Predicting Relative Populations of Protein Conformations without a Physics Engine Using AlphaFold2

Gabriel Monteiro da Silva and Jennifer Y. Cui
Brown University Department of Molecular Biology, Cell Biology, and Biochemistry, Providence, RI, USA

David C. Dalgarno
Dalgarno Scientific LLC, Brookline, MA, USA

George P. Lisi and Brenda M. Rubenstein
Brown University Department of Molecular Biology, Cell Biology, and Biochemistry
Brown University Department of Chemistry
Providence, RI, USA

July 25, 2023

Abstract

This paper presents a novel approach for predicting the relative populations of protein conformations using AlphaFold 2, an AI-powered method that has revolutionized biology by enabling the accurate prediction of protein structures. While AlphaFold 2 has shown exceptional accuracy and speed, it is designed to predict proteins' single ground state conformations and is limited in its ability to predict fold switching and the effects of mutations on conformational landscapes. Here, we demonstrate how AlphaFold 2 can directly predict the relative populations of different conformations of proteins and even accurately predict changes in those populations induced by mutations by subsampling multiple sequence alignments. We tested our method against NMR experiments on two proteins with drastically different amounts of available sequence data, Abl1 kinase and the granulocyte-macrophage colony-stimulating factor, and predicted their relative state populations with accuracies in excess of 80%. Our method offers a fast and cost-effective way to predict protein conformations and their relative populations at even single point mutation resolution, making it a useful tool for pharmacology, analyzing NMR data, and studying the effects of evolution.

Introduction

Proteins are essential biomolecules that carry out a wide range of functions in living organisms. Understanding their three-dimensional structures is critical for elucidating their functions and designing drugs that target them [1]. Historically, experimental techniques such as X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, and electron microscopy have been used to determine protein structures [2–4]. However, these methods can be time-consuming, technically challenging, and expensive, and may not work for all proteins [5]. To meet this challenge, *ab initio* structure prediction methods, which use computational algorithms to predict protein structures from their amino acid sequences, have been developed [6]. For many years, *ab initio* structure prediction methods have relied on physics-based algorithms that leverage statistical mechanics to calculate the most stable protein structures [7]. Although highly successful, these methods are limited by their steep computational expense, and consequently, have struggled to predict the structures of larger and more complex proteins [8].

The recent development of machine learning algorithms has significantly improved the speed of protein structure prediction [9, 10]. One of the most remarkable achievements in this area is the AlphaFold 2 (AF2) engine developed by DeepMind, which uses a deep neural network to predict ground state protein structures from amino acid sequences [11, 12]. AlphaFold 2 was trained using large amounts of experimental data and incorporates co-evolutionary information from massive metagenomic databases [11]. Its accuracy has revolutionized the field of protein structure prediction [11, 13, 14], opening up new possibilities for drug discovery and basic research with clear consequences for human health [15, 16].

However, studies have found that the default AF2 algorithm is limited in its capacity to predict alternative protein conformations and the effects of sequence variants [17]. Although AF2's inability to predict multiple conformations is unsurprising given its initial scope, the capacity to make predictions of different conformations would be as revolutionary as the capacity to accurately predict ground states. Phenomena that involve different conformations such as fold-switching and order-disorder transitions are ubiquitous across the proteome [18, 19] and are directly tied to the activity of many macromolecules [20]. Moreover, methods that can rapidly predict multiple conformations may have the potential to revolutionize drug discovery by uncovering substantially more drug targets [21]. To fully realize this potential, methods like AF2 will need to account for the relative populations of different conformations (states) since the conformational equilibrium of drug receptors is directly related to their affinities for drugs [22, 23]. A prime example of this relationship is Imatinib [24], a tyrosine inhibitor whose exceptional selectivity for Abl1 kinase was found to be caused by the enzyme's significant preference for conformational states that facilitate Imatinib binding and subsequent induced-fit steps [25].

In an attempt to realize this potential, researchers have devised new ways of employing the AF method to detect conformational changes, with significant success in a few test cases [26–30]. Although AF2 conventionally fails to predict fold-switching, researchers have found that sub-sampling the input multiple sequence alignments (MSAs) and increasing the number of predictions leads to structural ensembles that capture different physiologically-relevant conformations from the same sequence [26]. These predictions can be used as seeds in molecular dynamics (MD) simulations seeking to explore larger swaths of the conformational space and the relative populations of each predicted state [27]. Despite being a significant improvement over methods that only predict ground states, methods such as these still rely on expensive MD simulations to infer

most relative state population information, which comes at a significant cost compared to simply running a prediction engine.

