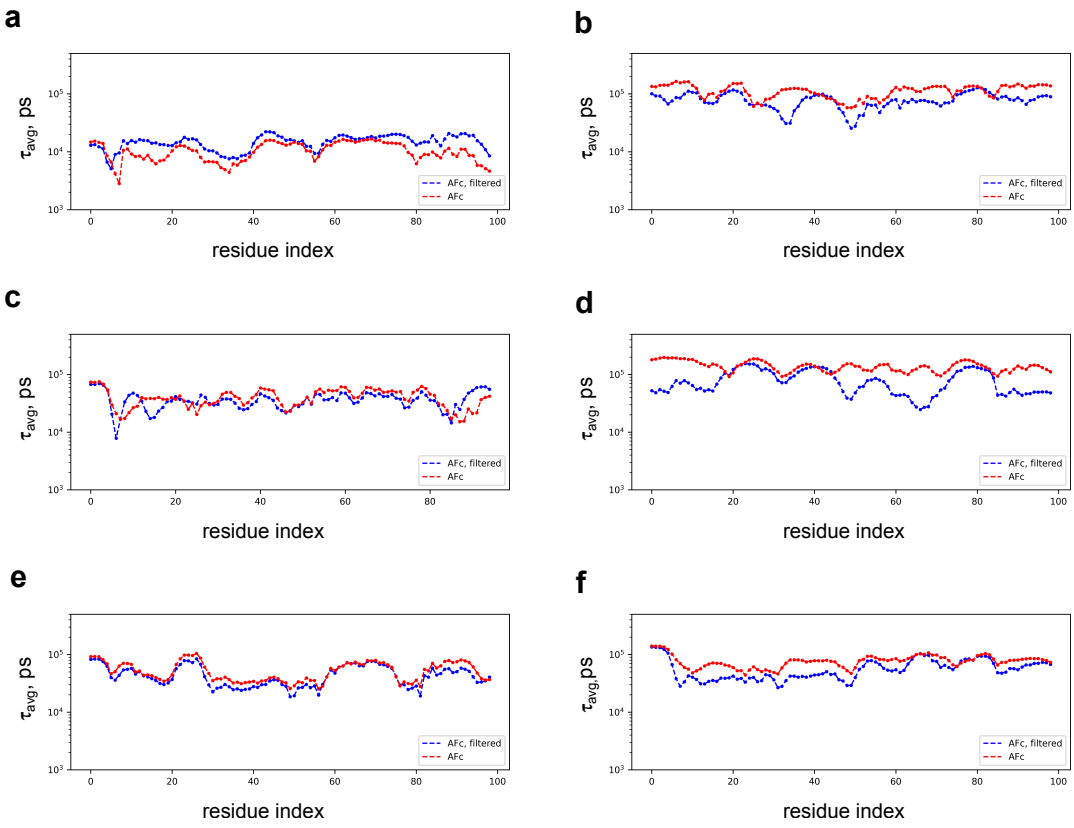


Figure I.20: Integrated timescales τ_{avg} predicted from the AlphaFold2-cluster ensemble with (red) and without (blue) out-of-set structures included in the calculation for each of six HIV-1 protease variants: a) 3TTP; b) 2PC0; c) 1Q9P; d) 1EBW; e) 4Z4X; f) 6P9A.

monomer, and the curl angle of the flap tip, measuring the angle formed by residues 48-49-50 or 147-148-149.[296] The histograms for these three measurements calculated over both the wild-type and mutant simulations are shown in Figure I.21. In I.21a) I show the histogram of the cosine of the flap tip curling angle, which I average over both residues because the difference between the two wild-type and mutant distributions for either monomer are not statistically significant, according to the KS test (data not shown). In either case, the averaged probability distribution between the two variants is not statistically significant either. Figure I.21b) shows the probability distribution of the flap-flap distance between the two monomers in the dimer. In this case, the mutant has, on average, a significantly larger distance between the two flap tips. Figures I.21c) and I.21d) show the flap tip to catalytic aspartate distance for the first and second monomer in the dimer, respectively. Due to the asymmetry of the dynamics observed in the MD simulations of both dimers, the probability distribution of the mutant is significantly different from the wild-type only for the first monomer, as verified by comparing the associated cumulative distribution functions (CDFs) using the KS test (CDFs not shown).

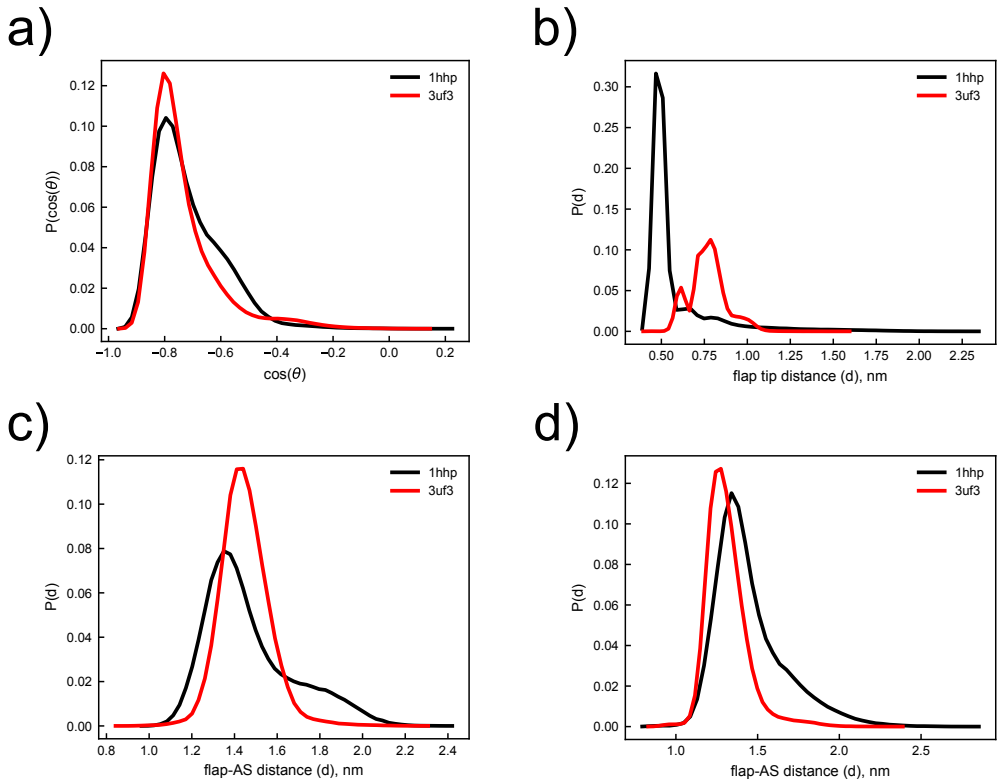


Figure I.21: Histograms of a) the curling angle of the three residues of the flap tip, averaged over the two monomers; b) the flap-flap distance between monomers in the HIV-1 protease dimer; c) the I50-D25 distance for the first monomer in the dimer; and d) the I149-D124 distance for the second monomer in the dimer.

I also perform an LE4PD analysis on these trajectories, and show the root-mean-squared fluctuations projected onto the six slowest LE4PD modes, otherwise known as the local mode lengthscale (LML). These plots are shown in Figure I.22 for the wild-type and mutant variants. For modes where the fluctuations are large enough, I superimpose structures of the dimer with

the LML projected onto the alpha-carbons onto the appropriate subplot. Compared to the LE4PD timescales of the monomers, the timescales of the dimer are more than an order of magnitude faster, at least in part due to the amplitude of the fluctuations of the dimers' slow modes being at least an order of magnitude lower than those of the monomers. This result aligns with the intuition that energetically favorable binding suppresses such entropically favorable conformational fluctuations from the bound state. The asymmetry of the LML plots also demonstrates previously seen asymmetry in the constituent monomers of the HIV-1 protease dimer.[288, 290, 296]

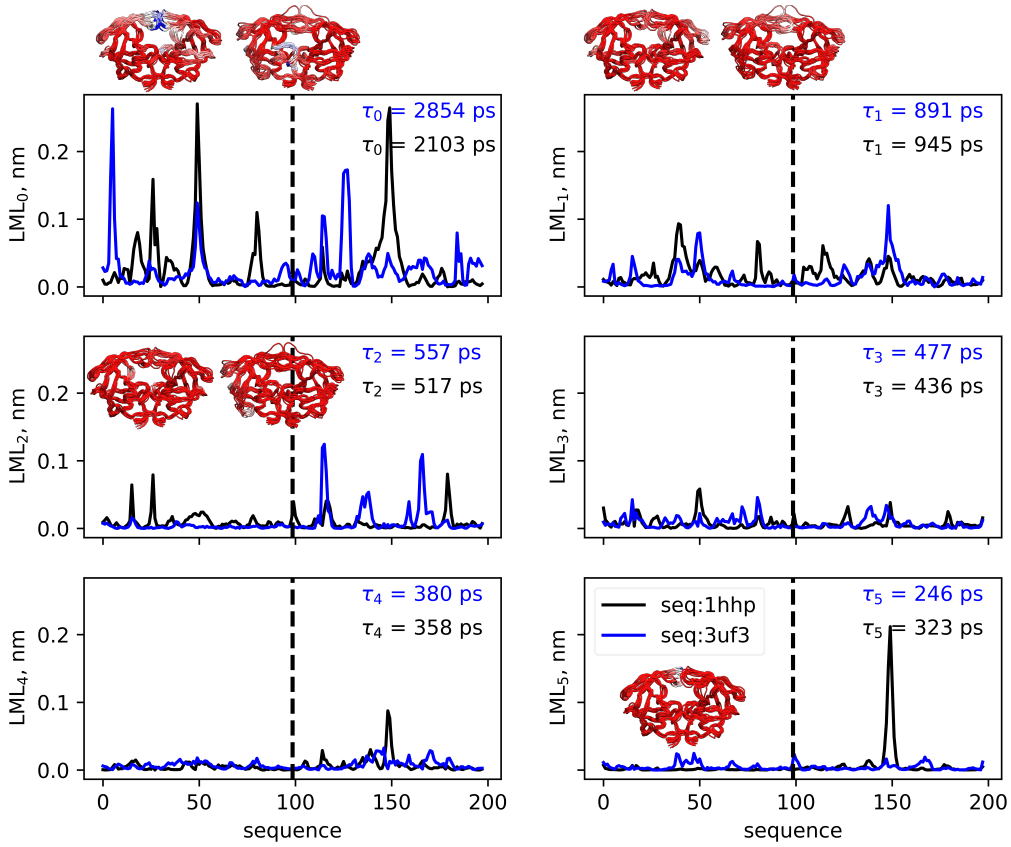


Figure I.22: Local mode lengthscale (LML) for the six slowest LE4PD modes calculated from MD simulations of the wild-type 1hhp (black) and mutant 3uf3 (blue) HIV-1 protease dimer. The LMLs projected onto the structural ensembles from the microsecond MD simulations are colored using a red-white-blue colorscale, where red indicates no dynamics and blue indicates large scale conformational dynamics. Note that the colorscale is scaled to the LML of the dimer simulations, and it does not correspond to the same scale used in the analogous plot for the monomer simulations. Here, the LMLs are around an order of magnitude smaller compared to the monomer results.

In Figure I.23, I show the projections of the flap-flap distance, flap tip-active site distance (I50-D25 distance, averaged over both residues), and the flap tip curl angle (again averaged over both residues) onto the two slowest LE4PD modes. I see that, for the mutant, changes in these collective variables are not encoded by the two slowest LE4PD modes. For the wildtype, the slowest LE4PD mode describes the shift from small to large flap-flap and flap-active site distances, but no coherent change in the flap tip curling angle.

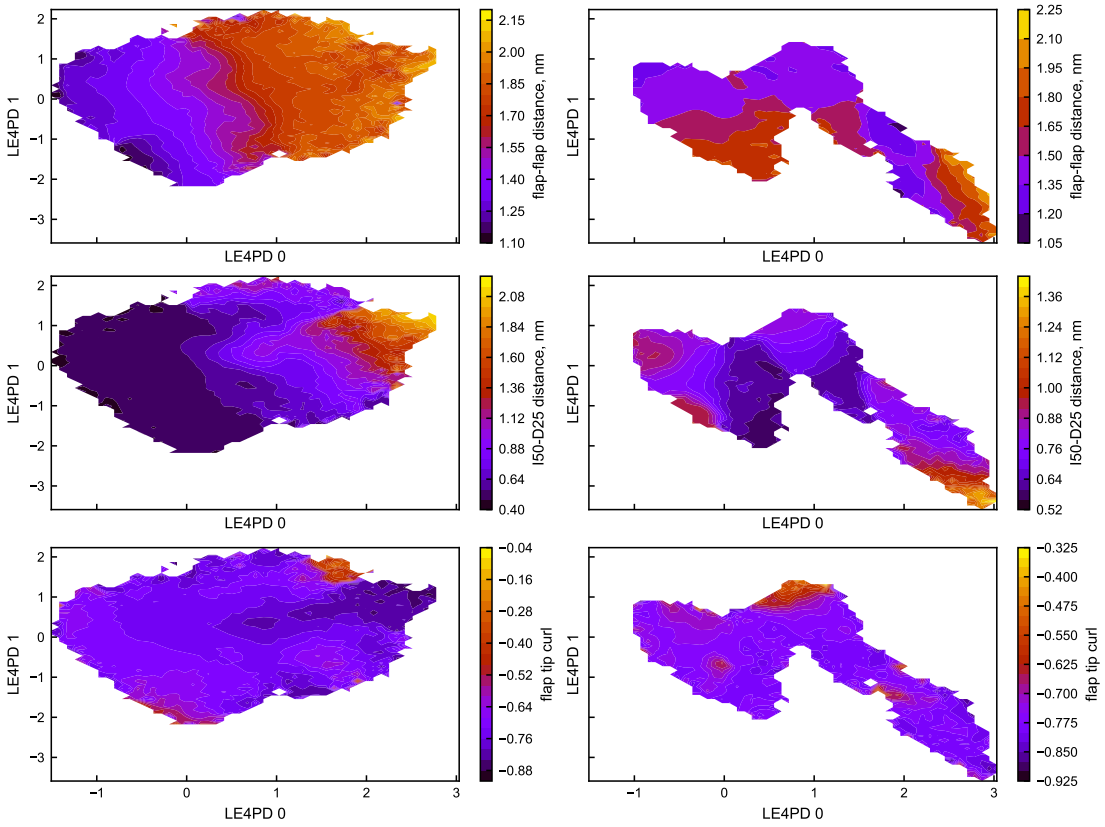


Figure I.23: Projection of the flap-flap distance (first row), flap tip-catalytic site distance (middle row), and the cosine of the flap tip curling angle (bottom row) onto the free-energy surface spanned by the first two LE4PD modes for the wildtype (1HHP; left column) and mutant (3UF3; right column) sequences.

I.11 Out-of-set-structure Analysis of Dimer Ensembles Generated from Co-folding ML Methods

I apply my protocol for the dimer ensembles generated by AlphaFold-multimer (AFmultimer).[184] I use the ColabFold implementation of AFmultimer to generate the conformational ensembles. I found that even with higher MSA depth, the co-folding models were generating structures with C_{2h}, instead of the functionally relevant C_2 symmetry. In Figure I.24 I show one structure for the WT sequence and MUT sequence where AFmultimer predicts a dimer with biologically non-functional symmetry, C_{2h}.

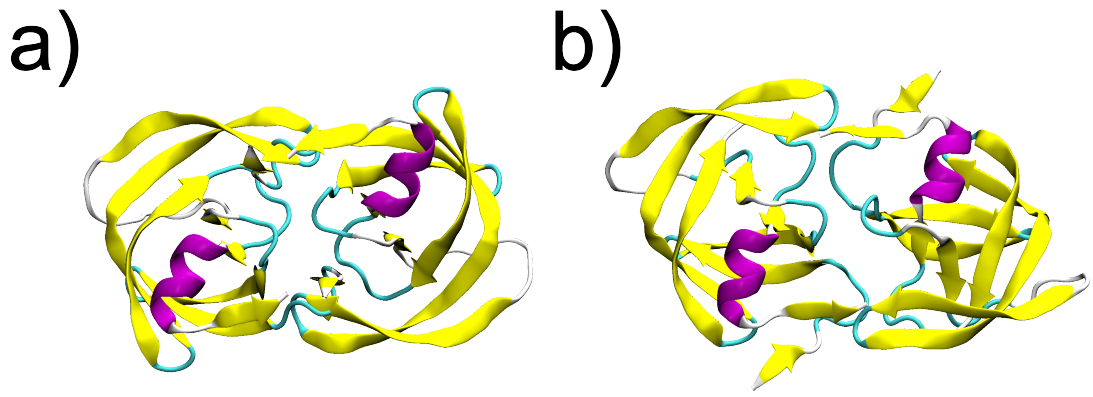


Figure I.24: Example C_{2h} symmetry structures for a) WT 1HHP sequence and b) MUT 3UF3 sequence predicted by AF2-multimer.

In Figure I.25a and b clearly show that co-folding models generally predict such structures with non-functional symmetry. Furthermore, the mutant has many more hallucinated structures even when the individual chains preserved the fold. This is seen in Figure I.25c. The presence of such hallucinated structures exaggerates the calculated ACFs, suggesting a slow relaxation as seen in Figure I.26. This effect is pronounced in the mutant case, where I have a significant fraction of hallucinated structures. While the prediction is biologically and functionally inaccurate, based on intuition, a symmetry-breaking motion should be characterized as the slowest process. Thus, the quality of the initial set of structures indeed matters as one would expect. However, reweighting methods or ensemble-driven generation are necessary to accurately predict these timescales.

I apply my proposed reweighting approach to remove the out-of-set samples. Figure I.26 shows the effect of this reweighting approach. I show the ACFs for functionally or dynamically relevant residues, such as the active site, 'ear', 'cheek', and flap tip.[296] As expected, the predicted timescales start to align with the directly calculated ACF from the 1 μs MD. Figure I.26b bottom panel clearly shows how reweighting aids these calculations.

