
Can Biomolecular Structure Predictors Generalize to Alternative Conformations?

Karson M. Chrispens

University of California, San Francisco
karson.chispens@ucsf.edu

Stephanie A. Wankowicz

Vanderbilt University
stephanie@wankowiczlab.com

Henry van den Bedem

Expedition Medicines, Cambridge, MA
University of California, San Francisco
henry.vandenbedem@ucsf.edu

James S. Fraser

University of California, San Francisco
james.fraser@ucsf.edu

Abstract

Proteins explore the conformational space around their native state to accomplish their function, so modeling the ensemble of accessible conformations is essential for understanding their behavior. Current structural models derived from macromolecular X-ray crystallography and single-particle cryogenic electron microscopy deposited in the Protein Data Bank are largely deposited as single states that fail to recognize the ensemble nature of proteins in experiments. Structure prediction models are trained on these single state structures, discarding cases where alternative conformations were manually built by an experimental modeler. Here, we apply steering of the Boltz-2 structure prediction model with synthetic density maps from manually modeled alternative conformations to explore the limitations of existing structure predictors. We find that sampling of alternative conformations is biased towards the single state structures in the training set despite the physical plausibility of the modeled alternative conformations, indicating limited ability to generalize in conformational space. To address this, we identify opportunities to overcome these limitations through augmenting training with alternative conformations and evaluating with experimental data.

1 Introduction

Protein structure prediction methods have advanced beyond monomeric protein predictions and now achieve high accuracy in predicting interactions between biomolecules. However, accurately modeling the details of these interactions and their functional impact requires a conformational ensemble. Modeling ensembles of accessible conformations has helped us understand the target specificity of drugs [Agafonov et al., 2014], reveal the allosteric coupling of activation in certain enzymes [Keedy et al., 2018], and target specific pockets with drug-like molecules during crystallographic fragment screening [Gahbauer et al., 2023, Pearce et al., 2017]. Existing computational methods for modeling conformational changes rely heavily on molecular dynamics simulations, but such simulations are often prohibitively expensive for conformational changes that occur only on longer timescales.

Experimental data in macromolecular crystallography and cryo-electron microscopy provide ensemble-averaged density maps that can reveal conformations important to function [Fraser et al., 2009]. Automated methods can tease apart ensemble-averaged density maps and explicitly model alternative conformations [Wankowicz et al., 2024, Flowers et al., 2024, Keedy et al., 2015], which are identified using the "altloc" identifier in the PDB and mmCIF specification. This data gets discarded during the preprocessing of training data for all currently available biomolecular deep

learning models [Abramson et al., 2024], providing a natural evaluation set for evaluating how well these methods capture conformational changes.

The most recent structure prediction methods [Abramson et al., 2024, Passaro et al., 2025, Chai Discovery et al., 2024, ByteDance AML AI4Science Team et al., 2025, Corley et al., 2025] use a diffusion model to generate structures from noisy atomic coordinates conditioned on a learned evolutionary embedding. Diffusion methods can be integrated into guidance frameworks [Ho et al., 2020, Maddipatla et al., 2025, Levy et al., 2024, Raghu et al., 2025] to draw samples consistent with experimental data, providing a powerful approach for solving complex inverse problems [Chung et al., 2024, Bendel et al., 2025].

Guidance relies on the diffusion model to be an accurate prior over macromolecular structures. Structure predictors are highly accurate at predicting the structure of held-out protein sequences, but it is equally as important to predict the multiple conformations that a single sequence can adopt as it carries out its function. It is not yet clear that AlphaFold 3 and related models are accurate priors over conformational space. We evaluate this here using synthetic density maps calculated from alternative locations ("altlocs") in deposited protein structures to evaluate the limitations of using these models for conformational sampling and identify opportunities for resolving them. We select four proteins (Fig. 1) with significant backbone conformational changes for evaluating the performance of the Boltz-2 [Passaro et al., 2025] structure prediction model with density guidance and steering.

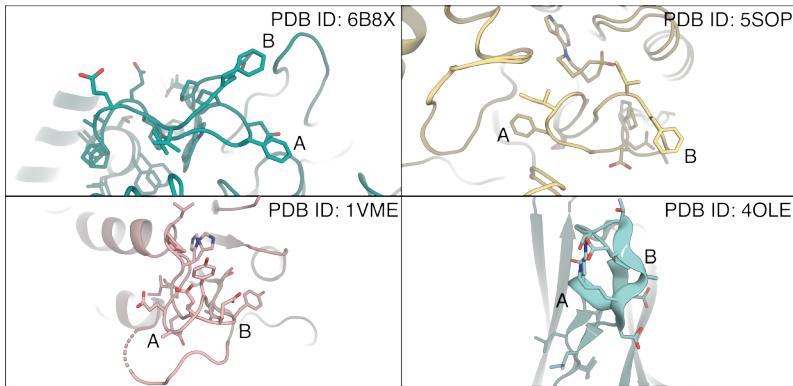


Figure 1: Conformational changes used to evaluate Boltz-2 in sampling alternative conformations using steering and guidance. A and B altlocs are shown for A) 6B8X, B) 5SOP, C) 1VME, D) 4OLE. Though 5SOP has 3 altlocs (A, B, and C), we only use altlocs A and B in calculating the synthetic density. We normalize the occupancies of A and B to 0.5 and 0.5 in all structures unless otherwise noted.

2 Results

We want to understand how proteins access their relevant functional states. In this paper, we evaluate performance on specific test cases with significant conformational changes (Figure 1). First, we chose Protein Tyrosine Phosphatase 1B (PTP1B) (Figure 1A), a protein previously studied with multitemperature crystallography and fragment screening [Keedy et al., 2018]. Two conformations were observed in the apo structure maps at temperatures ranging from 100K-278K, and both were manually built into the map. We were interested to see whether Boltz-2, which has been exposed to both conformations during training, could produce an ensemble that captured both conformations.