Here, we show that these MD simulation steps may be unnecessary if the goal is to discover major alternative conformations and their relative populations in a high-throughput fashion, such as for contrasting differences in the dynamics of orthologs of a protein of interest. We take inspiration for this work from the observation that proteins from the same evolutionary line can have differences in relative state populations that are strongly correlated with the genetic distance between them [25]. Since AF2 works by decoding co-evolutionary signals [15] and previous research has suggested that subsampling MSAs leads to accurate predictions of different conformations of the same protein [26], it seems reasonable to hypothesize that the instructions for fold-switching between major states could be decoded from sequence data alone. If this hypothesis holds true, AF2 and other AI-based methods could be capable of quantifying sequence-encoded fold-switching signals, which would make it possible to predict not only alternative conformations of the same protein but also changes in its relative state populations stemming from point mutations.

With this as motivation, we show how subsampling multiple sequence alignments can generate ensembles of protein conformations (see Figure 1 for an overview of our method) and systematically test AF2's capacity to predict differences in the populations of conformers in two illustrative proteins: the Abl1 tyrosine kinase core and the granulocyte-macrophage colony-stimulating factor (GMCSF). Diverging from previous works, as a first example, we focused on detecting changes in the active state population across the Src kinase to Abl1 evolutionary line and tested our method's ability to predict the effects of single and double point mutations known or suspected to shift state distributions. Crucially, we found that subsampled AF2 can qualitatively predict both the positive and negative effects of mutations on the active state populations of kinase cores with up to eighty percent accuracy. We also found that AF2 predicts most of the activation loop intermediate states in the active-to-inactive transition of the kinase core with an ensemble that is comparable to that obtained from multi-microsecond MD simulations. As a second example, we predicted changes in the relative state populations of GMCSF, a protein with minimal known homology, in response to point mutations. Our predictions strongly correlated with experimentally-determined NMR chemical shifts, further showcasing subsampled AF2's remarkable capacity to decode signals pertaining to conformational dynamics even when sequence data is scarce. Altogether, these results highlight the strong, yet untapped potential of AF2 for predicting conformational ensembles of mutated and evolutionarily-related proteins, which will have substantial impacts on the fields of biophysics, drug design, and NMR.