Finally, to quantify the discrepancy and the effect of reweighting, I used dynamic time warping distances between the predicted ACF and the ACFs directly from MD samples. Figure I.27a shows the DTW matrix for the WT and mutant. Each row of the DTW matrix contains information for different MSA lengths, and each column corresponds to a residue. As mentioned in the main text, a lower value indicates that the predictions are closer. I can easily see that all

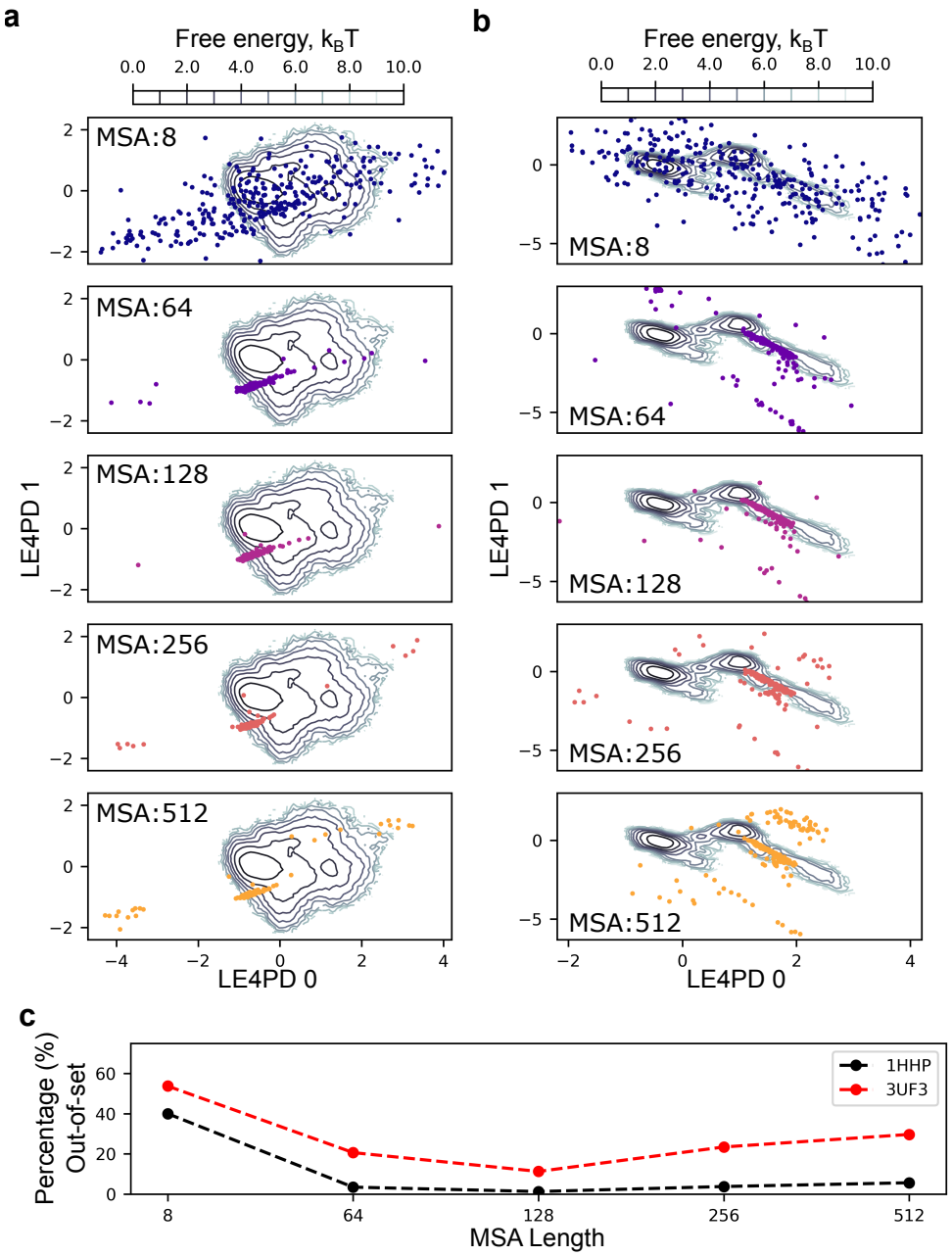


Figure I.25: Overlap of ML-generated ensemble with MD ensemble. a) Spread of the conformational ensemble is shown for the WT (PDB ID:1HHP) and b) mutant (PDB ID:3UF3). c) Shows the percentage out-of-set samples present for both WT and mutant HIV-1 protease dimers.

the predictions for the mutant behave similarly, as they are heavily biased by the hallucinated structures. Figure I.27b shows the same result after the reweighting procedure. Once these spurious structures were removed, patterns started to arise in the DTW matrix. This suggests that reweighting helps the predictions. Moreover, only after reweighting the good samples that were generated by the reduced MSA the approach shines, thus suggesting 1) there is indeed some relevant timescale information present in the generated ensembles, and 2) reweighting is required

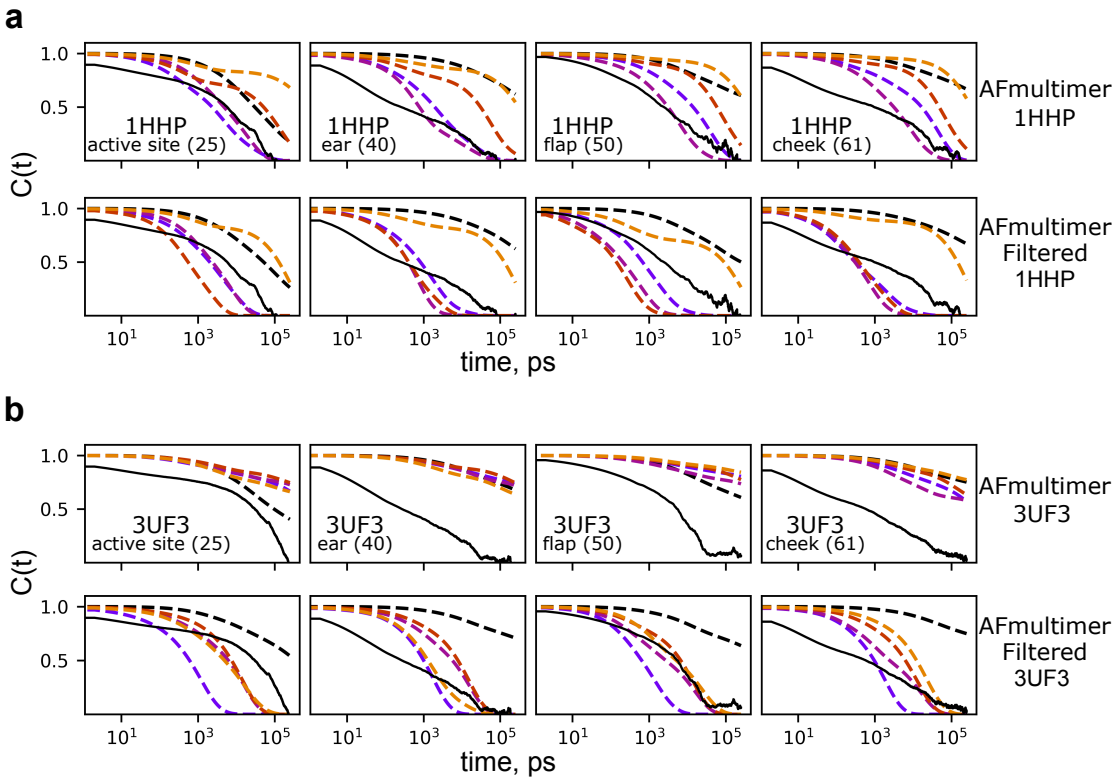


Figure I.26: Relaxation timescales for relevant residues. a) (Top) Shows the WT sequence's ACFs for the active site (residue 25), 'ear' (residue 25),flap tip (residue 25), and 'cheek' (residue 25). (Bottom) Shows the same ACFs calculated without the out-of-set distribution. b) Shows similar analysis for the mutant sequence.

to remove such conformations.

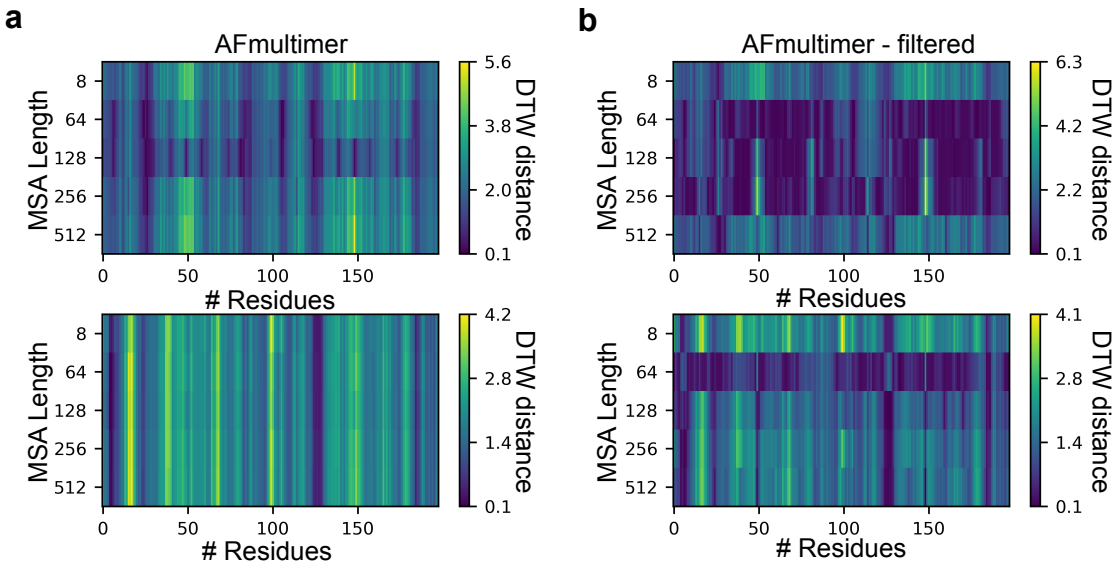


Figure I.27: DTW metric calculation. a) Shows the time warp distance between the MD ACF and my predicted ACF for the ensemble generated by the co-folding models. The top panel shows the calculated DTW matrix for the WT sequence, and the bottom panel shows that for the mutant. b) Shows the same analysis for the ACFs calculated without the out-of-set samples.

Bibliography

[1] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596:583–589, 2021.

[2] James AR Dalton, Isaias Lans, and Jesús Giraldo. Quantifying conformational changes in gpcrs: glimpse of a common functional mechanism. *BMC bioinformatics*, 16(1):1–15, 2015.

[3] Pratyush Tiwary, Jagannath Mondal, and B. J. Berne. How and when does an anticancer drug leave its binding site? *Science Advances*, 3(5), May 2017.

[4] Sonya M. Hanson, George Georghiou, Manish K. Thakur, W. Todd Miller, Joshua S. Rest, John D. Chodera, and Markus A. Seeliger. What makes a kinase promiscuous for inhibitors? *Cell Chemical Biology*, 26(3):390–399.e5, March 2019.

[5] Dedi Wang and Pratyush Tiwary. State predictive information bottleneck. *The Journal of Chemical Physics*, 154(13), April 2021.

[6] Pratyush Tiwary and Michele Parrinello. A time-independent free energy estimator for metadynamics. *The Journal of Physical Chemistry B*, 119(3):736–742, July 2014.

[7] Brajesh Narayan, Arman Fathizadeh, Clark Templeton, Peng He, Shima Arasteh, Ron Elber, Nicolae-Viorel Buchete, and Ron M. Levy. The transition between active and inactive conformations of abl kinase studied by rock climbing and milestoning. *Biochim. Biophys. Acta Gen. Subj.*, 1864(4):129508, April 2020.

[8] Vivek Modi and Roland L. Dunbrack. Defining a new nomenclature for the structures of active and inactive kinases. *Proceedings of the National Academy of Sciences*, 116(14):6818–6827, March 2019.

[9] Helen M. Berman, John Westbrook, Zukang Feng, Gary Gilliland, T. N. Bhat, Helge Weissig, Ilya N. Shindyalov, and Philip E. Bourne. The protein data bank. *Nucleic Acids Research*, 28(1):235–242, 2000.

[10] Minkyung Baek, Frank DiMaio, Ivan Anishchenko, Justas Dauparas, Sergey Ovchinnikov, Gyu Rie Lee, Jue Wang, Qian Cong, Lisa N. Kinch, R. Dustin Schaeffer, Claudia Millán, Hahnbeom Park, Carson Adams, Caleb R. Glassman, Andy DeGiovanni, Jose H. Pereira, Andria V. Rodrigues, Alberdina A. van Dijk, Ana C. Ebrecht, Diederik J. Opperman, Theo Sagmeister, Christoph Buhlheller, Tea Pavkov-Keller, Manoj K. Rathinaswamy, Udit Dalwadi, Calvin K. Yip, John E. Burke, K. Christopher Garcia, Nick V. Grishin, Paul D. Adams, Randy J. Read, and David Baker. Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, 373(6557):871–876, August 2021.

[11] Milot Mirdita, Konstantin Schütze, Yoshitaka Moriwaki, Lim Heo, Sergey Ovchinnikov, and Martin Steinegger. Colabfold: making protein folding accessible to all. *Nature Methods*, 19(6):679–682, May 2022.

[12] Gustaf Ahdritz, Nazim Bouatta, Sachin Kadyan, Qinghui Xia, William Gerecke, Mohammed AlQuraishi, et al. OpenFold: retraining AlphaFold2 yields new insights into its learning mechanisms and capacity for generalization. *Nature Methods*, 21:1514–1524, 2024.

[13] Zeming Lin, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Nikita Smetanin, Robert Verkuil, Ori Kabeli, Yaniv Shmueli, Allan Dos Santos Costa, Maryam Fazel-Zarandi, Tom Sercu, Salvatore Candido, and Alexander Rives. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637):1123–1130, 2023.

[14] Jeremy Wohlwend, Gabriele Corso, Saro Passaro, Mateo Reveiz, Kurt Leidal, Wojtek Swiderski, Tally Portnoi, Itamar Chinn, Jacob Silterra, Tommi Jaakkola, and Regina Barzilay. Boltz-1: Democratizing biomolecular interaction modeling. *bioRxiv*, 2024.

[15] Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J. Ballard, Joshua Bambrick, Sebastian Bodenstein, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 630:493–500, 2024.

[16] Katherine Henzler-Wildman and Dorothee Kern. Dynamic personalities of proteins. *Nature*, 450(7172):964–972, December 2007.

[17] Hans Frauenfelder, Stephen G. Sligar, and Peter G. Wolynes. The energy landscapes and motions of proteins. *Science*, 254(5038):1598–1603, December 1991.

[18] J D Bryngelson and P G Wolynes. Spin glasses and the statistical mechanics of protein folding. *Proceedings of the National Academy of Sciences*, 84(21):7524–7528, November 1987.

[19] Diego U. Ferreiro, Elizabeth A. Komives, and Peter G. Wolynes. Frustration in biomolecules. *Quarterly Reviews of Biophysics*, 47(4):285–363, 2014.

[20] Ken A Dill and Hue Sun Chan. From levinthal to pathways to funnels. *Nature structural biology*, 4(1):10–19, 1997.

[21] Murray Goodman. The bold legacy of emil fischer. *Journal of Peptide Science*, 9(9):594–603, 2003.

[22] F. Sanger and H. Tuppy. The amino-acid sequence in the phenylalanyl chain of insulin. 1. the identification of lower peptides from partial hydrolysates. *Biochemical Journal*, 49(4):463–481, September 1951.

[23] F. Sanger and H. Tuppy. The amino-acid sequence in the phenylalanyl chain of insulin. 2. the investigation of peptides from enzymic hydrolysates. *Biochemical Journal*, 49(4):481–490, September 1951.

[24] Linus Pauling, Robert B. Corey, and H. R. Branson. The structure of proteins: Two hydrogen-bonded helical configurations of the polypeptide chain. *Proceedings of the National Academy of Sciences*, 37(4):205–211, April 1951.

[25] G.N. Ramachandran, C. Ramakrishnan, and V. Sasisekharan. Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology*, 7(1):95–99, July 1963.

[26] J. C. KENDREW, R. E. DICKERSON, B. E. STRANDBERG, R. G. HART, D. R. DAVIES, D. C. PHILLIPS, and V. C. SHORE. Structure of myoglobin: A three-dimensional fourier synthesis at 2 Å. resolution. *Nature*, 185(4711):422–427, February 1960.

[27] M. F. PERUTZ, M. G. ROSSMANN, ANN F. CULLIS, HILARY MUIRHEAD, GEORG WILL, and A. C. T. NORTH. Structure of hæmoglobin: A three-dimensional fourier synthesis at 5.5-Å. resolution, obtained by x-ray analysis. *Nature*, 185(4711):416–422, February 1960.

[28] Christian B. Anfinsen. Principles that govern the folding of protein chains. *Science*, 181(4096):223–230, July 1973.

[29] Cyrus Levinthal. Are there pathways for protein folding? *Journal de Chimie Physique*, 65:44–45, 1968.

[30] Ken A. Dill, S. Banu Ozkan, M. Scott Shell, and Thomas R. Weikl. The protein folding problem. *Annual Review of Biophysics*, 37(1):289–316, June 2008.

[31] Diego U. Ferreiro, Elizabeth A. Komives, and Peter G. Wolynes. Frustration in biomolecules. *Quarterly Reviews of Biophysics*, 47(4):285–363, September 2014.

[32] Joseph D. Bryngelson, José Nelson Onuchic, Nicholas D. Socci, and Peter G. Wolynes. Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins: Structure, Function, and Bioinformatics*, 21(3):167–195, March 1995.

[33] José Nelson Onuchic, Zaida Luthey-Schulten, and Peter G. Wolynes. Theory of protein folding: The energy landscape perspective. *Annual Review of Physical Chemistry*, 48(1):545–600, October 1997.