Boltz-2 was trained with additional conditioning on the method that the structure was solved with. While other structure predictors based on the AlphaFold 3 architecture produce slightly different ensembles, Boltz-2 with different method conditioning inputs could generate both conformations in 6B8X (Supp. Fig. 1).

With MD conditioning, Boltz primarily captures the closed state (Fig. 2). The MD simulations that are in the training dataset (from Vander Meersche et al. [2024], Siebenmorgen et al. [2024], Mirarchi et al. [2024]) are initialized from the closed state and contain a maximum of 500ns of simulation time, which is not sufficient to sample the conformational transition. The sampling of the closed-to-open transition in PTP1B required ~2.6 ms of simulation time, and the open-to-closed

transition did not occur in unbiased simulation [Yeh et al., 2023]. With the default conditioning method (X-ray diffraction), Boltz-2 samples only the open state. The MD-closed and X-ray-open results show a clear bias towards the conformations present in the training set - the MD simulations do not contain the open state, whereas the majority of X-ray structures deposited for the protein are ligand-bound structures in the open state (mostly from the fragment screens in Keedy et al. [2018], Skaist Mehlman et al. [2023]).

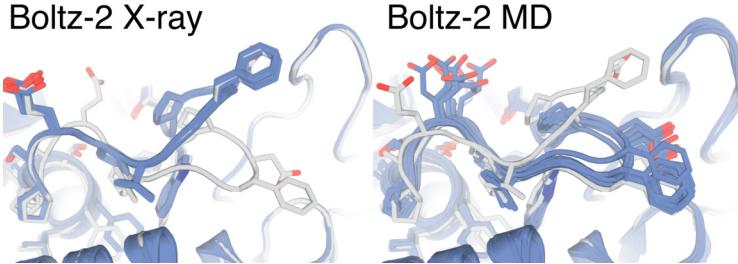


Figure 2: Boltz-2 predicts different conformations in PTP1B with X-ray vs. MD method conditioning. Ensemble of 8 structures produced through X-ray and MD conditioning. We observe that neither MD nor X-ray conditioning model both conformations in a given ensemble.

2.1 Guidance and steering of Boltz-2 with density enables sampling of alternative conformations

We reasoned that it might be possible to sample both conformations in a single ensemble by doing diffusion steering and guidance with a custom density potential, described in the Methods. Due to the challenges with solvent and background noise in experimental maps, as well as the perk of being able to control the occupancy of each state to unveil the strength of the model biases, we elected to use synthetic density maps calculated at the experimental resolution in the steering and guidance process.

We first tested our guidance and steering method using a simple synthetic example; a 15-residue Alanine helical peptide with the middle amino acid a Tryptophan. We provided a 2Å density map of two altlocs where the Trp is in a different rotamer state. Our guided sampling captured both conformations (Fig. 3). Next, we tested if the guidance method could infer conformations from the altlocs using full structures with synthetic data.

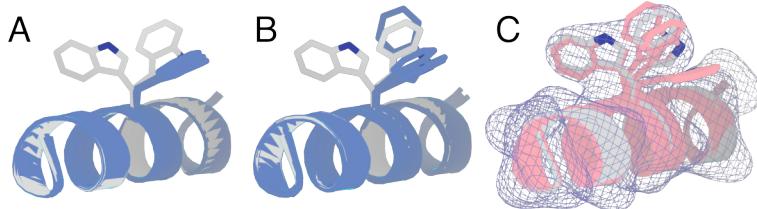


Figure 3: Guidance and steering on a synthetic peptide. A) The Boltz-2 model output (blue) with X-ray method conditioning and 4 diffusion samples. B) The Boltz-2 model output (blue) with MD method conditioning and 4 diffusion samples. C) The guided model output shown in pink with the 2Å map used during guidance and steering displayed as isomesh at contour 0.3 σ . The ground truth model with A and B altlocs is shown in gray in all panels.

For PTP1B (PDB ID: 6B8X, Fig. 1A), we are able to successfully capture both conformations (Fig. 4). We did a parameter sweep over guidance weights and ensemble sizes, which led to the observation that too high of a guidance weight leads to unphysical output structures (Supp. Fig. 3). With hyperparameters that maintained physical bond lengths and geometry, we were able to sample both conformations with MD but not X-ray conditioning. While not all the ensemble members fit well into the density, the two best fitting ensemble members by RSCC (Supp. Fig. 4) fit very well with the ground truth alternative conformations (0.903 Å RMSD to A and 0.860 Å RMSD to B over

residues 175-192). However, even with minimal occupancy in state A, we still do not capture a distribution biased towards the open state B when guiding Boltz-2 with MD conditioning (Fig. 4A). Conversely, with X-ray conditioning, we fail to capture the closed state without observing significant sample quality degradation (Supp. Fig. 3) that we attribute to guidance-induced drift from the data manifold, causing the denoising trajectory to traverse regions where the model’s score estimates are poorly calibrated.

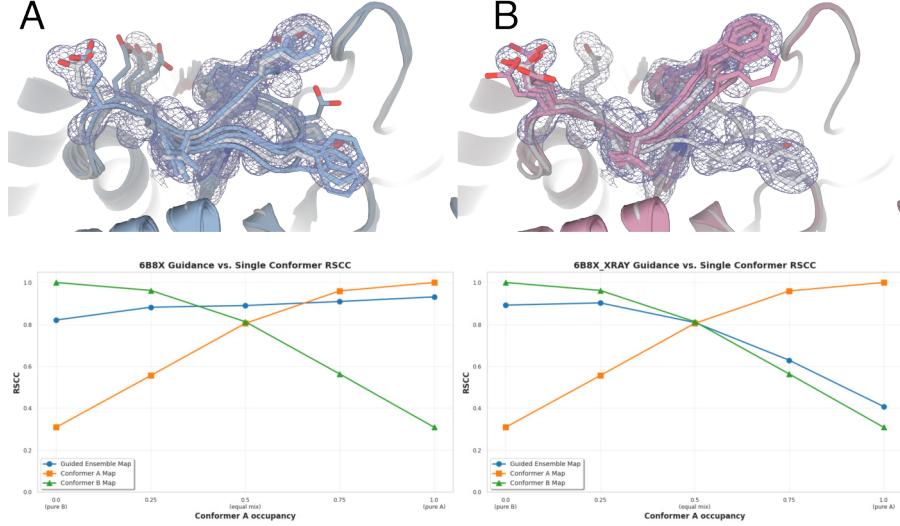


Figure 4: Ensemble guidance on MD but not X-ray conditioning achieves high correlation with density data. A) Guidance on Boltz-2 with MD conditioning recapitulates both underlying conformations and improves on just A or B conformations in the mixed density case. B) Guidance on X-ray conditioning suffers from bias towards the B conformation, and does not capture both conformations. Ensembles shown are from guidance with the 0.5 Å and 0.5 B density map, corresponding to the middle of the plots on the bottom row.