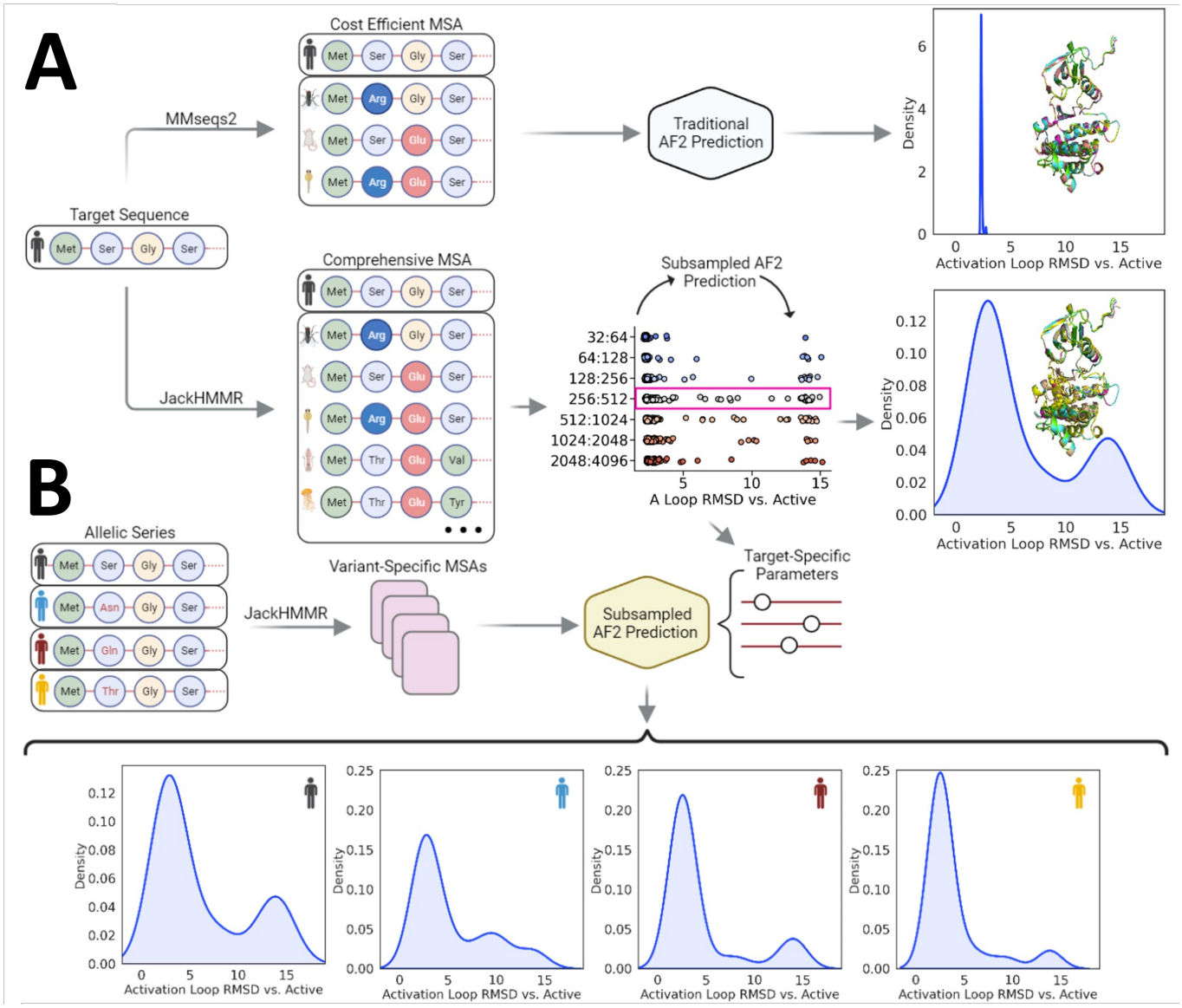

Figure 1: Summary of the subsampled AF2 workflow employed in this study. (A) Traditionally, AF2 predicts the structure of a target sequence by using a multiple sequence alignment (MSA) that is generated by methods that generally seek to balance sequence count, diversity, and speed. When running AF2 with such MSAs, the predicted structures are often similar to each other even with a large number of independent predictions (seeds). (B) In this study, we show that subsampling comprehensive MSAs causes AF2 to output predictions that occupy different conformations of the same protein, and the predicted frequency of each conformation based upon a range of random seeds strongly correlates with its experimentally-determined relative state population. In particular, we illustrate that subsampled AF2 is capable of predicting relative state populations of evolutionarily-related kinase cores and dynamics of the GMCSF protein with single substitution resolution.

Results

Optimizing MSA Subsampling to Predict Kinase Core Conformational Ensembles

In recent years, multiple groups have observed that AF2 with different parameters and inputs from MSAs is capable of predicting conformational changes based on sequence data alone [26, 27]. These alternative AF2 pipelines share the principle of subsampling MSAs to amplify co-evolutionary signals at different structural domains [26]. In its standard implementation, AF2 takes as input a target sequence and a corresponding multiple sequence alignment. An arbitrary number of sequences (defined by the *max_seq* parameter) are randomly selected from the master MSA (the target sequence is always selected), and the remaining sequences are clustered around each of the selected sequences using a Hamming distance. Both the cluster centers and a sample from each cluster with a length of *extra_seq* are used by AF2 for inference (see Figure 2). Previous works have shown that a significant reduction in the values of *max_seq* and *extra_seq* from their default values achieves ensemble prediction for a series of model systems [28].