[34] Ken A. Dill and Hue Sun Chan. From levinthal to pathways to funnels. *Nature Structural& Molecular Biology*, 4(1):10–19, January 1997.

[35] Andrej S˘ali, Eugene Shakhnovich, and Martin Karplus. How does a protein fold? *Nature*, 369(6477):248–251, May 1994.

[36] Eugene Shakhnovich and Alexander Gutin. Enumeration of all compact conformations of copolymers with random sequence of links. *The Journal of Chemical Physics*, 93(8):5967–5971, October 1990.

[37] Robert Zwanzig, Attila Szabo, and Biman Bagchi. Levinthal's paradox. *Proceedings of the National Academy of Sciences of the United States of America*, 89(1):20–22, 1992.

[38] Gregory R. Bowman, Xuhui Huang, and Vijay S. Pande. Using generalized ensemble simulations and markov state models to identify conformational states. *Methods*, 49(2):197–201, October 2009.

[39] Frank Noé, Christof Schütte, Eric Vanden-Eijnden, Lothar Reich, and Thomas R. Weikl. Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations. *Proceedings of the National Academy of Sciences*, 106(45):19011–19016, November 2009.

[40] Buyong Ma, Sandeep Kumar, Chung-Jung Tsai, and Ruth Nussinov. Folding funnels and binding mechanisms. *Protein Engineering, Design and Selection*, 12(9):713–720, September 1999.

[41] Chung-Jung Tsai, Sandeep Kumar, Buyong Ma, and Ruth Nussinov. Folding funnels, binding funnels, and protein function. *Protein Science*, 8(6):1181–1190, January 1999.

[42] David D Boehr, Ruth Nussinov, and Peter E Wright. The role of dynamic conformational ensembles in biomolecular recognition. *Nature Chemical Biology*, 5(11):789–796, October 2009.

[43] Christine A. Orengo and Janet M. Thornton. Protein families and their evolution—a structural perspective. *Annual Review of Biochemistry*, 74(1):867–900, June 2005.

[44] Andrej Šali and Tom L. Blundell. Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology*, 234(3):779–815, December 1993.

[45] Benjamin Webb and Andrej Sali. Comparative protein structure modeling using modeller. *Current Protocols in Bioinformatics*, 54(1), June 2016.

[46] D. T. Jones, W. R. Taylort, and J. M. Thornton. A new approach to protein fold recognition. *Nature*, 358(6381):86–89, July 1992.

[47] Kim T. Simons, Charles Kooperberg, Enoch Huang, and David Baker. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and bayesian scoring functions. *Journal of Molecular Biology*, 268(1):209–225, April 1997.

[48] Carol A. Rohl, Charlie E.M. Strauss, Kira M.S. Misura, and David Baker. *Protein Structure Prediction Using Rosetta*, page 66–93. Elsevier, 2004.

[49] Ulrike Göbel, Chris Sander, Reinhard Schneider, and Alfonso Valencia. Correlated mutations and residue contacts in proteins. *Proteins: Structure, Function, and Bioinformatics*, 18(4):309–317, April 1994.

[50] Faruck Morcos, Andrea Pagnani, Bryan Lunt, Arianna Bertolino, Debora S. Marks, Chris Sander, Riccardo Zecchina, José N. Onuchic, Terence Hwa, and Martin Weigt. Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proceedings of the National Academy of Sciences*, 108(49), November 2011.

[51] Magnus Ekeberg, Cecilia Lövkvist, Yueheng Lan, Martin Weigt, and Erik Aurell. Improved contact prediction in proteins: Using pseudolikelihoods to infer potts models. *Physical Review E*, 87(1), January 2013.

[52] Scott A Hollingsworth and Ron O Dror. Molecular dynamics simulation for all. *Neuron*, 99(6):1129–1143, September 2018.

[53] J. Andrew McCammon, Bruce R. Gelin, and Martin Karplus. Dynamics of folded proteins. *Nature*, 267(5612):585–590, June 1977.

[54] Martin Karplus and Gregory A. Petsko. Molecular dynamics simulations in biology. *Nature*, 347(6294):631–639, October 1990.

[55] Ron O. Dror, Robert M. Dirks, J.P. Grossman, Huafeng Xu, and David E. Shaw. Biomolecular simulation: A computational microscope for molecular biology. *Annual Review of Biophysics*, 41(1):429–452, June 2012.

[56] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The Journal of Physical Chemistry B*, 102(18):3586–3616, April 1998.

[57] Robert B. Best, Xiao Zhu, Jihyun Shim, Pedro E. M. Lopes, Jeetain Mittal, Michael Feig, and Alexander D. MacKerell. Optimization of the additive charmm all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain χ_1 and χ_2 dihedral angles. *Journal of Chemical Theory and Computation*, 8(9):3257–3273, August 2012.

[58] Giovanni Bussi, Davide Donadio, and Michele Parrinello. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics*, 126(1), January 2007.

[59] Giovanni Bussi, Tatyana Zykova-Timan, and Michele Parrinello. Isothermal-isobaric molecular dynamics using stochastic velocity rescaling. *The Journal of Chemical Physics*, 130(7), February 2009.

[60] Kresten Lindorff-Larsen, Stefano Piana, Ron O. Dror, and David E. Shaw. How fast-folding proteins fold. *Science*, 334(6055):517–520, October 2011.

[61] Stefano Piana, Kresten Lindorff-Larsen, and David E. Shaw. Protein folding kinetics and thermodynamics from atomistic simulation. *Proceedings of the National Academy of Sciences*, 109(44):17845–17850, July 2012.

[62] Yuji Sugita and Yuko Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters*, 314(1-2):141–151, November 1999.

[63] Massimiliano Meli and Giorgio Colombo. A hamiltonian replica exchange molecular dynamics (md) method for the study of folding, based on the analysis of the stabilization determinants of proteins. *International Journal of Molecular Sciences*, 14(6):12157–12169, June 2013.

[64] Omar Valsson, Pratyush Tiwary, and Michele Parrinello. Enhancing important fluctuations: Rare events and metadynamics from a conceptual viewpoint. *Annual Review of Physical Chemistry*, 67(1):159–184, May 2016.

[65] Jérôme Hénin, Tony Lelièvre, Michael R. Shirts, Omar Valsson, and Lucie Delemotte. Enhanced sampling methods for molecular dynamics simulations [article v1.0]. *Living Journal of Computational Molecular Science*, 4(1):1583, December 2022.

[66] G.M. Torrie and J.P. Valleau. Nonphysical sampling distributions in monte carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics*, 23(2):187–199, February 1977.

[67] Shankar Kumar, John M. Rosenberg, Djamal Bouzida, Robert H. Swendsen, and Peter A. Kollman. The weighted histogram analysis method for free-energy calculations on biomolecules. i. the method. *Journal of Computational Chemistry*, 13(8):1011–1021, October 1992.

[68] Benoît Roux. The calculation of the potential of mean force using computer simulations. *Computer Physics Communications*, 91(1-3):275–282, September 1995.

[69] Alessandro Laio and Michele Parrinello. Escaping free-energy minima. *Proceedings of the National Academy of Sciences*, 99(20):12562–12566, September 2002.

[70] Alessandro Barducci, Giovanni Bussi, and Michele Parrinello. Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Physical Review Letters*, 100(2), January 2008.

[71] Robert Zwanzig. Memory effects in irreversible thermodynamics. *Physical Review*, 124(4):983–992, November 1961.

[72] Robert B. Best, Gerhard Hummer, and William A. Eaton. Native contacts determine protein folding mechanisms in atomistic simulations. *Proceedings of the National Academy of Sciences*, 110(44):17874–17879, October 2013.

[73] Shams Mehdi, Zachary Smith, Lukas Herron, Ziyue Zou, and Pratyush Tiwary. Enhanced sampling with machine learning. *Annual Review of Physical Chemistry*, 75(1):347–370, June 2024.

[74] Shams Mehdi, Dedi Wang, Shashank Pant, and Pratyush Tiwary. Accelerating all-atom simulations and gaining mechanistic understanding of biophysical systems through state predictive information bottleneck. *Journal of Chemical Theory and Computation*, 18(5):3231–3238, April 2022.

[75] Suemin Lee, Dedi Wang, Markus A. Seeliger, and Pratyush Tiwary. Calculating protein–ligand residence times through state predictive information bottleneck based enhanced sampling. *Journal of Chemical Theory and Computation*, 20(14):6341–6349, July 2024.

[76] Ziyue Zou, Eric R. Beyerle, Sun-Ting Tsai, and Pratyush Tiwary. Driving and characterizing nucleation of urea and glycine polymorphs in water. *Proceedings of the National Academy of Sciences*, 120(7), February 2023.

[77] Ruiyu Wang, Shams Mehdi, Ziyue Zou, and Pratyush Tiwary. Is the Local Ion Density Sufficient to Drive NaCl Nucleation from the Melt and Aqueous Solution? *The Journal of Physical Chemistry B*, 128(4):1012–1021, January 2024.

[78] Ruiyu Wang and Pratyush Tiwary. Atomic scale insights into nacl nucleation in nanoconfined environments. *Chemical Science*, 15(37):15391–15398, 2024.

[79] Eric R. Beyerle, Shams Mehdi, and Pratyush Tiwary. Quantifying energetic and entropic pathways in molecular systems. *The Journal of Physical Chemistry B*, 126(21):3950–3960, May 2022.

[80] Eric R. Beyerle and Pratyush Tiwary. Thermodynamically optimized machine-learned reaction coordinates for hydrophobic ligand dissociation. *The Journal of Physical Chemistry B*, 128(3):755–767, January 2024.

[81] Vanessa J. Meraz, Ziyue Zou, and Pratyush Tiwary. Simulating crystallization in a colloidal system using state predictive information bottleneck based enhanced sampling. *The Journal of Physical Chemistry B*, 128(34):8207–8214, August 2024.

[82] S.D. Dunn, L.M. Wahl, and G.B. Gloor. Mutual information without the influence of phylogeny or entropy dramatically improves residue contact prediction. *Bioinformatics*, 24(3):333–340, December 2007.

[83] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding, 2018.

[84] Nicholas Bhattacharya, Neil Thomas, Roshan Rao, Justas Dauparas, Peter K. Koo, David Baker, Yun S. Song, and Sergey Ovchinnikov. Interpreting potts and transformer protein models through the lens of simplified attention. In *Biocomputing 2022*, page 34–45. WORLD SCIENTIFIC, November 2021.

[85] Riccardo Rende, Federica Gerace, Alessandro Laio, and Sebastian Goldt. Mapping of attention mechanisms to a generalized potts model. *Physical Review Research*, 6(2), April 2024.

[86] Roshan Rao, Jason Liu, Robert Verkuil, Joshua Meier, John F. Canny, Pieter Abbeel, Tom Sercu, and Alexander Rives. Msa transformer. *bioRχiv*, February 2021.

[87] Liwei Chang and Alberto Perez. Rapid estimation of protein folding pathways from sequence alone using alphafold2. *Nature Communications*, 17(1), December 2025.

[88] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. "why should i trust you?": Explaining the predictions of any classifier, 2016.

[89] Shams Mehdi and Pratyush Tiwary. Thermodynamics-inspired explanations of artificial intelligence. *Nature Communications*, 15(1), September 2024.

[90] Scott Lundberg and Su-In Lee. A unified approach to interpreting model predictions, 2017.

[91] Diego del Alamo, Davide Sala, Hassane S Mchaourab, and Jens Meiler. Sampling alternative conformational states of transporters and receptors with alphafold2. *eLife*, 11, March 2022.

[92] Vivek Modi and Roland L Dunbrack. Kincore: a web resource for structural classification of protein kinases and their inhibitors. *Nucleic Acids Research*, 50(D1):D654–D664, October 2021.

[93] James P. Roney and Sergey Ovchinnikov. State-of-the-art estimation of protein model accuracy using alphafold. *Physical Review Letters*, 129(23), November 2022.

[94] Bodhi P Vani, Akashnathan Aranganathan, Dedi Wang, and Pratyush Tiwary. Alphafold2-rave: From sequence to boltzmann ranking. *Journal of chemical theory and computation*, 19(14):4351–4354, 2023.

[95] Hannah K. Wayment-Steele, Adedolapo Ojoawo, Renee Otten, Julia M. Apitz, Warintra Pitsawong, Marc Hömberger, Sergey Ovchinnikov, Lucy Colwell, and Dorothee Kern. Predicting multiple conformations via sequence clustering and alphafold2. *Nature*, 625(7996):832–839, November 2023.

[96] Joseph W. Schafer, Myeongsang Lee, Devlina Chakravarty, Joseph F. Thole, Ethan A. Chen, and Lauren L. Porter. Sequence clustering confounds alphafold2. *Nature*, 638(8051):E8–E12, February 2025.

[97] Myeongsang Lee, Joseph W. Schafer, Jeshuwin Prabakaran, Devlina Chakravarty, Madeleine F. Clore, and Lauren L. Porter. Large-scale predictions of alternative protein conformations by alphafold2-based sequence association. *Nature Communications*, 16(1), July 2025.

[98] Richard A. Stein and Hassane S. Mchaourab. Speach_af: Sampling protein ensembles and conformational heterogeneity with alphafold2. *PLOS Computational Biology*, 18(8):e1010483, August 2022.

[99] Xingyue Guan, Qian-Yuan Tang, Weitong Ren, Mingchen Chen, Wei Wang, Peter G. Wolynes, and Wenfei Li. Predicting protein conformational motions using energetic frustration analysis and alphafold2. *Proceedings of the National Academy of Sciences*, 121(35), August 2024.

[100] Gwen R. Buel and Kylie J. Walters. Can alphafold2 predict the impact of missense mutations on structure? *Nature Structural& Molecular Biology*, 29(1):1–2, January 2022.

[101] Yihang Wang, João Marcelo Lamim Ribeiro, and Pratyush Tiwary. Past–future information bottleneck for sampling molecular reaction coordinate simultaneously with thermodynamics and kinetics. *Nature Communications*, 10(1), August 2019.

[102] Samuel L.C. Moors, Mario Hellings, Marc De Maeyer, Yves Engelborghs, and Arnout Ceulemans. Tryptophan rotamers as evidenced by x-ray, fluorescence lifetimes, and molecular dynamics modeling. *Biophysical Journal*, 91(3):816–823, August 2006.

[103] Dustin C. Becht, Moses J. Leavens, Baisen Zeng, Michael T. Rothfuss, Klára Briknarová, and Bruce E. Bowler. Residual structure in the denatured state of the fast-folding uba(1) domain from the human dna excision repair protein hhr23a. *Biochemistry*, 61(9):767–784, April 2022.

[104] Bonnie L. Bertolaet, Duncan J. Clarke, Meira Wolff, Mark H. Watson, Martha Henze, Gilles Divita, and Steven I. Reed. Uba domains of dna damage-inducible proteins interact with ubiquitin. *Nature Structural Biology*, 8(5):417–422, May 2001.

[105] Xin-heng He, Chong-zhao You, Hua-liang Jiang, Yi Jiang, H Eric Xu, and Xi Cheng. Alphafold2 versus experimental structures: evaluation on g protein-coupled receptors. *Acta Pharmacologica Sinica*, pages 1–7, 2022.

[106] Paul A Insel, Krishna Sriram, Shu Z Wiley, Andrea Wilderman, Trishna Katakia, Thalia McCann, Hiroshi Yokouchi, Lingzhi Zhang, Ross Corriden, Dongling Liu, et al. Gpcromics: Gpcr expression in cancer cells and tumors identifies new, potential biomarkers and therapeutic targets. *Frontiers in pharmacology*, 9:431, 2018.

[107] J Jiménez, S Doerr, G Martínez-Rosell, A S Rose, and G De Fabritiis. DeepSite: protein-binding site predictor using 3D-convolutional neural networks. *Bioinformatics*, 33(19):3036–3042, 05 2017.

[108] Radoslav Krivák and David Hoksza. P2rank: machine learning based tool for rapid and accurate prediction of ligand binding sites from protein structure. *Journal of Cheminformatics*, 10(1):39, Aug 2018.

[109] G A Huber and S Kim. Weighted-ensemble brownian dynamics simulations for protein association reactions. *Biophys J*, 70(1):97–110, January 1996.

[110] Bodhi P. Vani, Jonathan Weare, and Aaron R. Dinner. Computing transition path theory quantities with trajectory stratification. *The Journal of Chemical Physics*, 157(3):034106, 2022.

[111] Benoît Roux. String method with swarms-of-trajectories, mean drifts, lag time, and committor. *The Journal of Physical Chemistry A*, 125(34):7558–7571, August 2021.

[112] Ka Man Yip, Niels Fischer, Elham Paknia, Ashwin Chari, and Holger Stark. Atomic-resolution protein structure determination by cryo-em. *Nature*, 587(7832):157–161, 2020.

[113] Yinglong Miao and J Andrew McCammon. Gaussian accelerated molecular dynamics: Theory, implementation, and applications. In *Annual reports in computational chemistry*, volume 13, pages 231–278. Elsevier, 2017.

[114] Bodhi P Vani, Akashnathan Aranganathan, and Pratyush Tiwary. Exploring kinase asp-phe-gly (dfg) loop conformational stability with alphafold2-rave. *Journal of Chemical Information and Modeling*, 2023.

[115] G. Manning, D. B. Whyte, R. Martinez, T. Hunter, and S. Sudarsanam. The protein kinase complement of the human genome. *Science*, 298(5600):1912–1934, December 2002.

[116] Khushwant S. Bhullar, Naiara Orrego Lagarón, Eileen M. McGowan, Indu Parmar, Amitabh Jha, Basil P. Hubbard, and H. P. Vasantha Rupasinghe. Kinase-targeted cancer therapies: progress, challenges and future directions. *Mol. Cancer*, 17(1), February 2018.

[117] Philip Cohen, Darren Cross, and Pasi A. Jänne. Kinase drug discovery 20 years after imatinib: progress and future directions. *Nat. Rev. Drug Discov.*, 20(7):551–569, May 2021.