2.2 For proteins with few similar sequences in the training set and few representative conformations, guidance and steering fails to overcome bias

We also sought to evaluate the density method on more challenging cases. For Flavoprotein TM0755 (PDB ID 1VME) [van den Bedem et al., 2005], there are only 3 other proteins in the training cluster (4DIK, 4DIL, 5V8S) for Boltz-2. 4DIK, 4DIL, and 5V8S are all point mutants of 1VME and two (4DIK, 5V8S) contain the conformation B of 1VME, while 4DIL has conformation A. In this case, neither the MD and X-ray conditioned outputs capture A or B exactly (Supp. Fig. 5). We also fail to achieve an accurate recapitulation of both conformations with density guidance without degrading the model quality (Supp. Fig. 3). The guided model fails to achieve an adequate real space correlation coefficient (RSCC) in the mixed occupancy case, and does not improve over the correlation of just A or B with the mixed map (Fig. 5A).

The neighbor of BRCA1 gene 1 NBR1 (PDB ID 4OLE) presents an interesting case where there is only one structure in the training cluster. In this case, the MD and X-ray predictions from Boltz-2 do not cleanly align with either conformation (Supp. Fig. 6), and with density steering and guidance we are still unable to capture either conformations accurately (Fig. 5B). We note that this structure was also used in Maddipatla et al. [2025], and they observe successful capture of both conformations even though the AF3 prediction only captures the A state. When we restrict the diffusion to occur only in the altloc region, we are able to capture both backbone (though not sidechain) conformations. However, you may not know the location of the conformational change *a priori*, and when doing density guidance on the whole structure we are not able to capture both conformations. By default, models generally sample only the A conformation with multiple seeds (Supp. Fig. 2).

It is important to note that both the altloc regions in 1VME and 4OLE are relatively low confidence (Supp. Fig. 7). Ideally, this would mean that the region is more amenable to guidance because the model is uncertain about the exact atomic coordinates and thus would be tolerant of perturbations. This does not appear to be the case for these model currently, and we note that diffusion models

broadly are still limited in accurately representing prior knowledge of physics during sampling [Buttenschoen et al., 2024].

Another conformational change with functional impact we tested was observed from a fragment screen in the SARS-CoV-2 NSP3 macrodomain (PDB 5SOP). The deposited model contains 3 altlocs, one of which is a large backbone opening that accommodates binding of a fragment [Gahbauer et al., 2023]. Being able to model similar states across targets would enable in-silico structural screening against low population cryptic pockets [Pearce et al., 2017]. We observed that Boltz-2 only accurately predicts the ground state (altloc A) (Fig. 5C) with density guidance, and attempts to increase the weight of the density guidance lead to poor quality models (Supp. Fig. 3). Conformation B does not appear in any of the structures in the training set, and the model remains biased towards sampling only the conformation present in the structures it was trained on. This bias limits the application of these models to virtual screening, as it requires sampling structures that are physically plausible, even if they are only present at low population (as is the case with this structure).

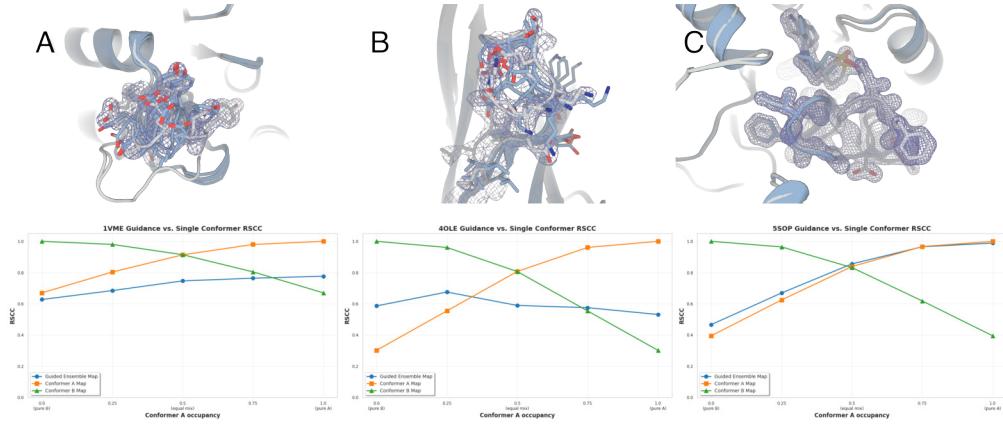


Figure 5: Boltz-2 is unable to recover conformations when the alternative conformations are not well represented in the training set. For PDBs 1VME, 4OLE, and 5SOP, guidance is not able to fit into the density as well as either A or B in the mixed case, and for 1VME, fails to fit as well as either conformation in any case except the only B conformation map compared to the only A conformation map. In the case of 5SOP, the generated ensemble only matches the A conformation, aligning completely with the A conformation map correlations.