[118] Wei ZHANG and Hui Tu LIU. MAPK signal pathways in the regulation of cell proliferation in mammalian cells. *Cell Res.*, 12(1):9–18, March 2002.

[119] Mayur Gadiya and Goutam Chakraborty. Signaling by discoidin domain receptor 1 in cancer metastasis. *Cell Adhesion& Migration*, page 1–9, September 2018.

[120] Yilin Meng, Yen-lin Lin, and Benoît Roux. Computational study of the "dfg-flip" conformational transition in c-abl and c-src tyrosine kinases. *The Journal of Physical Chemistry B*, 119(4):1443–1456, January 2015.

[121] Tao Jiang, Zhenhao Liu, Wenlang Liu, Jiawen Chen, Zheng Zheng, and Mojie Duan. The conformational transition pathways and hidden intermediates in dfg-flip process of c-met kinase revealed by metadynamics simulations. *Journal of Chemical Information and Modeling*, 62(15):3651–3663, July 2022.

[122] Giorgio Saladino and Francesco Luigi Gervasio. New insights in protein kinase conformational dynamics. *Curr. Top. Med. Chem.*, 12(17):1889–1895, September 2012.

[123] Yilin Meng, Cen Gao, David K. Clawson, Shane Atwell, Marijane Russell, Michal Vieth, and Benoît Roux. Predicting the conformational variability of abl tyrosine kinase using molecular dynamics simulations and markov state models. *J. Chem. Theory Comput.*, 14(5):2721–2732, February 2018.

[124] Zachary Smith, Michael Strobel, Bodhi P. Vani, and Pratyush Tiwary. Graph attention site prediction (grasp): Identifying druggable binding sites using graph neural networks with attention. *Journal of Chemical Information and Modeling*, 64(7):2637–2644, March 2024.

[125] Xinyu Gu, Akashnathan Aranganathan, and Pratyush Tiwary. Empowering alphafold2 for protein conformation selective drug discovery with alphafold2-rave. *Elife*, 13:RP99702, 2024.

[126] R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, D. E. Shaw, M. Shelley, J. K. Perry, P. Francis, and P. S. Shenkin. Glide: a new approach for rapid, accurate docking and scoring. 1. method and assessment of docking accuracy. *Journal of medicinal chemistry*, 47(7):1739–1749, 2004.

[127] Thomas A Halgren, Robert B Murphy, Richard A Friesner, Hege S Beard, Leah L Frye, W Thomas Pollard, and Jay L Banks. Glide: a new approach for rapid, accurate docking and scoring. 2. enrichment factors in database screening. *Journal of medicinal chemistry*, 47(7):1750–1759, 2004.

[128] Richard A Friesner, Robert B Murphy, Matthew P Repasky, Leah L Frye, Jeremy R Greenwood, Thomas A Halgren, Paul C Sanschagrin, and Daniel T Mainz. Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein- ligand complexes. *Journal of medicinal chemistry*, 49(21):6177–6196, 2006.

[129] Woody Sherman, Tyler Day, Matthew P Jacobson, Richard A Friesner, and Ramy Farid. Novel procedure for modeling ligand/receptor induced fit effects. *Journal of medicinal chemistry*, 49(2):534–553, 2006.

[130] Woody Sherman, Hege S Beard, and Ramy Farid. Use of an induced fit receptor structure in virtual screening. *Chemical biology & drug design*, 67(1):83–84, 2006.

[131] Gabriele Corso, Hannes Stärk, Bowen Jing, Regina Barzilay, and Tommi Jaakkola. Diffdock: Diffusion steps, twists, and turns for molecular docking. *arXiv preprint arXiv:2210.01776*, 2022.

[132] Wei Lu, Jixian Zhang, Weifeng Huang, Ziqiao Zhang, Xiangyu Jia, Zhenyu Wang, Leilei Shi, Chengtao Li, Peter G Wolynes, and Shuangjia Zheng. Dynamicbind: predicting ligand-specific protein-ligand complex structure with a deep equivariant generative model. *Nature Communications*, 15(1):1071, 2024.

[133] Susanne Müller, Apirat Chaikuad, Nathanael S Gray, and Stefan Knapp. The ins and outs of selective kinase inhibitor development. *Nature chemical biology*, 11(11):818–821, 2015.

[134] Joan Gizzio, Abhishek Thakur, Allan Haldane, and Ronald M Levy. Evolutionary divergence in the conformational landscapes of tyrosine vs serine/threonine kinases. *Elife*, 11:e83368, 2022.

[135] Abhishek Thakur, Joan Gizzio, and Ronald M Levy. Potts hamiltonian models and molecular dynamics free energy simulations for predicting the impact of mutations on protein kinase stability. *The Journal of Physical Chemistry B*, 2024.

[136] Joan Gizzio, Abhishek Thakur, Allan Haldane, and Ronald Levy. Evolutionary sequence and structural basis for the distinct conformational landscapes of tyr and ser/thr kinases. *bioRxiv*, pages 2024–03, 2024.

[137] Jiankun Lyu, Nicholas Kapolka, Ryan Gumpper, Assaf Alon, Liang Wang, Manish K. Jain, Ximena Barros-Álvarez, Kensuke Sakamoto, Yoojoong Kim, Jeffrey DiBerto, Kuglae Kim, Isabella S. Glenn, Tia A. Tummino, Sijie Huang, John J. Irwin, Olga O. Tarkhanova, Yurii Moroz, Georgios Skiniotis, Andrew C. Kruse, Brian K. Shoichet, and Bryan L. Roth. Alphafold2 structures guide prospective ligand discovery. *Science*, page eadn6354, 2024.

[138] Matthew Holcomb, Ya-Ting Chang, David S Goodsell, and Stefano Forli. Evaluation of alphafold2 structures as docking targets. *Protein Science*, 32(1):e4530, 2023.

[139] V Scardino, JI Di Filippo, and CN Cavasotto. How good are alphafold models for dockingbased virtual screening? *iScience 26*, 2023.

[140] Anna M Díaz-Rovira, Helena Martín, Thijs Beuming, Lucía Díaz, Victor Guallar, and Soumya S Ray. Are deep learning structural models sufficiently accurate for virtual screening? application of docking algorithms to alphafold2 predicted structures. *Journal of Chemical Information and Modeling*, 63(6):1668–1674, 2023.

[141] Feng Ren, Xiao Ding, Min Zheng, Mikhail Korzinkin, Xin Cai, Wei Zhu, Alexey Mantsyzov, Alex Aliper, Vladimir Aladinskiy, Zhongying Cao, Shanshan Kong, Xi Long, Bonnie Hei Man Liu, Yingtao Liu, Vladimir Naumov, Anastasia Shneyderman, Ivan V. Ozerov, Ju Wang, Frank W. Pun, Daniil A. Polykovskiy, Chong Sun, Michael Levitt, Alán Aspuru-Guzik, and Alex Zhavoronkov. Alphafold accelerates artificial intelligence powered drug discovery: efficient discovery of a novel cdk20 small molecule inhibitor. *Chemical Science*, 14(6):1443–1452, 2023.

[142] Davide Sala, Peter W Hildebrand, and Jens Meiler. Biasing alphafold2 to predict gpcrs and kinases with user-defined functional or structural properties. *Frontiers in Molecular Biosciences*, 10:1121962, 2023.

[143] Hugo Guterres, Sang-Jun Park, Wei Jiang, and Wonpil Im. Ligand-binding-site refinement to generate reliable holo protein structure conformations from apo structures. *Journal of chemical information and modeling*, 61(1):535–546, 2020.

[144] Thijs Beuming, Helena Martín, Anna M Díaz-Rovira, Lucía Díaz, Victor Guallar, and Soumya S Ray. Are deep learning structural models sufficiently accurate for free-energy calculations? application of fep+ to alphafold2-predicted structures. *Journal of Chemical Information and Modeling*, 62(18):4351–4360, 2022.

[145] Yuqi Zhang, Marton Vass, Da Shi, Esam Abualrous, Jennifer M. Chambers, Nikita Chopra, Christopher Higgs, Koushik Kasavajhala, Hubert Li, Prajwal Nandekar, Hideyuki Sato, Edward B. Miller, Matthew P. Repasky, and Steven V. Jerome. Benchmarking refined and unrefined alphafold2 structures for hit discovery. *Journal of Chemical Information and Modeling*, 63(6):1656–1667, 2023.

[146] Dilek Coskun, Muyun Lihan, João PGLM Rodrigues, Márton Vass, Daniel Robinson, Richard A Friesner, and Edward B Miller. Using alphafold and experimental structures for the prediction of the structure and binding affinities of gpcr complexes via induced fit docking and free energy perturbation. *Journal of Chemical Theory and Computation*, 20(1):477–489, 2023.

[147] Carmen Al-Masri, Francesco Trozzi, Shu-Hang Lin, Oanh Tran, Navriti Sahni, Marcel Patek, Anna Cichonska, Balaguru Ravikumar, and Rayees Rahman. Investigating the conformational landscape of alphafold2-predicted protein kinase structures. *Bioinformatics Advances*, 3(1):vbad129, 2023.

[148] Bowen Jing, Bonnie Berger, and Tommi Jaakkola. Alphafold meets flow matching for generating protein ensembles. *arXiv preprint arXiv:2402.04845*, 2024.

[149] Artur Meller, Soumendranath Bhakat, Shahlo Solieva, and Gregory R Bowman. Accelerating cryptic pocket discovery using alphafold. *Journal of Chemical Theory and Computation*, 19(14):4355–4363, 2023.

[150] Gabriel Monteiro da Silva, Jennifer Y Cui, David C Dalgarno, George P Lisi, and Brenda M Rubenstein. High-throughput prediction of protein conformational distributions with subsampled alphafold2. *Nature Communications*, 15(1):2464, 2024.

[151] Jiankun Lyu, Sheng Wang, Trent E. Balius, Isha Singh, Anat Levit, Yurii S. Moroz, Matthew J. O'Meara, Tao Che, Enkhjargal Algaa, Kateryna Tolmachova, Andrey A. Tolmachev, Brian K. Shoichet, Bryan L. Roth, and John J. Irwin. Ultra-large library docking for discovering new chemotypes. *Nature*, 566(7743):224–229, 2019.

[152] Xinyu Gu, Venkata Sai Sreyas Adury, Akashnathan Aranganathan, Xinhao Zhuang, Kristen M. Varney, David J. Weber, and Pratyush Tiwary. Hierarchical af2rave for multiconformation virtual screening targeting s100 ca2+-binding proteins. *Journal of Chemical Theory and Computation*, 21(18):9214–9225, September 2025.

[153] Brian J Bender, Stefan Gahbauer, Andreas Luttens, Jiankun Lyu, Chase M Webb, Reed M Stein, Elissa A Fink, Trent E Balius, Jens Carlsson, John J Irwin, et al. A practical guide to large-scale docking. *Nature protocols*, 16(10):4799–4832, 2021.

[154] Danna B Zimmer, Patti Wright Sadosky, and David J Weber. Molecular mechanisms of s100-target protein interactions. *Microscopy research and technique*, 60(6):552–559, 2003.

[155] Nathan T Wright, Brian R Cannon, Danna B Zimmer, and David J Weber. S100a1: structure, function, and therapeutic potential. *Current chemical biology*, 3(2):138–145, 2009.

[156] Rosario Donato, B R cannon, Guglielmo Sorci, Francesca Riuzzi, K Hsu, D J Weber, and C L Geczy. Functions of s100 proteins. *Current molecular medicine*, 13(1):24–57, 2013.

[157] Anne R Bresnick, David J Weber, and Danna B Zimmer. S100 proteins in cancer. *Nature Reviews Cancer*, 15(2):96–109, 2015.

[158] Lauriaselle Afanador, Emily A Roltsch, Leigh Holcomb, Kerry S Campbell, David A Keeling, Yan Zhang, and Danna B Zimmer. The ca2+ sensor s100a1 modulates neuroinflammation, histopathology and akt activity in the psapp alzheimer's disease mouse model. *Cell Calcium*, 56(2):68–80, 2014.

[159] Ryan G. Coleman, Michael Carchia, Teague Sterling, John J. Irwin, and Brian K. Shoichet. Ligand pose and orientational sampling in molecular docking. *PLOS ONE*, 8(10):1–19, 10 2013.

[160] William J. Allen, Trent E. Balius, Sudipto Mukherjee, Scott R. Brozell, Demetri T. Moustakas, P. Therese Lang, David A. Case, Irwin D. Kuntz, and Robert C. Rizzo. Dock 6: Impact of new features and current docking performance. *Journal of Computational Chemistry*, 36(15):1132–1156, 2015.

[161] Kristen M Varney and David J Weber. Tbd. *Under preparation*, xx(x):xx–xx, 2025.

[162] Yujin Wu, Fangyu Liu, Isabella Glenn, Karla Fonseca-Valencia, Lu Paris, Yuyue Xiong, Steven V Jerome, Charles L Brooks III, and Brian K Shoichet. Identifying artifacts from large library docking. *Journal of Medicinal Chemistry*, 67(18):16796–16806, 2024.

[163] Andrew R Leach, Brian K Shoichet, and Catherine E Peishoff. Prediction of protein- ligand interactions. docking and scoring: successes and gaps. *Journal of medicinal chemistry*, 49(20):5851–5855, 2006.

[164] Alberto Borsatto, Obaeda Akkad, Ioannis Galdadas, Shumeng Ma, Shymaa Damfo, Shozeb Haider, Frank Kozielski, Carolina Estarellas, and Francesco Luigi Gervasio. Revealing druggable cryptic pockets in the nsp1 of sars-cov-2 and other β-coronaviruses by simulations and crystallography. *Elife*, 11, 2022.

[165] Federico Comitani and Francesco Luigi Gervasio. Exploring cryptic pockets formation in targets of pharmaceutical interest with swish. *Journal of chemical theory and computation*, 14(6):3321–3331, 2018.

[166] Alberto Borsatto, Eleonora Gianquinto, Valerio Rizzi, and Francesco Luigi Gervasio. Swish-x, an expanded approach to detect cryptic pockets in proteins and at protein–protein interfaces. *Journal of Chemical Theory and Computation*, 20(8):3335–3348, 2024.

[167] Da Teng, Vanessa J. Meraz, Akashnathan Aranganathan, Xinyu Gu, and Pratyush Tiwary. af2rave: protein ensemble generation with physics-based sampling. *Digital Discovery*, 4(8):2052–2061, 2025.

[168] Akashnathan Aranganathan, Xinyu Gu, Dedi Wang, Bodhi P Vani, and Pratyush Tiwary. Modeling boltzmann-weighted structural ensembles of proteins using artificial intelligence–based methods. *Current Opinion in Structural Biology*, 91:103000, 2025.

[169] Matthew R Masters, Amr H Mahmoud, and Markus A Lill. Do deep learning models for co-folding learn the physics of protein-ligand interactions? *bioRxiv*, pages 2024–06, 2024.

[170] Valerio Mariani, Marco Biasini, Alessandro Barbato, and Torsten Schwede. lddt: a local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics*, 29(21):2722–2728, August 2013.

[171] Nathan T Wright, Kristen M Varney, Karen C Ellis, Joseph Markowitz, Rossitza K Gitti, Danna B Zimmer, and David J Weber. The three-dimensional solution structure of ca2+-bound s100a1 as determined by nmr spectroscopy. *Journal of molecular biology*, 353(2):410–426, 2005.

[172] Rui Lai, Guohui Li, and Qiang Cui. Flexibility of binding site is essential to the ca2+ selectivity in ef-hand calcium-binding proteins. *Journal of the American Chemical Society*, 146(11):7628–7639, 2024.

[173] Shankar Kumar, John M. Rosenberg, Djamal Bouzida, Robert H. Swendsen, and Peter A. Kollman. The weighted histogram analysis method for free-energy calculations on biomolecules. i. the method. *Journal of Computational Chemistry*, 13(8):1011–1021, 1992.

[174] Michael C Cavalier, Mohd Imran Ansari, Adam D Pierce, Paul T Wilder, Laura E McKnight, E Prabhu Raman, David B Neau, Padmavani Bezawada, Milad J Alasady, Thomas H Charpentier, et al. Small molecule inhibitors of ca2+-s100b reveal two protein conformations. *Journal of medicinal chemistry*, 59(2):592–608, 2016.

[175] Richard R Rustandi, Donna M Baldisseri, Alexander C Drohat, and David J Weber. Structural changes in the c-terminus of ca2+-bound rat s100b ($\beta\beta$) upon binding to a peptide derived from the c-terminal regulatory domain of p53. *Protein science*, 8(9):1743–1751, 1999.

[176] Douglas Kligman and Dana C Hilt. The s100 protein family. *Trends in biochemical sciences*, 13(11):437–443, 1988.

[177] Michael C. Cavalier, Adam D. Pierce, Paul T. Wilder, Milad J. Alasady, Kira G. Hartman, David B. Neau, Timothy L. Foley, Ajit Jadhav, David J. Maloney, Anton Simeonov,

Eric A. Toth, and David J. Weber. Covalent small molecule inhibitors of ca2+-bound s100b. *Biochemistry*, 53(42):6628–6640, 2014. PMID: 25268459.

[178] Xtra of University of Califonia at San Fransisco with support from NIH P41-GM103311. Ucsf chimera - a visualization system for exploratory research and analysis. *J. Comput. Chem.*, 25:1605–1612, 2004.

[179] Surbhi Sharma and Martin R. Schiller. The carboxy-terminus, a key regulator of protein function. *Critical Reviews in Biochemistry and Molecular Biology*, 54(2):85–102, 2019.

[180] Merissa Brousseau, Da Teng, Nathan E. Thomas, Gregory A. Voth, and Katherine A. Henzler-Wildman. The C-terminus of the multi-drug efflux pump EmrE prevents proton leak by gating transport. *eLife*, 14, 2025.