3 Discussion

While we see that Boltz-2 fails to reliably sample these conformational changes, it is possible that the specifics of the training procedures in the other structure predictors enable them to sample some of these conformations. We still anticipate that no current structure predictor will generalize to all conformational changes observed in altlocs in the PDB, and evaluating on altloc data will be an important test for future models. We expect that updated all-atom models trained on subsets of structural data that contain the type of diversity required to capture these conformational changes (a la BioEmu [Lewis et al., 2025]) will help improve conformational sampling. Furthermore, we expect that training on altlocs should enable the structural representations learned by these models to accommodate of the types of conformational changes observed in the underlying experimental data, enabling a flywheel of iterative retraining and improvement by increasing the agreement of structure prediction models with the data derived from X-ray crystallography, CryoEM, and NMR [Fadini et al., 2025].

We also acknowledge that our guidance and steering method could be limiting us from sampling the alternative conformations. We plan to apply more recent methods for inference-time augmentation of these models, as current methods often incur significant bias [Xu et al., 2025, Feng et al., 2025, Ren et al., 2025]. We are also actively expanding the size of the altloc dataset, and will release that updated data and evaluation framework in the coming months. We hope that by highlighting this set of failure modes, we can point the community towards specific interventions in model development that can enable the models to be broadly transformative for understanding protein function, in-silico drug screening, and protein design.

References

- Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J. Ballard, Joshua Bambrick, Sebastian W. Bodenstein, David A. Evans, Chia-Chun Hung, Michael O'Neill, David Reiman, Kathryn Tunyasuvunakool, Zachary Wu, Akvilé Žemgulytė, Eirini Arvaniti, Charles Beattie, Ottavia Bertolli, Alex Bridgland, Alexey Cherepanov, Miles Congreve, Alexander I. Cowen-Rivers, Andrew Cowie, Michael Figurnov, Fabian B. Fuchs, Hannah Gladman, Rishub Jain, Yousuf A. Khan, Caroline M. R. Low, Kuba Perlin, Anna Potapenko, Pascal Savy, Sukhdeep Singh, Adrian Stecula, Ashok Thillaisundaram, Catherine Tong, Sergei Yakneen, Ellen D. Zhong, Michal Zielinski, Augustin Žídek, Victor Bapst, Pushmeet Kohli, Max Jaderberg, Demis Hassabis, and John M. Jumper. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 630(8016):493–500, June 2024. ISSN 1476-4687. doi: 10.1038/s41586-024-07487-w.
- Roman V. Agafonov, Christopher Wilson, Renee Otten, Vanessa Buosi, and Dorothee Kern. Energetic dissection of Gleevec's selectivity toward human tyrosine kinases. *Nature Structural & Molecular Biology*, 21(10):848–853, October 2014. ISSN 1545-9985. doi: 10.1038/nsmb.2891.
- Matt C. Bendel, Saurav K. Shastri, Rizwan Ahmad, and Philip Schniter. Solving Inverse Problems using Diffusion with Fast Iterative Renoising, January 2025.
- Martin Buttenschoen, Garrett M. Morris, and Charlotte M. Deane. PoseBusters: AI-based docking methods fail to generate physically valid poses or generalise to novel sequences. *Chemical Science*, 15(9):3130–3139, February 2024. ISSN 2041-6539. doi: 10.1039/D3SC04185A.
- ByteDance AML AI4Science Team, Xinshi Chen, Yuxuan Zhang, Chan Lu, Wenzhi Ma, Jiaqi Guan, Chengyue Gong, Jincai Yang, Hanyu Zhang, Ke Zhang, Shenghao Wu, Kuangqi Zhou, Yanping Yang, Zhenyu Liu, Lan Wang, Bo Shi, Shaochen Shi, and Wenzhi Xiao. Protenix - Advancing Structure Prediction Through a Comprehensive AlphaFold3 Reproduction, January 2025.
- 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.
- Hyungjin Chung, Jeongsol Kim, Michael T. Mccann, Marc L. Klasky, and Jong Chul Ye. Diffusion Posterior Sampling for General Noisy Inverse Problems, May 2024.
- Nathaniel Corley, Simon Mathis, Rohith Krishna, Magnus S. Bauer, Tuscan R. Thompson, Woody Ahern, Maxwell W. Kazman, Rafael I. Brent, Kieran Didi, Andrew Kubaney, Lilian McHugh, Arnav Nagle, Andrew Favor, Meghana Kshirsagar, Pascal Sturmels, Yanjing Li, Jasper Butcher, Bo Qiang, Lars L. Schaaf, Raktim Mitra, Katelyn Campbell, Odin Zhang, Roni Weissman, Ian R. Humphreys, Qian Cong, Jonathan Funk, Shreyash Sonthalia, Pietro Lio, David Baker, and Frank DiMaio. Accelerating Biomolecular Modeling with AtomWorks and RF3, August 2025. ISSN 2692-8205.
- Alisia Fadini, Minhuan Li, Airlie J. McCoy, 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, February 2025.
- Hanqi Feng, Peng Qiu, Meng-Chun Zhang, You Fan, Yiran Tao, and Barnabas Poczos. Soft Metropolis-Hastings Correction for Generative Model Sampling, September 2025. ISSN 2692-8205.
- Jessica Flowers, Nathaniel Echols, Galen Correy, Priya Jaishankar, Takaya Togo, Adam R. Renslo, Henry van den Bedem, James S. Fraser, and Stephanie A. Wankowicz. Expanding Automated Multiconformer Ligand Modeling to Macrocycles and Fragments, September 2024.
- James S. Fraser, Michael W. Clarkson, Sheena C. Degnan, Renske Erion, Dorothee Kern, and Tom Alber. Hidden alternative structures of proline isomerase essential for catalysis. *Nature*, 462(7273):669–673, December 2009. ISSN 1476-4687. doi: 10.1038/nature08615.

Stefan Gahbauer, Galen J. Correy, Marion Schuller, Matteo P. Ferla, Yagmur Umay Doruk, Moira Rachman, Taiasean Wu, Morgan Diolaiti, Siyi Wang, R. Jeffrey Neitz, Daren Fearon, Dmytro S. Radchenko, Yurii S. Moroz, John J. Irwin, Adam R. Renslo, Jenny C. Taylor, Jason E. Gestwicki, Frank von Delft, Alan Ashworth, Ivan Ahel, Brian K. Shoichet, and James S. Fraser. Iterative computational design and crystallographic screening identifies potent inhibitors targeting the Nsp₃ macromdomain of SARS-CoV-2. *Proceedings of the National Academy of Sciences*, 120(2): e2212931120, January 2023. doi: 10.1073/pnas.2212931120.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising Diffusion Probabilistic Models, December 2020.