[181] K. Christopher Garcia, Massimo Degano, Larry R. Pease, Mingdong Huang, Per A. Peterson, Luc Teyton, and Ian A. Wilson. Structural basis of plasticity in t cell receptor recognition of a self peptide-mhc antigen. *Science*, 279(5354):1166–1172, February 1998.

[182] Ellis L. Reinherz, Kemin Tan, Lei Tang, Petra Kern, Jin-huan Liu, Yi Xiong, Rebecca E. Hussey, Alex Smolyar, Brian Hare, Rongguang Zhang, Andrzej Joachimiak, Hsiu-Ching Chang, Gerhard Wagner, and Jia-huai Wang. The crystal structure of a t cell receptor in complex with peptide and mhc class ii. *Science*, 286(5446):1913–1921, December 1999.

[183] Christopher Szeto, Christian A. Lobos, Andrea T. Nguyen, and Stephanie Gras. Tcr recognition of peptide–mhc-i: Rule makers and breakers. *International Journal of Molecular Sciences*, 22(1):68, December 2020.

[184] Richard Evans, Michael O'Neill, Alexander Pritzel, Natasha Antropova, Andrew Senior, Tim Green, Augustin Žídek, Russ Bates, Sam Blackwell, Jason Yim, Olaf Ronneberger, Sebastian Bodenstein, Michal Zielinski, Alex Bridgland, Anna Potapenko, Andrew Cowie, Kathryn Tunyasuvunakool, Rishub Jain, Ellen Clancy, Pushmeet Kohli, John Jumper, and Demis Hassabis. Protein complex prediction with alphafold-multimer. October 2021.

[185] Saro Passaro, Gabriele Corso, Jeremy Wohlwend, Mateo Reveiz, Stephan Thaler, Vignesh Ram Somnath, Noah Getz, Tally Portnoi, Julien Roy, Hannes Stark, David Kwabi-Addo, Dominique Beaini, Tommi Jaakkola, and Regina Barzilay. Boltz-2: Towards accurate and efficient binding affinity prediction. June 2025.

[186] Chai Discovery, Jacques Boitreaud, Jack Dent, Matthew McPartlon, Joshua Meier, Vinicius Reis, Alex Rogozhnikov, and Kevin Wu. Chai-1: Decoding the molecular interactions of life. October 2024.

[187] Philip Bradley. Structure-based prediction of t cell receptor:peptide-mhc interactions. *eLife*, 12, January 2023.

[188] Rui Yin, Helder V Ribeiro-Filho, Valerie Lin, Ragul Gowthaman, Melyssa Cheung, and Brian G Pierce. Tcrmodel2: high-resolution modeling of t cell receptor recognition using deep learning. *Nucleic Acids Research*, 51(W1):W569–W576, May 2023.

[189] Jason R. Devlin, Jesus A. Alonso, Cory M. Ayres, Grant L. J. Keller, Sara Bobisse, Craig W. Vander Kooi, George Coukos, David Gfeller, Alexandre Harari, and Brian M. Baker. Structural dissimilarity from self drives neoepitope escape from immune tolerance. *Nature Chemical Biology*, 16(11):1269–1276, August 2020.

[190] Roy A. Mariuzza, Daichao Wu, and Brian G. Pierce. Structural basis for t cell recognition of cancer neoantigens and implications for predicting neoepitope immunogenicity. *Frontiers in Immunology*, 14, November 2023.

[191] Valerie Lin, Melyssa Cheung, Ragul Gowthaman, Maya Eisenberg, Brian M Baker, and Brian G Pierce. Tcr3d 2.0: expanding the t cell receptor structure database with new structures, tools and interactions. *Nucleic Acids Research*, 53(D1):D604–D608, 2024.

[192] L. Evans, M. K. Cameron, and P. Tiwary. Computing committors via mahalanobis diffusion maps with enhanced sampling data. *The Journal of Chemical Physics*, 157(21):214107, 12 2022.

[193] Anthony J. Clark, Pratyush Tiwary, Ken Borrelli, Shulu Feng, Edward B. Miller, Robert Abel, Richard A. Friesner, and B. J. Berne. Prediction of protein–ligand binding poses via a combination of induced fit docking and metadynamics simulations. *Journal of Chemical Theory and Computation*, 12(6):2990–2998, May 2016.

[194] Pavan Ravindra, Zachary Smith, and Pratyush Tiwary. Automatic mutual information noise omission (amino): generating order parameters for molecular systems. *Molecular Systems Design & Engineering*, 5(1):339–348, 2020.

[195] Justin M. Mirazee, Akashnathan Aranganathan, Sooraj Achar, Dongya Jia, Xiang Chen, Bodhi P. Vani, Christopher D. Chien, Marie Pouzolles, Kniya DeDe, Philippe Youkharibache, Kylie Walters, Pratyush Tiwary, Gregoire Altan-Bonnet, and Naomi Taylor. Enhancing car t cell therapy through in silico modeling of hinge sequence length based on epitope location. *Biophysical Journal*, 123(3):552a, February 2024.

[196] Z Eshhar, T Waks, G Gross, and D G Schindler. Specific activation and targeting of cytotoxic lymphocytes through chimeric single chains consisting of antibody-binding domains and the gamma or zeta subunits of the immunoglobulin and t-cell receptors. *Proc. Natl. Acad. Sci. U. S. A.*, 90(2):720–724, January 1993.

[197] Helene M Finney, Alastair D G Lawson, Christopher R Bebbington, and A Neil C Weir. Chimeric receptors providing both primary and costimulatory signaling in t cells from a single gene product. *The Journal of Immunology*, 161(6):2791–2797, September 1998.

[198] G Gross, T Waks, and Z Eshhar. Expression of immunoglobulin-t-cell receptor chimeric molecules as functional receptors with antibody-type specificity. *Proc. Natl. Acad. Sci. U. S. A.*, 86(24):10024–10028, December 1989.

[199] John Maher, Renier J Brentjens, Gertrude Gunset, Isabelle Rivière, and Michel Sadelain. Human t-lymphocyte cytotoxicity and proliferation directed by a single chimeric TCRzeta /CD28 receptor. *Nat. Biotechnol.*, 20(1):70–75, January 2002.

[200] Robbie G Majzner and Crystal L Mackall. Tumor antigen escape from CAR t-cell therapy. *Cancer Discov.*, 8(10):1219–1226, October 2018.

[201] Shannon L Maude, Theodore W Laetsch, Jochen Buechner, Susana Rives, Michael Boyer, Henrique Bittencourt, Peter Bader, Michael R Verneris, Heather E Stefanski, Gary D Myers, Muna Qayed, Barbara De Moerloose, Hidefumi Hiramatsu, Krysta Schlis, Kara L Davis, Paul L Martin, Eneida R Nemecek, Gregory A Yanik, Christina Peters, Andre Baruchel, Nicolas Boissel, Francoise Mechinaud, Adriana Balduzzi, Joerg Krueger, Carl H June, Bruce L Levine, Patricia Wood, Tetiana Taran, Mimi Leung, Karen T Mueller, Yiyun Zhang, Kapildeb Sen, David Lebwohl, Michael A Pulsipher, and Stephan A Grupp. Tisagenlecleucel in children and young adults with b-cell lymphoblastic leukemia. *N. Engl. J. Med.*, 378(5):439–448, February 2018.

[202] Jae H Park, Isabelle Rivière, Mithat Gonen, Xiuyan Wang, Brigitte Sénéchal, Kevin J Curran, Craig Sauter, Yongzeng Wang, Bianca Santomasso, Elena Mead, Mikhail Roshal, Peter Maslak, Marco Davila, Renier J Brentjens, and Michel Sadelain. Long-term follow-up of CD19 CAR therapy in acute lymphoblastic leukemia. *N. Engl. J. Med.*, 378(5):449–459, February 2018.

[203] Nirav N Shah, Bryon D Johnson, Dina Schneider, Fenlu Zhu, Aniko Szabo, Carolyn A Keever-Taylor, Winfried Krueger, Andrew A Worden, Michael J Kadan, Sharon Yim, Ashley Cunningham, Mehdi Hamadani, Timothy S Fenske, Boro Dropulić, Rimas Orentas, and Parameswaran Hari. Bispecific anti-CD20, anti-CD19 CAR T cells for relapsed B cell malignancies: a phase 1 dose escalation and expansion trial. *Nat. Med.*, 26(10):1569–1575, October 2020.

[204] Jay Y Spiegel, Shabnum Patel, Lori Muffly, Nasheed M Hossain, Jean Oak, John H Baird, Matthew J Frank, Parveen Shiraz, Bita Sahaf, Juliana Craig, Maria Iglesias, Sheren Younes, Yasodha Natkunam, Michael G Ozawa, Eric Yang, John Tamaresis, Harshini Chinnasamy, Zach Ehlinger, Warren Reynolds, Rachel Lynn, Maria Caterina Rotiroti, Nikolaos Gkitsas, Sally Arai, Laura Johnston, Robert Lowsky, Robbie G Majzner, Everett Meyer, Robert S Negrin, Andrew R Rezvani, Surbhi Sidana, Judith Shizuru, Wen-Kai Weng, Chelsea Mullins, Allison Jacob, Ilan Kirsch, Magali Bazzano, Jing Zhou, Sean Mackay, Scott J Bornheimer, Liora Schultz, Sneha Ramakrishna, Kara L Davis, Katherine A Kong, Nirali N Shah, Haiying Qin, Terry Fry, Steven Feldman, Crystal L Mackall, and David B Miklos. CAR T cells with dual targeting of CD19 and CD22 in adult patients with recurrent or refractory B cell malignancies: a phase 1 trial. *Nat. Med.*, 27(8):1419–1431, August 2021.

[205] Venugopal Gudipati, Julian Rydzek, Iago Doel-Perez, Vasco Dos Reis Gonçalves, Lydia Scharf, Sebastian Königsberger, Elisabeth Lobner, Renate Kunert, Hermann Einsele, Hannes Stockinger, Michael Hudecek, and Johannes B Huppa. Inefficient CAR-proximal signaling blunts antigen sensitivity. *Nat. Immunol.*, 21(8):848–856, August 2020.

[206] Daniel T Harris, Marlies V Hager, Sheena N Smith, Qi Cai, Jennifer D Stone, Philipp Kruger, Melissa Lever, Omer Dushek, Thomas M Schmitt, Philip D Greenberg, and

David M Kranz. Comparison of T cell activities mediated by human TCRs and CARs that use the same recognition domains. *J. Immunol.*, 200(3):1088–1100, February 2018.

[207] Dina Schneider, Ying Xiong, Darong Wu, Peirong Hu, Leah Alabanza, Brittany Steimle, Hasan Mahmud, Kim Anthony-Gonda, Winfried Krueger, Zhongyu Zhu, Dimiter S Dimitrov, Rimas J Orentas, and Boro Dropulić. Trispecific CD19-CD20-CD22-targeting duoCAR-T cells eliminate antigen-heterogeneous B cell tumors in preclinical models. *Sci. Transl. Med.*, 13(586):eabc6401, March 2021.

[208] Haneen Shalabi, Haiying Qin, Angela Su, Bonnie Yates, Pamela L Wolters, Seth M Steinberg, John A Ligon, Sara Silbert, Kniya DéDé, Mehdi Benzaoui, Sophia Goldberg, Sooraj Achar, Dina Schneider, Shilpa A Shahani, Lauren Little, Toni Foley, John C Molina, Sandhya Panch, Crystal L Mackall, Daniel W Lee, Christopher D Chien, Marie Pouzolles, Mark Ahlman, Constance M Yuan, Hao-Wei Wang, Yanyu Wang, Jon Inglefield, Mary Anne Toledo-Tamula, Staci Martin, Steven L Highfill, Gregoire Altan-Bonnet, David Stroncek, Terry J Fry, Naomi Taylor, and Nirali N Shah. CD19/22 CAR T cells in children and young adults with B-ALL: phase 1 results and development of a novel bicistronic CAR. *Blood*, 140(5):451–463, August 2022.

[209] Patrick A Baeuerle, Jian Ding, Ekta Patel, Niko Thorausch, Holly Horton, Jessica Gierut, Irene Scarfo, Rashmi Choudhary, Olga Kiner, Janani Krishnamurthy, Bonnie Le, Anna Morath, G Christian Baldeviano, Justin Quinn, Patrick Tavares, Qi Wei, Solly Weiler, Marcela V Maus, Daniel Getts, Wolfgang W Schamel, and Robert Hofmeister. Synthetic TRuC receptors engaging the complete T cell receptor for potent anti-tumor response. *Nat. Commun.*, 10(1):2087, May 2019.

[210] Yue Liu, Guangna Liu, Jiasheng Wang, Zhe-Yu Zheng, Lemei Jia, Wei Rui, Daosheng Huang, Zhi-Xiao Zhou, Liqun Zhou, Xin Wu, Song Lin, Xueqiang Zhao, and Xin Lin. Chimeric STAR receptors using TCR machinery mediate robust responses against solid tumors. *Sci. Transl. Med.*, 13(586):eabb5191, March 2021.

[211] Jorge Mansilla-Soto, Justin Eyquem, Sascha Haubner, Mohamad Hamieh, Judith Feucht, Noémie Paillon, Andrés Ernesto Zucchetti, Zhuoning Li, Maria Sjöstrand, Pieter L Lindenbergh, Michelle Saetersmoen, Anton Dobrin, Mathieu Maurin, Archana Iyer, Andreina Garcia Angus, Matthew M Miele, Zeguo Zhao, Theodoros Giavridis, Sjoukje J C van der Stegen, Fella Tamzalit, Isabelle Rivière, Morgan Huse, Ronald C Hendrickson, Claire Hivroz, and Michel Sadelain. HLA-independent T cell receptors for targeting tumors with low antigen density. *Nat. Med.*, 28(2):345–352, February 2022.

[212] Taisuke Kondo, François X P Bourassa, Sooraj Achar, Justyn DuSold, Pablo F Céspedes, Makoto Ando, Alka Dwivedi, Josquin Moraly, Christopher Chien, Saliha Majdoul, Adam L Kenet, Madison Wahlsten, Audun Kvalvaag, Edward Jenkins, Sanghyun P Kim, Catherine M Ade, Zhiya Yu, Guillaume Gaud, Marco Davila, Paul Love, James C Yang, Michael L Dustin, Grégoire Altan-Bonnet, Paul François, and Naomi Taylor. Engineering TCR-controlled fuzzy logic into CAR T cells enhances therapeutic specificity. *Cell*, 188(9):2372–2389.e35, May 2025.

[213] Robbie G Majzner, Skyler P Rietberg, Elena Sotillo, Rui Dong, Vipul T Vachharajani, Louai Labanieh, June H Myklebust, Meena Kadapakkam, Evan W Weber, Aidan M Tousley, Rebecca M Richards, Sabine Heitzeneder, Sang M Nguyen, Volker Wiebking, Johanna Theruvath, Rachel C Lynn, Peng Xu, Alexander R Dunn, Ronald D Vale, and Crystal L Mackall. Tuning the antigen density requirement for CAR t-cell activity. *Cancer Discov.*, 10(5):702–723, May 2020.

[214] Michael C Milone, Jonathan D Fish, Carmine Carpenito, Richard G Carroll, Gwendolyn K Binder, David Teachey, Minu Samanta, Mehdi Lakhal, Brian Gloss, Gwenn Danet-Desnoyers, Dario Campana, James L Riley, Stephan A Grupp, and Carl H June. Chimeric receptors containing CD137 signal transduction domains mediate enhanced survival of T cells and increased antileukemic efficacy in vivo. *Mol. Ther.*, 17(8):1453–1464, August 2009.

[215] Varvara Folimonova, Xiang Chen, Hitendra Negi, Charles D Schwieters, Jess Li, R Andrew Byrd, Naomi Taylor, Philippe Youkharibache, and Kylie J Walters. CD28 hinge used in chimeric antigen receptor (CAR) t-cells exhibits local structure and conformational exchange amidst global disorder. *Commun. Biol.*, 7(1):1072, August 2024.

[216] Xiang Chen, Justin M Mirazee, Katarzyna A Skorupka, Hiroshi Matsuo, Philippe Youkharibache, Naomi Taylor, and Kylie J Walters. The CD8α hinge is intrinsically disordered with a dynamic exchange that includes proline cis-trans isomerization. *J. Magn. Reson.*, 340(107234):107234, July 2022.

[217] Xiuli Wang, Wen-Chung Chang, Chinglam W Wong, David Colcher, Mark Sherman, Julie R Ostberg, Stephen J Forman, Stanley R Riddell, and Michael C Jensen. A transgene-encoded cell surface polypeptide for selection, in vivo tracking, and ablation of engineered cells. *Blood*, 118(5):1255–1263, August 2011.

[218] Raanan Greenman, Yoav Pizem, Maya Haus-Cohen, Alona Goor, Guy Horev, Galit Denkberg, Keren Sinik, Yael Elbaz, Vered Bronner, Anat Globerson Levin, Galit Horn, Shai Shen-Orr, and Yoram Reiter. Shaping functional avidity of CAR T cells: Affinity, avidity, and antigen density that regulate response. *Mol. Cancer Ther.*, 20(5):872–884, May 2021.

[219] Mark B Leick, Harrison Silva, Irene Scarfò, Rebecca Larson, Bryan D Choi, Amanda A Bouffard, Kathleen Gallagher, Andrea Schmidts, Stefanie R Bailey, Michael C Kann, Max Jan, Marc Wehrli, Korneel Grauwet, Nora Horick, Matthew J Frigault, and Marcela V Maus. Non-cleavable hinge enhances avidity and expansion of CAR-T cells for acute myeloid leukemia. *Cancer Cell*, 40(5):494–508.e5, May 2022.

[220] Leidy D Caraballo Galva, Xiaotao Jiang, Mohamed S Hussein, Huajun Zhang, Rui Mao, Pierce Brody, Yibing Peng, Aiwu Ruth He, Mercy Kehinde-Ige, Ramses Sadek, Xiangguo Qiu, Huidong Shi, and Yukai He. Novel low-avidity glypican-3 specific CARTs resist exhaustion and mediate durable antitumor effects against HCC. *Hepatology*, 76(2):330–344, August 2022.

[221] Rui Mao, Wanqing Kong, and Yukai He. The affinity of antigen-binding domain on the antitumor efficacy of CAR T cells: Moderate is better. *Front. Immunol.*, 13:1032403, October 2022.

[222] Changhao He, Jorge Mansilla-Soto, Nandish Khanra, Mohamad Hamieh, Victor Bustos, Alice J Paquette, Andreina Garcia Angus, Derek M Shore, William J Rice, George Khelashvili, Michel Sadelain, and Joel R Meyerson. CD19 CAR antigen engagement mechanisms and affinity tuning. *Sci. Immunol.*, 8(81):eadf1426, March 2023.

[223] Michael L Olson, Erica R Vander Mause, Sabarinath V Radhakrishnan, Joshua D Brody, Aaron P Rapoport, Alana L Welm, Djordje Atanackovic, and Tim Luetkens. Low-affinity CAR T cells exhibit reduced trogocytosis, preventing rapid antigen loss, and increasing CAR T cell expansion. *Leukemia*, 36(7):1943–1946, July 2022.

[224] Sooraj R Achar, François X P Bourassa, Thomas J Rademaker, Angela Lee, Taisuke Kondo, Emanuel Salazar-Cavazos, John S Davies, Naomi Taylor, Paul François, and Grégoire Altan-Bonnet. Universal antigen encoding of T cell activation from high-dimensional cytokine dynamics. *Science*, 376(6595):880–884, May 2022.

[225] Chengcheng Liao, Lin Zeng, Shengjuan Lu, Shaocu Zheng, Baoping Guo, Qing Ke, Mingyue Wang, Jie Sun, Chao Rong, Sha He, Dani Zhong, Mei Huang, Xiaohong Tan, and Hong Cen. Comparison of the efficacy and safety of axi-cel and Tisa-Cel based on meta-analysis. *J. Cancer*, 15(17):5729–5741, September 2024.

[226] Caron A Jacobson, Javier Munoz, Fang Sun, Steve Kanters, Eve H Limbrick-Oldfield, Clare Spooner, Krystal Mignone, Francis Ayuk, Robin Sanderson, James Whitmore, Yuanyuan Wang, Hairong Xu, and Michael Dickinson. Real-world outcomes with chimeric antigen receptor T cell therapies in large B cell lymphoma: A systematic review and meta-analysis. *Transplant. Cell. Ther.*, 30(1):77.e1–77.e15, January 2024.

[227] Emmanuel Bachy, Steven Le Gouill, Roberta Di Blasi, Pierre Sesques, Guillaume Manson, Guillaume Cartron, David Beauvais, Louise Roulin, François Xavier Gros, Marie Thérèse Rubio, Pierre Bories, Jacques Olivier Bay, Cristina Castilla Llorente, Sylvain Choquet, René-Olivier Casasnovas, Mohamad Mohty, Stéphanie Guidez, Magalie Joris, Michaël Loschi, Sylvain Carras, Julie Abraham, Adrien Chauchet, Laurianne Drieu La Rochelle, Bénédicte Deau-Fischer, Olivier Hermine, Thomas Gastinne, Jean Jacques Tudesq, Elodie Gat, Florence Broussais, Catherine Thieblemont, Roch Houot, and Franck Morschhauser. A real-world comparison of tisagenlecleucel and axicabtagene ciloleucel CAR T cells in relapsed or refractory diffuse large B cell lymphoma. *Nat. Med.*, 28(10):2145–2154, October 2022.

[228] Qian Xiao, Xinyan Zhang, Liqun Tu, Jian Cao, Christian S Hinrichs, and Xiaolei Su. Size-dependent activation of CAR-T cells. *Sci. Immunol.*, 7(74):eabl3995, August 2022.

[229] Chiara Cassioli, Laura Patrussi, Salvatore Valitutti, and Cosima T Baldari. Learning from TCR signaling and immunological synapse assembly to build new chimeric antigen receptors (CARs). *Int. J. Mol. Sci.*, 23(22):14255, November 2022.

[230] Yiwei Xiong, Kendra A Libby, and Xiaolei Su. The physical landscape of CAR-T synapse. *Biophys. J.*, 123(15):2199–2210, August 2024.

[231] A J Davenport, R S Cross, K A Watson, Y Liao, W Shi, H M Prince, P A Beavis, J A Trapani, M H Kershaw, D S Ritchie, P K Darcy, P J Neeson, and M R Jenkins. Chimeric antigen receptor T cells form nonclassical and potent immune synapses driving rapid cytotoxicity. *Proc. Natl. Acad. Sci. U. S. A.*, 115(9):E2068–E2076, February 2018.

[232] Lakshmi Balagopalan, Taylor Moreno, Haiying Qin, Benjamin C Angeles, Taisuke Kondo, Jason Yi, Katherine M McIntire, Neriah Alvinez, Sandeep Pallikkuth, Mariah E Lee, Hidehiro Yamane, Andy D Tran, Philippe Youkharibache, Raul E Cachau, Naomi Taylor, and Lawrence E Samelson. Generation of antitumor chimeric antigen receptors incorporating T cell signaling motifs. *Sci. Signal.*, 17(846):eadp8569, July 2024.

[233] Alexander I Salter, Anusha Rajan, Jacob J Kennedy, Richard G Ivey, Sarah A Shelby, Isabel Leung, Megan L Templeton, Vishaka Muhunthan, Valentin Voillet, Daniel Sommermeyer, Jeffrey R Whiteaker, Raphael Gottardo, Sarah L Veatch, Amanda G Paulovich, and Stanley R Riddell. Comparative analysis of TCR and CAR signaling informs CAR designs with superior antigen sensitivity and in vivo function. *Sci. Signal.*, 14(697):eabe2606, August 2021.

[234] Markus Barden, Astrid Holzinger, Lukas Velas, Marianna Mezősi-Csaplár, Árpád Szöőr, György Vereb, Gerhard J Schütz, Andreas A Hombach, and Hinrich Abken. CAR and TCR form individual signaling synapses and do not cross-activate, however, can co-operate in T cell activation. *Front. Immunol.*, 14:1110482, February 2023.

[235] Maria C Ramello, Ismahène Benzaïd, Brent M Kuenzi, Maritza Lienlaf-Moreno, Wendy M Kandell, Daniel N Santiago, Mibel Pabón-Saldaña, Lancia Darville, Bin Fang, Uwe Rix, Sean Yoder, Anders Berglund, John M Koomen, Eric B Haura, and Daniel Abate-Daga. An immunoproteomic approach to characterize the CAR interactome and signalosome. *Sci. Signal.*, 12(568):eaap9777, February 2019.

[236] Ling Wu, Joanna Brzostek, Previtha Dawn Sakthi Vale, Qianru Wei, Clara K T Koh, June Xu Hui Ong, Liang-Zhe Wu, Jia Chi Tan, Yen Leong Chua, Jiawei Yap, Yuan Song, Vivian Jia Yi Tan, Triscilla Y Y Tan, Junyun Lai, Paul A MacAry, and Nicholas R J Gascoigne. CD28-CAR-T cell activation through FYN kinase signaling rather than LCK enhances therapeutic performance. *Cell Rep. Med.*, 4(2):100917, February 2023.

[237] Kyle G Daniels, Shangying Wang, Milos S Simic, Hersh K Bhargava, Sara Capponi, Yurie Tonai, Wei Yu, Simone Bianco, and Wendell A Lim. Decoding CAR T cell phenotype using combinatorial signaling motif libraries and machine learning. *Science*, 378(6625):1194–1200, December 2022.

[238] Jake Burton, Jesús A Siller-Farfán, Johannes Pettmann, Benjamin Salzer, Mikhail Kutuzov, P Anton van der Merwe, and Omer Dushek. Inefficient exploitation of accessory receptors reduces the sensitivity of chimeric antigen receptors. *Proc. Natl. Acad. Sci. U. S. A.*, 120(2):e2216352120, January 2023.

[239] Michael L Dustin. The immunological synapse. *Cancer Immunol. Res.*, 2(11):1023–1033, November 2014.

[240] Michael L Dustin. Modular design of immunological synapses and kinapses. *Cold Spring Harb. Perspect. Biol.*, 1(1):a002873, July 2009.

[241] Giacomo Janson, Gilberto Valdes-Garcia, Lim Heo, and Michael Feig. Direct generation of protein conformational ensembles via machine learning. *Nat. Commun.*, 14(1):774, February 2023.

[242] Giacomo Janson and Michael Feig. Transferable deep generative modeling of intrinsically disordered protein conformations. *PLoS Comput. Biol.*, 20(5):e1012144, May 2024.

[243] Giulio Tesei, Thea K Schulze, Ramon Crehuet, and Kresten Lindorff-Larsen. Accurate model of liquid-liquid phase behavior of intrinsically disordered proteins from optimization of single-chain properties. *Proc. Natl. Acad. Sci. U. S. A.*, 118(44):e2111696118, November 2021.

[244] Sarah Lewis, Tim Hempel, José Jiménez-Luna, Michael Gastegger, Yu Xie, Andrew Y K Foong, Victor García Satorras, Osama Abdin, Bastiaan S Veeling, Iryna Zaporozhets, Yaoyi Chen, Soojung Yang, Adam E Foster, Arne Schneuing, Jigyasa Nigam, Federico Barbero, Vincent Stimper, Andrew Campbell, Jason Yim, Marten Lienen, Yu Shi, Shuxin Zheng, Hannes Schulz, Usman Munir, Roberto Sordillo, Ryota Tomioka, Cecilia Clementi, and Frank Noé. Scalable emulation of protein equilibrium ensembles with generative deep learning. *Science*, 389(6761):eadv9817, August 2025.

[245] Kaushik Choudhuri, David Wiseman, Marion H Brown, Keith Gould, and P Anton van der Merwe. T-cell receptor triggering is critically dependent on the dimensions of its peptide-MHC ligand. *Nature*, 436(7050):578–582, July 2005.

[246] Danielle M Paul, Judith Mantell, Ufuk Borucu, Jennifer Coombs, Katherine J Surridge, John M Squire, Paul Verkade, and Mark P Dodding. In situ cryo-electron tomography reveals filamentous actin within the microtubule lumen. *J. Cell Biol.*, 219(9):e201911154, September 2020.

[247] Terry J Fry, Nirali N Shah, Rimas J Orentas, Maryalice Stetler-Stevenson, Constance M Yuan, Sneha Ramakrishna, Pamela Wolters, Staci Martin, Cindy Delbrook, Bonnie Yates, Haneen Shalabi, Thomas J Fountaine, Jack F Shern, Robbie G Majzner, David F Stroncek, Marianna Sabatino, Yang Feng, Dimiter S Dimitrov, Ling Zhang, Sang Nguyen, Haiying Qin, Boro Dropulic, Daniel W Lee, and Crystal L Mackall. CD22-targeted CAR T cells induce remission in B-ALL that is naive or resistant to CD19-targeted CAR immunotherapy. *Nat. Med.*, 24(1):20–28, January 2018.

[248] A M Wu, G J Tan, M A Sherman, P Clarke, T Olafsen, S J Forman, and A A Raubitschek. Multimerization of a chimeric anti-CD20 single-chain Fv-Fc fusion protein is mediated through variable domain exchange. *Protein Eng.*, 14(12):1025–1033, December 2001.

[249] Nirali N Shah, Sarah K Tasian, M Eric Kohler, Emily M Hsieh, Susanne H C Baumeister, Corinne Summers, Haneen Shalabi, Jessica A Pollard, Bonnie Yates, Ruta Brazauskas, Eric Ndifon, Constance M Yuan, Hao-Wei Wang, Hati K Kobusingye, Kayla Christianson, Allison A Weiss, Erin Leckrone, Bronwen E Shaw, Jeffery J Auletta, Hemalatha G Rangarajan, John Roach, Julie Blake, Dennis Michiel, Brian Bowser, Sheryl Ruppel, George Mitra, Anthony Welch, Jason Yovandich, Rachelle Salomon, Leslie Kean, Katherine Tarlock, Michael A Pulsipher, Michael R Verneris, Larisa Broglie, Terry J Fry, Steven M Devine, and Richard Aplenc. CD33 CAR t-cells (CD33CART) for children and young adults with relapsed/refractory AML: Dose-escalation results from a phase I/II multicenter trial. *Blood*, 142(Supplement 1):771–771, November 2023.

[250] Terry J Fry, Nirali N Shah, Rimas J Orentas, Maryalice Stetler-Stevenson, Constance M Yuan, Sneha Ramakrishna, Pamela Wolters, Staci Martin, Cindy Delbrook, Bonnie Yates, Haneen Shalabi, Thomas J Fountaine, Jack F Shern, Robbie G Majzner, David F Stroncek, Marianna Sabatino, Yang Feng, Dimiter S Dimitrov, Ling Zhang, Sang Nguyen, Haiying Qin, Boro Dropulic, Daniel W Lee, and Crystal L Mackall. CD22-targeted CAR T cells induce remission in B-ALL that is naive or resistant to CD19-targeted CAR immunotherapy. *Nat. Med.*, 24(1):20–28, January 2018.

[251] Nirali N Shah, Steven L Highfill, Haneen Shalabi, Bonnie Yates, Jianjian Jin, Pamela L Wolters, Amanda Ombrello, Seth M Steinberg, Staci Martin, Cindy Delbrook, Leah Hoffman, Lauren Little, Anusha Ponduri, Haiying Qin, Haris Qureshi, Alina Dulau-Florea, Dalia Salem, Hao-Wei Wang, Constance Yuan, Maryalice Stetler-Stevenson, Sandhya Panch, Minh Tran, Crystal L Mackall, David F Stroncek, and Terry J Fry. CD4/CD8 t-cell selection affects chimeric antigen receptor (CAR) t-cell potency and toxicity: Updated results from a phase I anti-CD22 CAR t-cell trial. *J. Clin. Oncol.*, 38(17):1938–1950, June 2020.

[252] Nicholas J Haradhvala, Mark B Leick, Katie Maurer, Satyen H Gohil, Rebecca C Larson, Ning Yao, Kathleen M E Gallagher, Katelin Katsis, Matthew J Frigault, Jackson Southard, Shuqiang Li, Michael C Kann, Harrison Silva, Max Jan, Kahn Rhrissorrakrai, Filippo Utro, Chaya Levovitz, Raquel A Jacobs, Kara Slowik, Brian P Danysh, Kenneth J Livak, Laxmi Parida, Judith Ferry, Caron Jacobson, Catherine J Wu, Gad Getz, and Marcela V Maus. Distinct cellular dynamics associated with response to CAR-T therapy for refractory B cell lymphoma. *Nat. Med.*, 28(9):1848–1859, September 2022.

[253] D Moritz and B Groner. A spacer region between the single chain antibody- and the CD3 zeta-chain domain of chimeric T cell receptor components is required for efficient ligand binding and signaling activity. *Gene Ther*, 2(8):539–546, October 1995.

[254] Colin D Godwin, George S Laszlo, Salvatore Fiorenza, Eliotte E Garling, Tinh-Doan Phi, Olivia M Bates, Colin E Correnti, Benjamin G Hoffstrom, Margaret C Lunn, Olivier Humbert, Hans-Peter Kiem, Cameron J Turtle, and Roland B Walter. Targeting the membrane-proximal c2-set domain of CD33 for improved CD33-directed immunotherapy. *Leukemia*, 35(9):2496–2507, September 2021.

[255] Salvatore Fiorenza, Sheryl Y T Lim, George S Laszlo, Erik L Kimble, Tinh-Doan Phi, Margaret C Lunn-Halbert, Delaney R Kirchmeier, Jenny Huo, Hans-Peter Kiem, Cameron J Turtle, and Roland B Walter. Targeting the membrane-proximal c2-set domain of CD33 for improved CAR T cell therapy. *Mol. Ther. Oncol.*, 32(3):200854, September 2024.

[256] Scott McComb, Tina Nguyen, Alex Shepherd, Kevin A Henry, Darin Bloemberg, Anne Marcil, Susanne Maclean, Ahmed Zafer, Rénald Gilbert, Christine Gadoury, Robert A Pon, Traian Sulea, Qin Zhu, and Risini D Weeratna. Programmable attenuation of antigenic sensitivity for a nanobody-based EGFR chimeric antigen receptor through hinge domain truncation. *Front. Immunol.*, 13:864868, July 2022.

[257] Nan Li, Alex Quan, Dan Li, Jiajia Pan, Hua Ren, Gerard Hoeltzel, Natalia de Val, Dana Ashworth, Weiming Ni, Jing Zhou, Sean Mackay, Stephen M Hewitt, Raul Cachau, and Mitchell Ho. The IgG4 hinge with CD28 transmembrane domain improves VHH-based CAR T cells targeting a membrane-distal epitope of GPC1 in pancreatic cancer. *Nat. Commun.*, 14(1):1986, April 2023.

[258] David T Rodgers, Magdalena Mazagova, Eric N Hampton, Yu Cao, Nitya S Ramadoss, Ian R Hardy, Andrew Schulman, Juanjuan Du, Feng Wang, Oded Singer, Jennifer Ma, Vanessa Nunez, Jiayin Shen, Ashley K Woods, Timothy M Wright, Peter G Schultz, Chan Hyuk Kim, and Travis S Young. Switch-mediated activation and retargeting of CAR-T cells for b-cell malignancies. *Proc. Natl. Acad. Sci. U. S. A.*, 113(4):E459–68, January 2016.

[259] Leah Alabanza, Melissa Pegues, Claudia Geldres, Victoria Shi, Jed J W Wiltzius, Stuart A Sievers, Shicheng Yang, and James N Kochenderfer. Function of novel anti-CD19 chimeric antigen receptors with human variable regions is affected by hinge and transmembrane domains. *Mol. Ther.*, 25(11):2452–2465, November 2017.