Daniel A. Keedy, James S. Fraser, and Henry van den Bedem. Exposing Hidden Alternative Backbone Conformations in X-ray Crystallography Using qFit. *PLOS Computational Biology*, 11(10):e1004507, October 2015. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1004507.

Daniel A Keedy, Zachary B Hill, Justin T Biel, Emily Kang, T Justin Rettenmaier, José Brandão-Neto, Nicholas M Pearce, Frank von Delft, James A Wells, and James S Fraser. An expanded allosteric network in PTP1B by multitemperature crystallography, fragment screening, and covalent tethering. *eLife*, 7:e36307, June 2018. ISSN 2050-084X. doi: 10.7554/eLife.36307.

Axel Levy, Eric R. Chan, Sara Fridovich-Keil, Frédéric Poitevin, Ellen D. Zhong, and Gordon Wetzstein. Solving Inverse Problems in Protein Space Using Diffusion-Based Priors, June 2024.

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): ead9817, July 2025. doi: 10.1126/science.adv9817.

Advaith Maddipatla, Nadav Bojan Sellam, Meital Bojan, Sanketh Vedula, Paul Schanda, Ailie Marx, and Alex M. Bronstein. Inverse problems with experiment-guided AlphaFold, February 2025.

Antonio Mirarchi, Toni Giorgino, and Gianni De Fabritiis. mdCATH: A Large-Scale MD Dataset for Data-Driven Computational Biophysics. *Scientific Data*, 11(1):1299, November 2024. ISSN 2052-4463. doi: 10.1038/s41597-024-04140-z.

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.

N. M. Pearce, T. Krojer, and F. von Delft. Proper modelling of ligand binding requires an ensemble of bound and unbound states. *Acta Crystallographica Section D: Structural Biology*, 73(3):256–266, March 2017. ISSN 2059-7983. doi: 10.1107/S2059798317003412.

Edward Prince, editor. *International Tables for Crystallography. C: Mathematical, Physical and Chemical Tables / Hrsg. E. Prince*. Springer Netherland, Berlin, 3. ed., 1. online ed edition, 2007. ISBN 978-1-4020-1900-5 978-1-4020-5408-2. doi: 10.1107/97809553602060000103.

Rishwanth Raghu, Axel Levy, Gordon Wetzstein, and Ellen D. Zhong. Multiscale guidance of AlphaFold3 with heterogeneous cryo-EM data, June 2025.

Yinuo Ren, Wenhao Gao, Lexing Ying, Grant M. Rotskoff, and Jiequn Han. DriftLite: Lightweight Drift Control for Inference-Time Scaling of Diffusion Models, September 2025.

Till Siebenmorgen, Filipe Menezes, Sabrina Benassou, Erinc Merdivan, Kieran Didi, André Santos Dias Mourão, Radosław Kitel, Pietro Liò, Stefan Kesselheim, Marie Piraud, Fabian J. Theis, Michael Sattler, and Grzegorz M. Popowicz. MISATO: Machine learning dataset of protein–ligand complexes for structure-based drug discovery. *Nature Computational Science*, 4(5):367–378, May 2024. ISSN 2662-8457. doi: 10.1038/s43588-024-00627-2.

Raghav Singhal, Zachary Horvitz, Ryan Teehan, Mengye Ren, Zhou Yu, Kathleen McKeown, and Rajesh Ranganath. A General Framework for Inference-time Scaling and Steering of Diffusion Models, January 2025.

Tamar Skaist Mehlman, Justin T Biel, Syeda Maryam Azeem, Elliot R Nelson, Sakib Hossain, Louise Dunnett, Neil G Paterson, Alice Douangamath, Romain Talon, Danny Axford, Helen Orins, Frank von Delft, and Daniel A Keedy. Room-temperature crystallography reveals altered binding of small-molecule fragments to PTP1B. *eLife*, 12:e84632, March 2023. ISSN 2050-084X. doi: 10.7554/eLife.84632.

H. van den Bedem, I. Lotan, J.-C. Latombe, and A. M. Deacon. Real-space protein-model completion: An inverse-kinematics approach. *Acta Crystallographica Section D: Biological Crystallography*, 61(1):2–13, January 2005. ISSN 0907-4449. doi: 10.1107/S0907444904025697.

Yann Vander Meersche, Gabriel Cretin, Aria Gheeraert, Jean-Christophe Gelly, and Tatiana Galochkina. ATLAS: Protein flexibility description from atomistic molecular dynamics simulations. *Nucleic Acids Research*, 52(D1):D384–D392, January 2024. ISSN 0305-1048. doi: 10.1093/nar/gkad1084.

Stephanie A Wankowicz, Ashraya Ravikumar, Shivani Sharma, Blake Riley, Akshay Raju, Daniel W Hogan, Jessica Flowers, Henry Van Den Bedem, Daniel A Keedy, and James S Fraser. Automated multiconformer model building for X-ray crystallography and cryo-EM. *eLife*, 12:RP90606, June 2024. ISSN 2050-084X. doi: 10.7554/eLife.90606.

Jeremy Wohlwend, Gabriele Corso, Saro Passaro, Noah Getz, Mateo Reveiz, Ken Leidal, Wojtek Swiderski, Liam Atkinson, Tally Portnoi, Itamar Chinn, Jacob Silterra, Tommi Jaakkola, and Regina Barzilay. Boltz-1: Democratizing Biomolecular Interaction Modeling, April 2025.

Tongda Xu, Xiyan Cai, Xinjie Zhang, Xingtong Ge, Dailan He, Ming Sun, Jingjing Liu, Ya-Qin Zhang, Jian Li, and Yan Wang. Rethinking Diffusion Posterior Sampling: From Conditional Score Estimator to Maximizing a Posterior, January 2025.