[260] Michael Hudecek, Daniel Sommermeyer, Paula L Kosasih, Anne Silva-Benedict, Lingfeng Liu, Christoph Rader, Michael C Jensen, and Stanley R Riddell. The nonsignaling extracellular spacer domain of chimeric antigen receptors is decisive for in vivo antitumor activity. *Cancer Immunol. Res.*, 3(2):125–135, February 2015.

[261] Sylvain Simon and Stanley R Riddell. Dual targeting with CAR T cells to limit antigen escape in multiple myeloma. *Blood Cancer Discov.*, 1(2):130–133, September 2020.

[262] J. Copperman and M. G. Guenza. Predicting protein dynamics from structural ensembles. *The Journal of Chemical Physics*, 143(24):243131, 2015.

[263] Guillaume Bouvignies, Pramodh Vallurupalli, D. Flemming Hansen, Bruno E. Correia, Oliver Lange, Alaji Bah, Robert M. Vernon, Frederick W. Dahlquist, David Baker, and Lewis E. Kay. Solution structure of a minor and transiently formed state of a t4 lysozyme mutant. *Nature*, 477(7362):111–114, Sep 2011.

[264] Sandro Bottaro and Kresten Lindorff-Larsen. Biophysical experiments and biomolecular simulations: A perfect match? *Science*, 361(6400):355–360, 2018.

[265] Kresten Lindorff-Larsen, Stefano Piana, Ron O. Dror, and David E. Shaw. How fast-folding proteins fold. *Science*, 334(6055):517–520, 2011.

[266] Valentina Tozzini, Joanna Trylska, Chia en Chang, and J. Andrew McCammon. Flap opening dynamics in hiv-1 protease explored with a coarse-grained model. *Journal of Structural Biology*, 157(3):606–615, 2007. Advances in Molecular Dynamics Simulations.

[267] Paulo C. T. Souza, Luís Borges-Araújo, Christopher Brasnett, Rodrigo A. Moreira, Fabian Grünewald, Peter Park, Liguo Wang, Hafez Razmazma, Ana C. Borges-Araújo, Luis Fernando Cofas-Vargas, Luca Monticelli, Raúl Mera-Adasme, Manuel N. Melo, Sangwook Wu, Siewert J. Marrink, Adolfo B. Poma, and Sebastian Thallmair. Gōmartini 3: From large conformational changes in proteins to environmental bias corrections. *Nature Communications*, 16(1):4051, Apr 2025.

[268] G Marius Clore, Attila Szabo, Ad Bax, Lewis E Kay, Paul C Driscoll, and Angela M Gronenborn. Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins. *Journal of the American Chemical Society*, 112(12):4989–4991, 1990.

[269] G. Marius Clore, Paul C. Driscoll, Paul T. Wingfield, and Angela M. Gronenborn. Analysis of the backbone dynamics of interleukin-1.beta. using two-dimensional inverse detected heteronuclear nitrogen-15-proton nmr spectroscopy. *Biochemistry*, 29(32):7387–7401, Aug 1990.

[270] Paul Schanda and Gilad Haran. Nmr and single-molecule fret insights into fast protein motions and their relation to function. *Annual Review of Biophysics*, 53(53):247–273, 2024.

[271] David E. Shaw, Peter J. Adams, Asaph Azaria, Joseph A. Bank, Brannon Batson, Alistair Bell, Michael Bergdorf, Jhanvi Bhatt, J. Adam Butts, Timothy Correia, Robert M. Dirks, Ron O. Dror, Michael P. Eastwood, Bruce Edwards, Amos Even, Peter Feldmann, Michael Fenn, Christopher H. Fenton, Anthony Forte, Joseph Gagliardo, Gennette Gill, Maria Gorlatova, Brian Greskamp, J.P. Grossman, Justin Gullingsrud, Anissa Harper, William Hasenplaugh, Mark Heily, Benjamin Colin Heshmat, Jeremy Hunt, Douglas J. Ierardi, Lev Iserovich, Bryan L. Jackson, Nick P. Johnson, Mollie M. Kirk, John L. Klepeis, Jeffrey S. Kuskin, Kenneth M. Mackenzie, Roy J. Mader, Richard McGowen, Adam McLaughlin, Mark A. Moraes, Mohamed H. Nasr, Lawrence J. Nociolo, Lief O'Donnell, Andrew Parker, Jon L. Peticolas, Goran Pocina, Cristian Predescu, Terry Quan, John K. Salmon, Carl Schwink, Keun Sup Shim, Naseer Siddique, Jochen Spengler, Tamas Szalay, Raymond Tabladillo, Reinhard Tartler, Andrew G. Taube, Michael Theobald, Brian Towles, William Vick, Stanley C. Wang, Michael Wazlowski, Madeleine J. Weingarten, John M. Williams, and Kevin A. Yuh. Anton 3: twenty microseconds of molecular dynamics simulation before lunch. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, SC '21, New York, NY, USA, 2021. Association for Computing Machinery.

[272] Shuxin Zheng, Jiyan He, Chang Liu, Yu Shi, Ziheng Lu, Weitao Feng, Fusong Ju, Jiaxi Wang, Jianwei Zhu, Yaosen Min, He Zhang, Shidi Tang, Hongxia Hao, Peiran Jin, Chi Chen, Frank Noé, Haiguang Liu, and Tie-Yan Liu. Predicting equilibrium distributions for molecular systems with deep learning. *Nature Machine Intelligence*, 6(5):558–567, May 2024.

[273] Esther Caballero-Manrique, Jenelle K Bray, William A Deutschman, Frederick W Dahlquist, and Marina G Guenza. A theory of protein dynamics to predict nmr relaxation. *Biophysical Journal*, 93(12):4128–4140, 2007.

[274] E. R. Beyerle and M. G. Guenza. Comparison between slow anisotropic le4pd fluctuations and the principal component analysis modes of ubiquitin. *The Journal of Chemical Physics*, 154(12):124111, 03 2021.

[275] Eric R. Beyerle and Marina G. Guenza. Kinetics analysis of ubiquitin local fluctuations with markov state modeling of the le4pd normal modes. *The Journal of Chemical Physics*, 151(16):164119, 10 2019.

[276] Ronald R Coifman, Stephane Lafon, Ann B Lee, Mauro Maggioni, Boaz Nadler, Frederick Warner, and Steven W Zucker. Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. *Proceedings of the national academy of sciences*, 102(21):7426–7431, 2005.

[277] Ronald R Coifman, Ioannis G Kevrekidis, Stéphane Lafon, Mauro Maggioni, and Boaz Nadler. Diffusion maps, reduction coordinates, and low dimensional representation of stochastic systems. *Multiscale Modeling & Simulation*, 7(2):842–864, 2008.

[278] Tuğçe Gökdemir and Jakub Rydzewski. Machine learning of slow collective variables and enhanced sampling via spatial techniques. *Chemical Physics Reviews*, 6(1):011304, 02 2025.

[279] Jakub Rydzewski, Ming Chen, and Omar Valsson. Manifold learning in atomistic simulations: a conceptual review. *Machine Learning: Science and Technology*, 4(3):031001, sep 2023.

[280] Boaz Nadler, Stephane Lafon, Ioannis Kevrekidis, and Ronald Coifman. Diffusion maps, spectral clustering and eigenfunctions of fokker-planck operators. *Advances in neural information processing systems*, 18, 2005.

[281] Frank Noé and Cecilia Clementi. Kinetic distance and kinetic maps from molecular dynamics simulation. *Journal of Chemical Theory and Computation*, 11(10):5002–5011, Oct 2015.

[282] Mary A Rohrdanz, Wenwei Zheng, Mauro Maggioni, and Cecilia Clementi. Determination of reaction coordinates via locally scaled diffusion map. *The Journal of chemical physics*, 134(12), 2011.

[283] Jakub Rydzewski. Spectral map: Embedding slow kinetics in collective variables. *The Journal of Physical Chemistry Letters*, 14(22):5216–5220, 2023.

[284] Rieko Ishima, Rodolfo Ghirlando, József Tözsér, Angela M. Gronenborn, Dennis A. Torchia, and John M. Louis. Folded monomer of hiv-1 protease. *Journal of Biological Chemistry*, 276(52):49110–49116, 2001.

[285] Heike I. Rösner, Martina Caldarini, Gregory Potel, Daniel Malmodin, Maria A. Vanoni, Alessandro Aliverti, Ricardo A. Broglia, Birthe B. Kragelund, and Guido Tiana. The denatured state of hiv-1 protease under native conditions. *Proteins: Structure, Function, and Bioinformatics*, 90(1):96–109, 2022.

[286] Sébastien Bubeck, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Peter Lee, Yin Tat Lee, Yuanzhi Li, Scott Lundberg, et al. Sparks of artificial general intelligence: Early experiments with gpt-4. *arXiv preprint arXiv:2303.12712*, 2023.

[287] Lei Huang, Weijiang Yu, Weitao Ma, Weihong Zhong, Zhangyin Feng, Haotian Wang, Qianglong Chen, Weihua Peng, Xiaocheng Feng, Bing Qin, and Ting Liu. A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. *ACM Trans. Inf. Syst.*, 43(2), January 2025.

[288] Sarentha Chetty, Soumendranath Bhakat, Alberto J.M. Martin, and Mahmoud E.S. Soliman. Multi-drug resistance profile of pr20 hiv-1 protease is attributed to distorted conformational and drug binding landscape: molecular dynamics insights. *Journal of Biomolecular Structure and Dynamics*, 34(1):135–151, 2016. PMID: 25671669.

[289] Michael Souffrant, Xin-Qiu Yao, and Donald Hamelberg. Evolving mutational buildup in hiv-1 protease shifts conformational dynamics to gain drug resistance. *Journal of Chemical Information and Modeling*, 63(12):3892–3902, Jun 2023.

[290] Xi Huang, Manuel D. Britto, Jamie L. Kear-Scott, Christopher D. Boone, James R. Rocca, Carlos Simmerling, Robert Mckenna, Michael Bieri, Paul R. Gooley, Ben M. Dunn, and Gail E. Fanucci. The role of select subtype polymorphisms on hiv-1 protease conformational sampling and dynamics*. *Journal of Biological Chemistry*, 289(24):17203–17214, 2014.

[291] Rieko Ishima, Darón I Freedberg, Yun-Xing Wang, John M Louis, and Dennis A Torchia. Flap opening and dimer-interface flexibility in the free and inhibitor-bound hiv protease, and their implications for function. *Structure*, 7(9):1047–S12, 1999.

[292] Milan Kožíšek, Martin Lepšík, Klára Grantz Šašková, Jiří Brynda, Jan Konvalinka, and Pavlína Řezáčová. Thermodynamic and structural analysis of hiv protease resistance to darunavir – analysis of heavily mutated patient-derived hiv-1 proteases. *The FEBS Journal*, 281(7):1834–1847, 2014.

[293] Holly Heaslet, Robin Rosenfeld, Mike Giffin, Ying-Chuan Lin, Karen Tam, Bruce E. Torbett, John H. Elder, Duncan E. McRee, and C. David Stout. Conformational flexibility in the flap domains of ligand-free HIV protease. *Acta Crystallographica Section D*, 63(8):866–875, Aug 2007.

[294] Chen-Hsiang Shen, Yu-Chung Chang, Johnson Agniswamy, Robert W. Harrison, and Irene T. Weber. Conformational variation of an extreme drug resistant mutant of hiv protease. *Journal of Molecular Graphics and Modelling*, 62:87–96, 2015.

[295] Daniel W. Kneller, Johnson Agniswamy, Robert W. Harrison, and Irene T. Weber. Highly drug-resistant hiv-1 protease reveals decreased intra-subunit interactions due to clusters of mutations. *The FEBS Journal*, 287(15):3235–3254, 2020.

[296] Alexander L Perryman, Jung-Hsin Lin, and J Andrew McCammon. Hiv-1 protease molecular dynamics of a wild-type and of the v82f/i84v mutant: Possible contributions to drug resistance and a potential new target site for drugs. *Protein Science*, 13(4):1108–1123, 2004.

[297] Yuqi Yu, Jinan Wang, Zhaoqiang Chen, Guimin Wang, Qiang Shao, Jiye Shi, and Weiliang Zhu. Structural insights into hiv-1 protease flap opening processes and key intermediates. *RSC Adv.*, 7:45121–45128, 2017.

[298] Kannan Sankar, Jie Liu, Yuan Wang, and Robert L. Jernigan. Distributions of experimental protein structures on coarse-grained free energy landscapes. *The Journal of Chemical Physics*, 143(24):243153, 12 2015.

[299] Da Lu, Shuhong Yu, Yixiang Huang, and Xinqi Gong. Multimeric protein interaction and complex prediction: Structure, dynamics and function. *Computational and Structural Biotechnology Journal*, 27:1975–1997, 2025.

[300] S. Spinelli, Q.Z. Liu, P.M. Alzari, P.H. Hirel, and R.J. Poljak. The three-dimensional structure of the aspartyl protease from the hiv-1 isolate bru. *Biochimie*, 73(11):1391–1396, 1991.

[301] Johnson Agniswamy, Chen-Hsiang Shen, Annie Aniana, Jane M. Sayer, John M. Louis, and Irene T. Weber. Hiv-1 protease with 20 mutations exhibits extreme resistance to clinical inhibitors through coordinated structural rearrangements. *Biochemistry*, 51(13):2819–2828, 2012. PMID: 22404139.

[302] I. Lyubimov and M. G. Guenza. First-principle approach to rescale the dynamics of simulated coarse-grained macromolecular liquids. *Phys. Rev. E*, 84:031801, Sep 2011.

[303] R Zwanzig. Diffusion in a rough potential. *Proceedings of the National Academy of Sciences*, 85(7):2029–2030, 1988.

[304] J. Copperman, M. Dinpajooh, E. R. Beyerle, and M. G. Guenza. Universality and specificity in protein fluctuation dynamics. *Phys. Rev. Lett.*, 119:158101, Oct 2017.

[305] S. Siegel. *Nonparametric Statistics for the Behavioral Sciences*. International student edition. McGraw-Hill, 1956.

[306] Dhiman Ray and Michele Parrinello. Data-driven classification of ligand unbinding pathways. *Proceedings of the National Academy of Sciences*, 121(10):e2313542121, 2024.

[307] Wannes Meert, Kilian Hendrickx, Toon Van Craenendonck, Pieter Robberechts, Hendrik Blockeel, and Jesse Davis. Dtaidistance, August 2020.

[308] Elana Simon and James Zou. Interplm: discovering interpretable features in protein language models via sparse autoencoders. *Nature Methods*, 22(10):2107–2117, September 2025.

[309] Zhidian Zhang, Hannah K. Wayment-Steele, Garyk Brixi, Haobo Wang, Dorothee Kern, and Sergey Ovchinnikov. Protein language models learn evolutionary statistics of interacting sequence motifs. *Proceedings of the National Academy of Sciences*, 121(45), October 2024.

[310] Da Teng, Yunrui Qiu, Gokulakannan Sakthivel, Akashnathan Aranganathan, Lukas Herron, and Pratyush Tiwary. Detecting and quantifying overparametrization in rna language models with redial. May 2026.

[311] Heqin Zhu, Ruifeng Li, Ao Chang, Haobin Chen, Feng Zhang, Fenghe Tang, Tong Ye, Xin Li, Yunjie Gu, Peng Xiong, and S. Kevin Zhou. A fully open structure-guided rna foundation model for robust structural and functional inference. August 2025.

[312] Alisia Fadini, Minhuan Li, Airlie J. McCoy, Suresh Banjara, Hiroki Okumura, Eve Napier, Pietro Fontana, Amir R. Khan, Luca Jovine, Thomas C. Terwilliger, Randy J. Read, Doeke R. Hekstra, and Mohammed AlQuraishi. Alphafold as a prior: experimental structure determination conditioned on a pretrained neural network. *Nature Methods*, 23(4):785–795, April 2026.

[313] Kolja Stahl, Andrea Graziadei, Therese Dau, Oliver Brock, and Juri Rappsilber. Protein structure prediction with in-cell photo-crosslinking mass spectrometry and deep learning. *Nature Biotechnology*, 41(12):1810–1819, March 2023.

[314] Tianqi Wu, Richard A. Stein, Te-Yu Kao, Benjamin Brown, and Hassane S. Mchaourab. Modeling protein conformational ensembles by guiding alphafold2 with double electron electron resonance (deer) distance distributions. *Nature Communications*, 16(1), August 2025.

[315] Akashnathan Aranganathan and Eric R. Beyerle. Applied causality to infer protein dynamics and kinetics. *Journal of Chemical Information and Modeling*, 66(3):1661–1674, January 2026.

[316] Simon Olsson and Frank Noé. Mechanistic models of chemical exchange induced relaxation in protein nmr. *Journal of the American Chemical Society*, 139(1):200–210, December 2016.

[317] Simon Olsson, Hao Wu, Fabian Paul, Cecilia Clementi, and Frank Noé. Combining experimental and simulation data of molecular processes via augmented markov models. *Proceedings of the National Academy of Sciences*, 114(31):8265–8270, July 2017.

[318] Davide Branduardi, Francesco Luigi Gervasio, and Michele Parrinello. From a to b in free energy space. *The Journal of Chemical Physics*, 126(5), February 2007.

[319] Arthur France-Lanord, Hadrien Vroylandt, Mathieu Salanne, Benjamin Rotenberg, A. Marco Saitta, and Fabio Pietrucci. Data-driven path collective variables. *Journal of Chemical Theory and Computation*, 20(8):3069–3084, April 2024.

[320] Tim Hempel, Mauricio J. del Razo, Christopher T. Lee, Bryn C. Taylor, Rommie E. Amaro, and Frank Noé. Independent markov decomposition: Toward modeling kinetics of biomolecular complexes. *Proceedings of the National Academy of Sciences*, 118(31), July 2021.

[321] Andreas Mardt, Tim Hempel, Cecilia Clementi, and Frank Noé. Deep learning to decompose macromolecules into independent markovian domains. *Nature Communications*, 13(1), November 2022.