Christine Y. Yeh, Jesus A. Izaguirre, Jack B. Greisman, Lindsay Willmore, Paul Maragakis, and David E. Shaw. A Conserved Local Structural Motif Controls the Kinetics of PTP1B Catalysis. *Journal of Chemical Information and Modeling*, 63(13):4115–4124, July 2023. ISSN 1549-9596. doi: 10.1021/acs.jcim.3c00286.

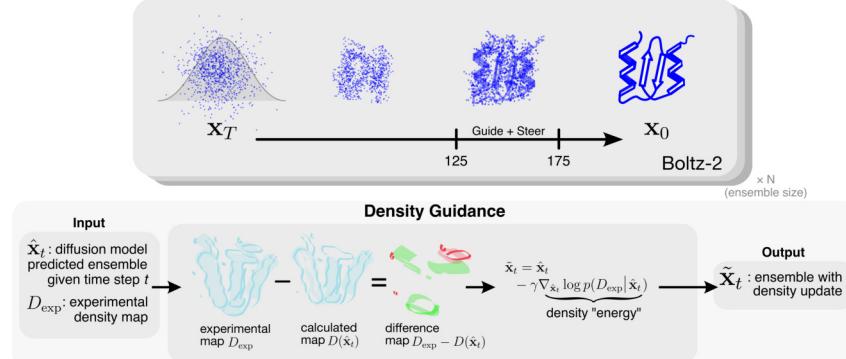
A Methods

A.1 Guidance and steering

We use the Boltz-2 (version 2.2.1) [Passaro et al., 2025] model along with the Feynman-Kac steering [Singhal et al., 2025] framework and guidance that was implemented in Boltz-1x to steer the model to sample structures consistent with the density map. To do this, we use a custom density potential defined by differentiably calculating a density map based on the estimated \hat{x}_0 ensemble from the model at the given time step and calculating the L_2 -norm of the difference between the calculated map and the experimental map. We then update the estimated \hat{x}_0 ensemble with the gradient of the L_2 -norm of the difference map. In order to maintain alignment with the density map, we compute a Kabsch alignment of each ensemble member to the ground-truth structure. For the density, non-i.i.d. guidance and resampling is applied over the whole ensemble as is done in Maddipatla et al. [2025] and resampling is done over ensemble-particles based on the density fit, though we plan to experiment with different guidance and steering frameworks in the future. The algorithmic flow of guidance and steering is documented in Algorithm 1, adapted from Wohlwend et al. [2025] Algorithm 2.

We apply guidance only in timesteps 125-175 (out of 200 total), split up in two sections. The first section guides on a downsampled 8Å map from steps 125 to 150, and the second on the full resolution from 150-175. During tuning of the guidance process, we noticed that running guidance during the whole trajectory caused significant increases in the score early on, and the guidance objective was not able to recover plausible structures. This was also done in Raghu et al. [2025]. We make our code available and open on GitHub.

We note that this guidance framework is sensitive to the choice of the diffusion model prior. We implemented a version of Maddipatla et al. [2025] using the Boltz-1 model, but in the one structure that overlaps with our work here (4OLE) we were not able to accurately sample the A and B conformations. We also do not sample both conformations using Feynman-Kac steering with Boltz-2, but in recent tests with Protenix [ByteDance AML AI4Science Team et al., 2025], the AF3 replication they use for the underlying diffusion model, we see improved results 8. We observe that the diversity of samples from different seeds may impact this (shown for PTP1B in Supp. Fig. 1), as the sampled ensembles will change depending on the exact training process even if the model architectures are similar.



Methods Figure 1: Using Boltz-2 [Passaro et al., 2025], we sample protein ensembles by optimizing for fit with experimental density maps. At each step, we compute the density map for the diffusion prediction and subtract it from the experimental density map to obtain a difference map. We take gradient steps to minimize the L_2 -norm of the difference map and use the updated coordinates for the next diffusion step. After sampling an ensemble x_0 from the method, we evaluate it against the ground truth density map and ground truth coordinates when available.

A.2 Synthetic density calculation

We calculate synthetic density maps using the commonly used approximation of 5 Gaussians centered on each atomic coordinate, each scaled according to empirically derived coefficients defined in Prince [2007]. To accomplish density guidance with a low memory burden over the structure prediction

model, we implemented a CUDA kernel that computes each atom’s contribution to the density in each voxel. We implement our framework in PyTorch and make the custom CUDA kernel available as a PyTorch extension.

Algorithm 1 Density and substructure steering for Boltz-2

Require: $g(\cdot)$ (density forward model), y (density map), \mathbf{x}_{gt} (substructure) $p_\theta(\cdot, t)$, T , $f_\theta(\cdot, t)$, $\{\zeta_t\}_{t=1}^T$ (guidance weight schedule), k (steering particles) ▷ Input parameters

Initialize $\Delta\hat{\mathbf{x}}_T^{(i)} \leftarrow \mathbf{0}$ for $i \in [k]$ ▷ Initial gradient updates, for $i \in [k]$

for $t = T, \dots, 0$ **do**

- if** $t = T$ **then**

 - Sample $\mathbf{x}_T^{(i)} \sim p_\theta(\mathbf{x}_T^{(i)} | T)$ ▷ Sample from initial distribution

- else**

 - Sample $\mathbf{x}_t^{(i)} \sim \tau(\mathbf{x}_t^{(i)} | \hat{\mathbf{x}}_{t+1}^{(i)} + \Delta\hat{\mathbf{x}}_{t+1}^{(i)}, \mathbf{x}_{t+1}^{(i)}, t)$ ▷ Sample from proposal distribution

- end if**
- Predict $\hat{\mathbf{x}}_0^{(i)} \leftarrow f_\theta(\mathbf{x}_t^{(i)}, t)$
- if** $T - t \text{ (mod } 3) = 0$ **then**

 - Compute $G^{(i)} = \frac{\tau(\mathbf{x}_t^{(i)} | \hat{\mathbf{x}}_{t+1}^{(i)}, \mathbf{x}_{t+1}^{(i)}, t)}{\tau(\mathbf{x}_t^{(i)} | \hat{\mathbf{x}}_{t+1}^{(i)} + \Delta\hat{\mathbf{x}}_{t+1}^{(i)}, \mathbf{x}_{t+1}^{(i)}, t)} G_t(\mathbf{x}_T^{(i)}, \dots, \mathbf{x}_t^{(i)})$
 - Resample $\mathbf{x}_t^{(i)} \sim \text{Multinomial}(\mathbf{x}_t^{(j)}, G^{(j)})$

- end if**
- Initialize $\Delta\hat{\mathbf{x}}_t^{(i)} \leftarrow \mathbf{0}$ for $i \in [k]$
- $\Delta\hat{\mathbf{x}}_t^{(i)} \leftarrow \Delta\hat{\mathbf{x}}_t^{(i)} - \nabla \|g(\hat{\mathbf{x}}_t^{(i)}) - y\|^2$ ▷ Density potential
- $\Delta\hat{\mathbf{x}}_t^{(i)} \leftarrow \Delta\hat{\mathbf{x}}_t^{(i)} - \nabla \|\hat{\mathbf{x}}_t^{(i)} - \mathbf{x}_{gt}\|^2$ ▷ Substructure potential
- for** $j = 1, \dots, m$ **do**

 - $\Delta\hat{\mathbf{x}}_t^{(i)} \leftarrow \Delta\hat{\mathbf{x}}_t^{(i)} - \nabla E_j(\hat{\mathbf{x}}_t^{(i)} + \Delta\hat{\mathbf{x}}_t^{(i)})$ ▷ Base Boltz-1x potentials

- end for**

end for

Output $\mathbf{x}_0^{(i)} + \Delta\hat{\mathbf{x}}_0^{(i)}$ ▷ Return final ensemble

A.3 Metrics

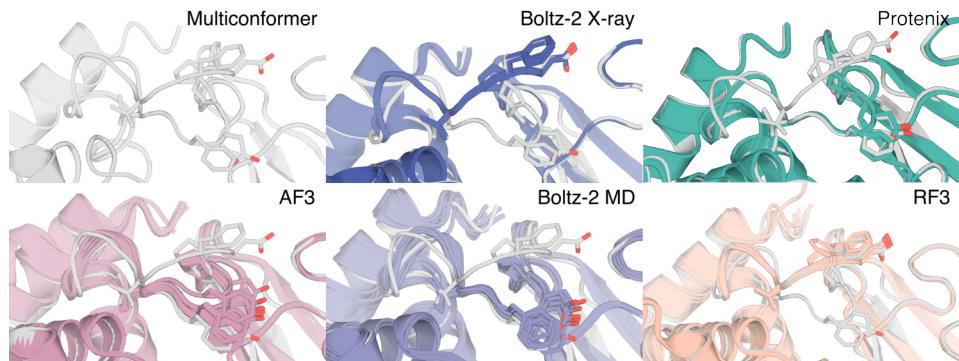
To compare the generated ensembles with the synthetic density and the ground truth conformations, we use the real-space correlation coefficient between the voxels bounded by the comparison segments in Table 1. Where we report RMSD to the ground truth alternative conformations, we also use these same segments.

For all proteins, we also determined the size of the training cluster. We use the same 40% sequence identity clustering and the cutoffs described in Passaro et al. [2025].

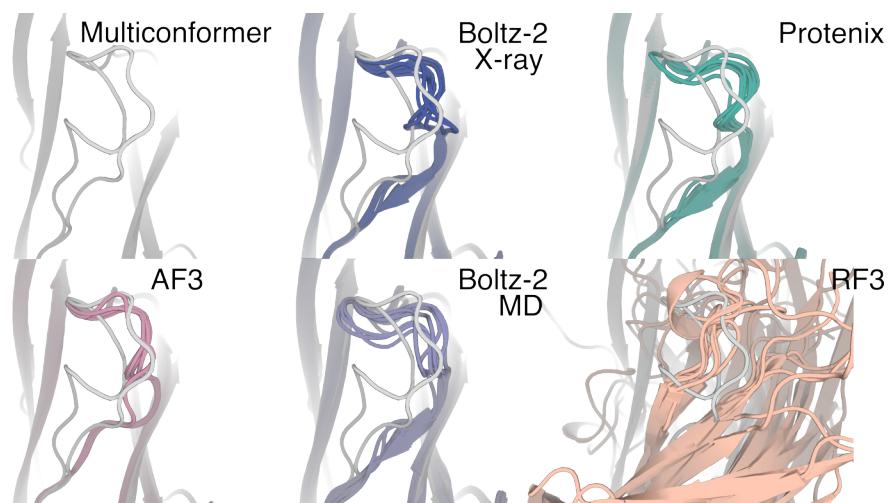
Table 1: Segments from deposited proteins used for comparison metrics. All residue numbers are starting from the the first modeled residue as residue 1.

PDB ID	Comparison residue segment	Guidance residue segment	Training cluster size
6B8X	180-184	175-192	349
1VME	326-339	326-339	4
4OLE	60-67	58-70	1
5SOP	129-135	120-140	481

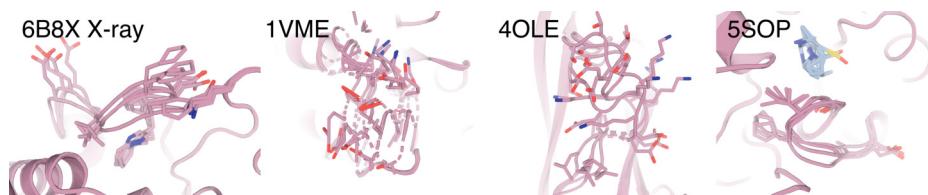
B Supplemental Figures



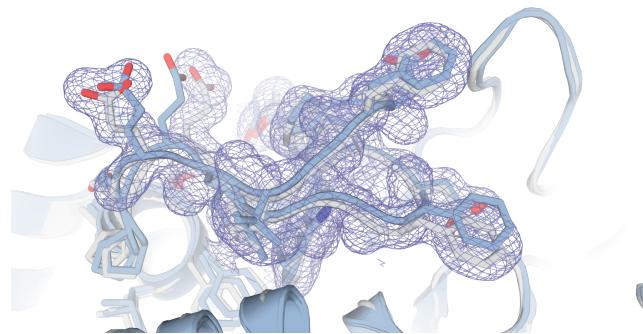
Supplementary Figure 1: AlphaFold 3 (AF3), Boltz-2 X-ray and MD, Protenix, and RosettaFold 3 (RF3) output with seeds 1, 2, 4, 8, 16, 32, 64, 128 on PTP1B (using sequence from PDB 6B8X).