[322] Tim Hempel, Simon Olsson, and Frank Noé. Markov field models: Scaling molecular kinetics approaches to large molecular machines. *Current Opinion in Structural Biology*, 77:102458, December 2022.

[323] Frank Noé, Simon Olsson, Jonas Köhler, and Hao Wu. Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning. *Science*, 365(6457), September 2019.

[324] Lukas Herron, Kinjal Mondal, John S. Schneekloth, and Pratyush Tiwary. Inferring phase transitions and critical exponents from limited observations with thermodynamic maps. *Proceedings of the National Academy of Sciences*, 121(52), December 2024.

[325] Suemin Lee, Ruiyu Wang, Lukas Herron, and Pratyush Tiwary. Exponentially tilted thermodynamic maps (exptm): Predicting phase transitions across temperature, pressure, and chemical potential, 2025.

[326] Yunrui Qiu, Richard John, Lukas Herron, and Pratyush Tiwary. Latent thermodynamic flows: unified representation learning and generative modeling of temperature-dependent behaviors from limited data. *Chemical Science*, 17(7):3743–3757, 2026.

[327] Pratyush Tiwary and Michele Parrinello. From metadynamics to dynamics. *Physical Review Letters*, 111(23), December 2013.

[328] Pratyush Tiwary and Axel van de Walle. *A Review of Enhanced Sampling Approaches for Accelerated Molecular Dynamics*, page 195–221. Springer International Publishing, 2016.

[329] Giovanni Bussi and Alessandro Laio. Using metadynamics to explore complex free-energy landscapes. *Nature Reviews Physics*, 2(4):200–212, March 2020.

[330] Yong Duan, Chun Wu, Shibasish Chowdhury, Mathew C. Lee, Guoming Xiong, Wei Zhang, Rong Yang, Piotr Cieplak, Ray Luo, Taisung Lee, James Caldwell, Junmei Wang, and Peter Kollman. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *Journal of Computational Chemistry*, 24(16):1999–2012, 2003.

[331] Mark James Abraham, Teemu Murtola, Roland Schulz, Szilárd Páll, Jeremy C. Smith, Berk Hess, and Erik Lindahl. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1-2:19–25, 2015.

[332] Berk Hess, Henk Bekker, Herman J. C. Berendsen, and Johannes G. E. M. Fraaije. Lincs: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, 18(12):1463–1472, 1997.

[333] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh ewald: An n log(n) method for ewald sums in large systems. *The Journal of Chemical Physics*, 98(12):10089–10092, June 1993.

[334] Massimiliano Bonomi, Giovanni Bussi, Carlo Camilloni, Gareth A. Tribello, Pavel Banáš, Alessandro Barducci, Mattia Bernetti, Peter G. Bolhuis, Sandro Bottaro, Davide Branduardi, Riccardo Capelli, and Paolo Carloni. Promoting transparency and reproducibility in enhanced molecular simulations. *Nature Methods*, 16(8):670–673, aug 2019.

[335] Li-Nan Chen, Wei-Wei Wang, Ying-Jun Dong, Dan-Dan Shen, Jia Guo, Xuefei Yu, Jiao Qin, Su-Yu Ji, Huibing Zhang, Qingya Shen, Qiaojun He, Bo Yang, Yan Zhang, Qinglin Li, and Chunyou Mao. Structures of the endogenous peptide- and selective non-peptide agonist-bound sstr2 signaling complexes. *Cell Research*, 32(8):785–788, May 2022.

[336] Jan-Hendrik Prinz, Hao Wu, Marco Sarich, Bettina Keller, Martin Senne, Martin Held, John D. Chodera, Christof Schütte, and Frank Noé. Markov models of molecular kinetics: Generation and validation. *The Journal of Chemical Physics*, 134(17), May 2011.

[337] Teixeira AL and Alves NA. The high stability of the three-helix bundle uba domain of p62 protein as revealed by molecular dynamics simulations. *Journal of molecular modeling*, 27(4):102, Mar 2021.

[338] Qinghua Wang, Amanda M. Goh, Peter M. Howley, and Kylie J. Walters. Ubiquitin recognition by the dna repair protein hhr23a. *Biochemistry*, 42(46):13529–13535, October 2003.

[339] Wolfgang Kabsch and Christian Sander. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22(12):2577–2637, December 1983.

[340] Benedict Leimkuhler and Charles Matthews. Robust and efficient configurational molecular sampling via langevin dynamics. *J. Chem. Phys.*, 138(17):174102, May 2013.

[341] Peter Eastman, Jason Swails, John D. Chodera, Robert T. McGibbon, Yutong Zhao, Kyle A. Beauchamp, Lee-Ping Wang, Andrew C. Simmonett, Matthew P. Harrigan, Chaya D. Stern, Rafal P. Wiewiora, Bernard R. Brooks, and Vijay S. Pande. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLOS Computational Biology*, 13(7):e1005659, July 2017.

[342] Mrinal Shekhar, Zachary Smith, Markus A. Seeliger, and Pratyush Tiwary. Protein flexibility and dissociation pathway differentiation can explain onset of resistance mutations in kinases. *Angew. Chem. Int. Ed.*, 61(28), May 2022.

[343] David A Case, Thomas E Cheatham III, Tom Darden, Holger Gohlke, Ray Luo, Kenneth M Merz Jr, Alexey Onufriev, Carlos Simmerling, Bing Wang, and Robert J Woods. The amber biomolecular simulation programs. *Journal of computational chemistry*, 26(16):1668–1688, 2005.

[344] Kresten Lindorff-Larsen, Stefano Piana, Kim Palmo, Paul Maragakis, John L Klepeis, Ron O Dror, and David E Shaw. Improved side-chain torsion potentials for the amber ff99sb protein force field. *Proteins: Structure, Function, and Bioinformatics*, 78(8):1950–1958, 2010.

[345] William L Jorgensen, Jayaraman Chandrasekhar, Jeffry D Madura, Roger W Impey, and Michael L Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*, 79(2):926–935, 1983.

[346] Zhijun Zhang, Xinzijian Liu, Kangyu Yan, Mark E Tuckerman, and Jian Liu. Unified efficient thermostat scheme for the canonical ensemble with holonomic or isokinetic constraints via molecular dynamics. *The Journal of Physical Chemistry A*, 123(28):6056–6079, 2019.

[347] Pengfei Li and Kenneth M Merz Jr. Taking into account the ion-induced dipole interaction in the nonbonded model of ions. *Journal of chemical theory and computation*, 10(1):289–297, 2014.

[348] Bernard R Brooks, Charles L Brooks III, Alexander D Mackerell Jr, Lennart Nilsson, Robert J Petrella, Benoît Roux, Youngdo Won, Georgios Archontis, Christian Bartels, Stefan Boresch, et al. Charmm: the biomolecular simulation program. *Journal of computational chemistry*, 30(10):1545–1614, 2009.

[349] Jing Huang, Sarah Rauscher, Grzegorz Nawrocki, Ting Ran, Michael Feig, Bert L De Groot, Helmut Grubmüller, and Alexander D MacKerell Jr. Charmm36m: an improved force field for folded and intrinsically disordered proteins. *Nature methods*, 14(1):71–73, 2017.

[350] Sunhwan Jo, Taehoon Kim, Vidyashankara G Iyer, and Wonpil Im. Charmm-gui: a web-based graphical user interface for charmm. *Journal of computational chemistry*, 29(11):1859–1865, 2008.

[351] Jumin Lee, Xi Cheng, Sunhwan Jo, Alexander D MacKerell, Jeffery B Klauda, and Wonpil Im. Charmm-gui input generator for namd, gromacs, amber, openmm, and charmm/openmm simulations using the charmm36 additive force field. *Biophysical journal*, 110(3):641a, 2016.

[352] Aram Davtyan, Nicholas P Schafer, Weihua Zheng, Cecilia Clementi, Peter G Wolynes, and Garegin A Papoian. Awsem-md: protein structure prediction using coarse-grained physical potentials and bioinformatically based local structure biasing. *The Journal of Physical Chemistry B*, 116(29):8494–8503, 2012.

[353] Thomas H Charpentier, Laura E Thompson, Melissa A Liriano, Kristen M Varney, Paul T Wilder, Edwin Pozharski, Eric A Toth, and David J Weber. The effects of capz peptide (trtk-12) binding to s100b–ca2+ as examined by nmr and x-ray crystallography. *Journal of molecular biology*, 396(5):1227–1243, 2010.

[354] Gergő Gógl, Anita Alexa, Bence Kiss, Gergely Katona, Mihály Kovács, Andrea Bodor, Attila Reményi, and László Nyitray. Structural basis of ribosomal s6 kinase 1 (rsk1) inhibition by s100b protein: Modulation of the extracellular signal-regulated kinase (erk) signaling cascade in a calcium-dependent way. *Journal of Biological Chemistry*, 291(1):11–27, 2016.

[355] Paul T Wilder, Thomas H Charpentier, Melissa A Liriano, Kira Gianni, Kristen M Varney, Edwin Pozharski, Andrew Coop, Eric A Toth, Alex D MacKerell, and David J Weber. In vitro screening and structural characterization of inhibitors of the s100b-p53 interaction. *International journal of high throughput screening*, pages 109–126, 2010.

[356] Laura E. McKnight, E. Prabhu Raman, Padmavani Bezawada, Sucheta Kudrimoti, Paul T. Wilder, Kira G. Hartman, Raquel Godoy-Ruiz, Eric A. Toth, Andrew Coop, Alexander D. Jr. MacKerell, and David J. Weber. Structure-based discovery of a novel pentamidine-related inhibitor of the calcium-binding protein s100b. *ACS Medicinal Chemistry Letters*, 3(12):975–979, 2012. PMID: 23264854.

[357] Michael C. Cavalier, Zephan Melville, Ehson Aligholizadeh, E. Prabhu Raman, Wenbo Yu, Lei Fang, Milad Alasady, Adam D. Pierce, Paul T. Wilder, Alexander D. MacKerell, Jr, and David J. Weber. Novel protein–inhibitor interactions in site 3 of Ca^2+-bound S100B as discovered by X-ray crystallography. *Acta Crystallographica Section D*, 72(6):753–760, Jun 2016.

[358] Thomas H Charpentier, Paul T Wilder, Melissa A Liriano, Kristen M Varney, Edwin Pozharski, Alexander D MacKerell Jr, Andrew Coop, Eric A Toth, and David J Weber. Divalent metal ion complexes of s100b in the absence and presence of pentamidine. *Journal of molecular biology*, 382(1):56–73, 2008.

[359] Mariangela Agamennone, Lucia Cesari, Daniela Lalli, Elisa Turlizzi, Rebecca Del Conte, Paola Turano, Stefano Mangani, and Alessandro Padova. Fragmenting the s100b–p53 interaction: combined virtual/biophysical screening approaches to identify ligands. *ChemMedChem*, 5(3):428–435, 2010.

[360] Thomas H Charpentier, Paul T Wilder, Melissa A Liriano, Kristen M Varney, Shijun Zhong, Andrew Coop, Edwin Pozharski, Alexander D MacKerell Jr, Eric A Toth, and David J Weber. Small molecules bound to unique sites in the target protein binding cleft of calcium-bound s100b as characterized by nuclear magnetic resonance and x-ray crystallography. *Biochemistry*, 48(26):6202–6212, 2009.

[361] Jaime L Jensen, Venkata SK Indurthi, David B Neau, Stefan W Vetter, and Christopher L Colbert. Structural insights into the binding of the human receptor for advanced glycation end products (rage) by s100b, as revealed by an s100b–rage-derived peptide complex. *Biological Crystallography*, 71(5):1176–1183, 2015.

[362] Nathan T Wright, Benjamin L Prosser, Kristen M Varney, Danna B Zimmer, Martin F Schneider, and David J Weber. S100a1 and calmodulin compete for the same binding site on ryanodine receptor. *Journal of Biological Chemistry*, 283(39):26676–26683, 2008.

[363] Maestro. Schrödinger release 2023-3: Glide; induced fit docking protocol; prime. *Schrödinger, LLC: New York, NY*, 2023.

[364] Enade P Istyastono, Muhammad Radifar, Nunung Yuniarti, Vivitri D Prasasty, and Sudi Mungkasi. Pyplif hippos: A molecular interaction fingerprinting tool for docking results of autodock vina and plants. *Journal of Chemical Information and Modeling*, 60(8):3697–3702, 2020.

[365] Scott R. Brozell, Sudipto Mukherjee, Trent E. Balius, Daniel R. Roe, David A. Case, and Robert C. Rizzo. Evaluation of dock 6 as a pose generation and database enrichment tool. *Journal of Computer-Aided Molecular Design*, 26(6):749–773, 2012.

[366] P. Jaccard. *Bull. Soc. Vaudoise Sci. Nat.*, 37:547– 589, 1901.

[367] Greg Landrum. Rdkit: Open-source cheminformatics software. 2016.

[368] Jean-François Truchon and Christopher I. Bayly. Evaluating virtual screening methods: Good and bad metrics for the "early recognition" problem. *Journal of Chemical Information and Modeling*, 47(2):488–508, 2007.

[369] Robert P. Sheridan, Suresh B. Singh, Eugene M. Fluder, and Simon K. Kearsley. Protocols for bridging the peptide to nonpeptide gap in topological similarity searches. *Journal of Chemical Information and Computer Sciences*, 41(5):1395–1406, 2001.

[370] Wei Zhao, Kirk E. Hevener, Stephen W. White, Richard E. Lee, and James M. Boyett. A statistical framework to evaluate virtual screening. *BMC Bioinformatics*, 10, 2009.

[371] Benedict Leimkuhler and Charles Matthews. Efficient molecular dynamics using geodesic integration and solvent-solute splitting. *Proc. Math. Phys. Eng. Sci.*, 472(2189):20160138, May 2016.

[372] Gareth A Tribello, Massimiliano Bonomi, Davide Branduardi, Carlo Camilloni, and Giovanni Bussi. PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.*, 185(2):604–613, February 2014.

[373] David Regan, Joseph Williams, Paola Borri, and Wolfgang Langbein. Lipid bilayer thickness measured by quantitative DIC reveals phase transitions and effects of substrate hydrophilicity. *Langmuir*, 35(43):13805–13814, October 2019.

[374] João Marcelo Lamim Ribeiro, Pablo Bravo, Yihang Wang, and Pratyush Tiwary. Reweighted autoencoded variational bayes for enhanced sampling (RAVE). *J. Chem. Phys.*, 149(7):072301, August 2018.

[375] Arnd Hoeveler and Bernard Malissen. The cysteine residues in the cytoplasmic tail of cd8 α are required for its coreceptor function. *Molecular Immunology*, 30(8):755–764, June 1993.

[376] Emilia L Wu, Xi Cheng, Sunhwan Jo, Huan Rui, Kevin C Song, Eder M Dávila-Contreras, Yifei Qi, Jumin Lee, Viviana Monje-Galvan, Richard M Venable, Jeffery B Klauda, and Wonpil Im. CHARMM-GUI membrane builder toward realistic biological membrane simulations. *J. Comput. Chem.*, 35(27):1997–2004, October 2014.

[377] Lunna Li, Tommaso Casalini, Paolo Arosio, and Matteo Salvalaglio. Modeling the structure and interactions of intrinsically disordered peptides with multiple replica, metadynamics-based sampling methods and force-field combinations. *Journal of Chemical Theory and Computation*, 18(3):1915–1928, February 2022.

[378] Andrei L. Lomize, Spencer C. Todd, and Irina D. Pogozheva. Spatial arrangement of proteins in planar and curved membranes by ppm 3.0. *Protein Science*, 31(1):209–220, November 2021.

[379] Rieko Ishima, Dennis A. Torchia, Shannon M. Lynch, Angela M. Gronenborn, and John M. Louis. Solution structure of the mature hiv-1 protease monomer: Insight into the tertiary fold and stability of a precursor. *Journal of Biological Chemistry*, 278(44):43311–43319, Oct 2003.

[380] Hans O. Andersson, Kerstin Fridborg, Seved Löwgren, Mathias Alterman, Anna Mühlman, Magnus Björsne, Neeraj Garg, Ingmar Kvarnström, Wesley Schaal, Björn Classon, Anders Karlén, U. Helena Danielsson, Göran Ahlsén, Ullrika Nillroth, Lotta Vrang, Bo öberg, Bertil Samuelsson, Anders Hallberg, and Torsten Unge. Optimization of p1–p3 groups in symmetric and asymmetric hiv-1 protease inhibitors. *European Journal of Biochemistry*, 270(8):1746–1758, 2003.

[381] Stefano Piana, Paolo Carloni, and Ursula Rothlisberger. Drug resistance in hiv-1 protease: flexibility-assisted mechanism of compensatory mutations. *Protein Science*, 11(10):2393–2402, 2002.

[382] Luis Galiano, Fangyu Ding, Angelo M Veloro, Mandy E Blackburn, Carlos Simmerling, and Gail E Fanucci. Drug pressure selected mutations in hiv-1 protease alter flap conformations. *Journal of the American Chemical Society*, 131(2):430–431, 2009.

[383] Rajeswari Appadurai and Sanjib Senapati. Dynamical network of hiv-1 protease mutants reveals the mechanism of drug resistance and unhindered activity. *Biochemistry*, 55(10):1529–1540, Mar 2016.

[384] Ying Liu, Eran Eyal, and Ivet Bahar. Analysis of correlated mutations in hiv-1 protease using spectral clustering. *Bioinformatics*, 24(10):1243–1250, 03 2008.

[385] Shuichi Nosé. A unified formulation of the constant temperature molecular dynamics methods. *The Journal of chemical physics*, 81(1):511–519, 1984.

[386] William G Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Physical review A*, 31(3):1695, 1985.

[387] M. Parrinello and A. Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics*, 52(12):7182–7190, 12 1981.

[388] Lluís A. Belanche-Muñoz and Małgorzata Wiejacha. Analysis of kernel matrices via the von neumann entropy and its relation to rvm performances. *Entropy*, 25(1), 2023.