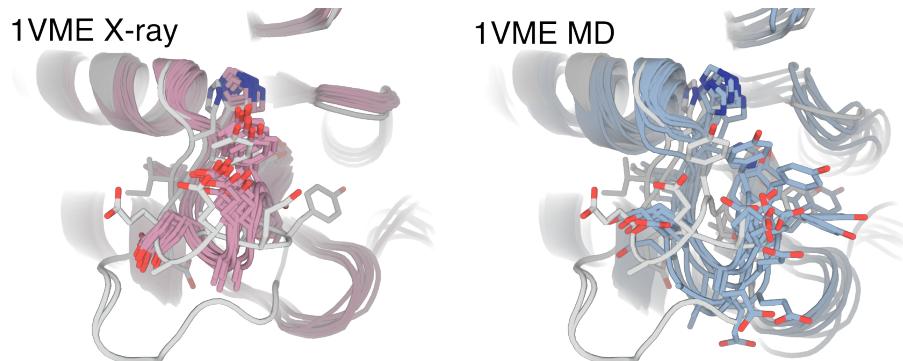
Supplementary Figure 2: AlphaFold 3 (AF3), Boltz-2 X-ray and MD, Protenix, and RosettaFold 3 (RF3) output with seeds 1, 2, 4, 8, 16, 32, 64, 128 on 4OLE. Most conformations align better with conformation A (right) than B (left), though RF3 fails to capture either and has a significant overall RMSD from the deposited structure.



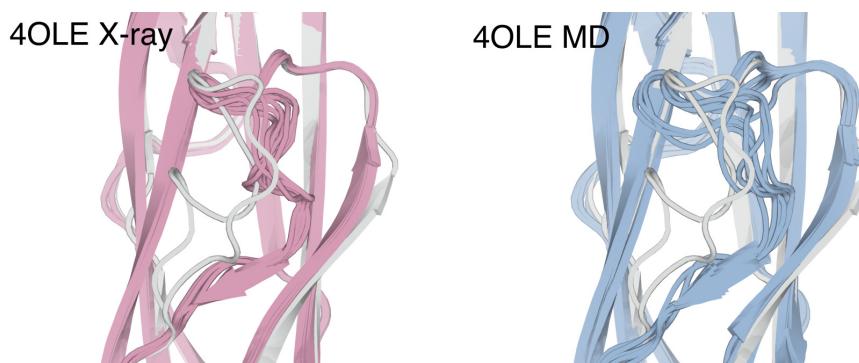
Supplementary Figure 3: Increasing strength of density guidance does not improve fidelity to density and instead damages model quality, evidenced by clashes and ring pucker in 6B8X with X-ray method conditioning, bond breakage in 1VME, chain break in 4OLE, and ring pucker and ligand distortion in 5SOP.



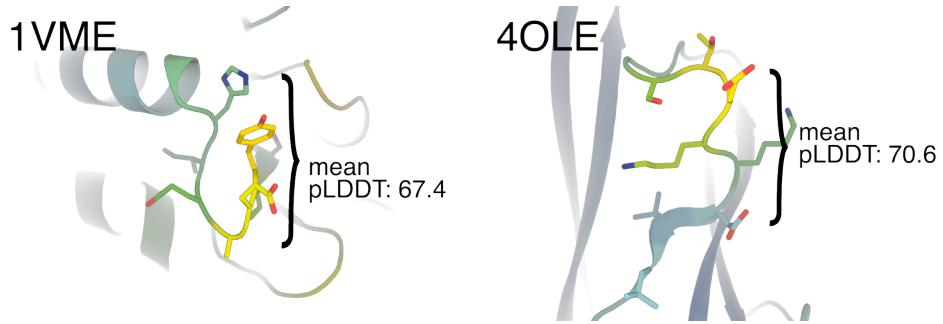
Supplementary Figure 4: Best 2 ensemble members by RSCC from guided and steered sampling from the MD-conditioned model for 6B8X.



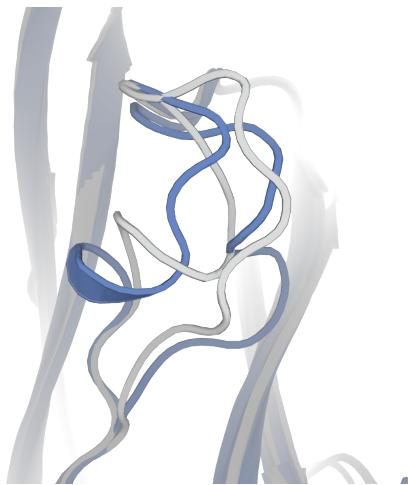
Supplementary Figure 5: Eight samples of 1VME from Boltz-2 with X-ray and MD method conditioning. (MD min RMSD to A: 3.081 Å, min RMSD to B: 2.426 Å, X-ray min RMSD to A: 1.486 Å, min RMSD to B: 1.504 Å)



Supplementary Figure 6: Eight samples of 4OLE from Boltz-2 with X-ray and MD method conditioning.



Supplementary Figure 7: 1VME and 4OLE mean pLDDT in the altloc region (residues 316-323 in 1VME, 423-431 in 4OLE).



Supplementary Figure 8: Best 2 ensemble members from steered 4OLE ensemble predicted by Protenix.