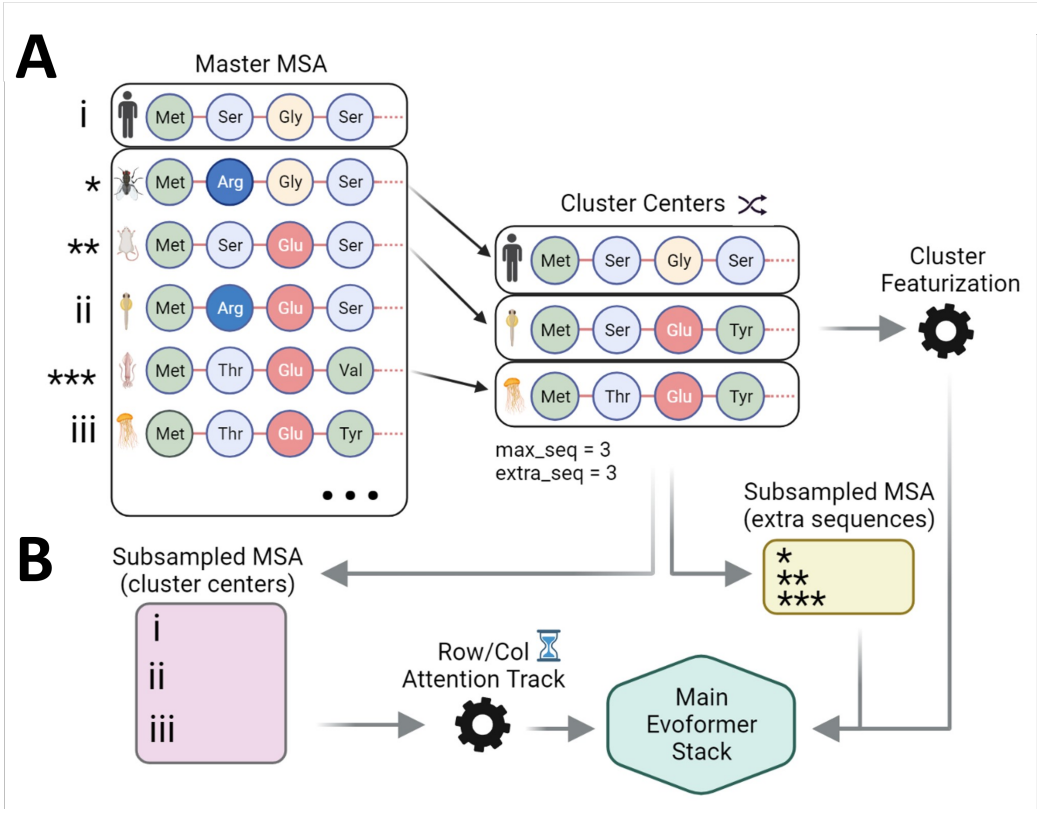

Figure 2: AF2's multiple sequence alignment (MSA) clustering heuristic. (A) An MSA of arbitrary length is built from a target sequence and passed to AF2, which randomly selects a number of sequences (defined by *max_seq*) from the input MSA. Each of the selected sequences becomes a cluster center around which the sequences not selected in the previous step are distributed. The target sequence is always selected as a cluster center. The clusters obtained through this process are featurized and relevant statistics are calculated. (B) All of the previously discussed elements are used by AF2 for inference. Cluster features and a number of random non-cluster-center sequences (defined by *extra_seq*) are processed and passed to the Main Evoformer Stack, while the MSA containing the cluster centers is processed, passed to the comparatively expensive row/col attention track, and then finally passed to the Main Evoformer Stack as well.

Motivated by these observations, we started our work by systematically testing the accuracy of different AF2 parameter combinations for predicting the Abl1 kinase core structural ensemble. We chose the Abl1 kinase core (residues 235-497) as our first test case due to this system's extensively documented conformational dynamics [22]. Abl1 is thought to occupy three major conformations with different populations (see Figure 3). In solution, Abl1 primarily exists in an active (ground) state, which as the name suggests, is capable of productive phosphorylation. Infrequently, Abl1 will switch to inactive state 1 (I1), and then to inactive state 2 (I2) [22], which strongly binds to Imatinib (Gleevec) [25]. While the change from ground to I1 is subtle and mostly involves shifting side-chain dihedrals, the transition from I1 to I2 involves considerable backbone rearrangements: the activation loop, composed of about 20 residues, detaches from its resting position below the C-helix and folds on itself, a change that involves the activation loop shifting by over 15 Å from

its original position as shown in Figure 3.

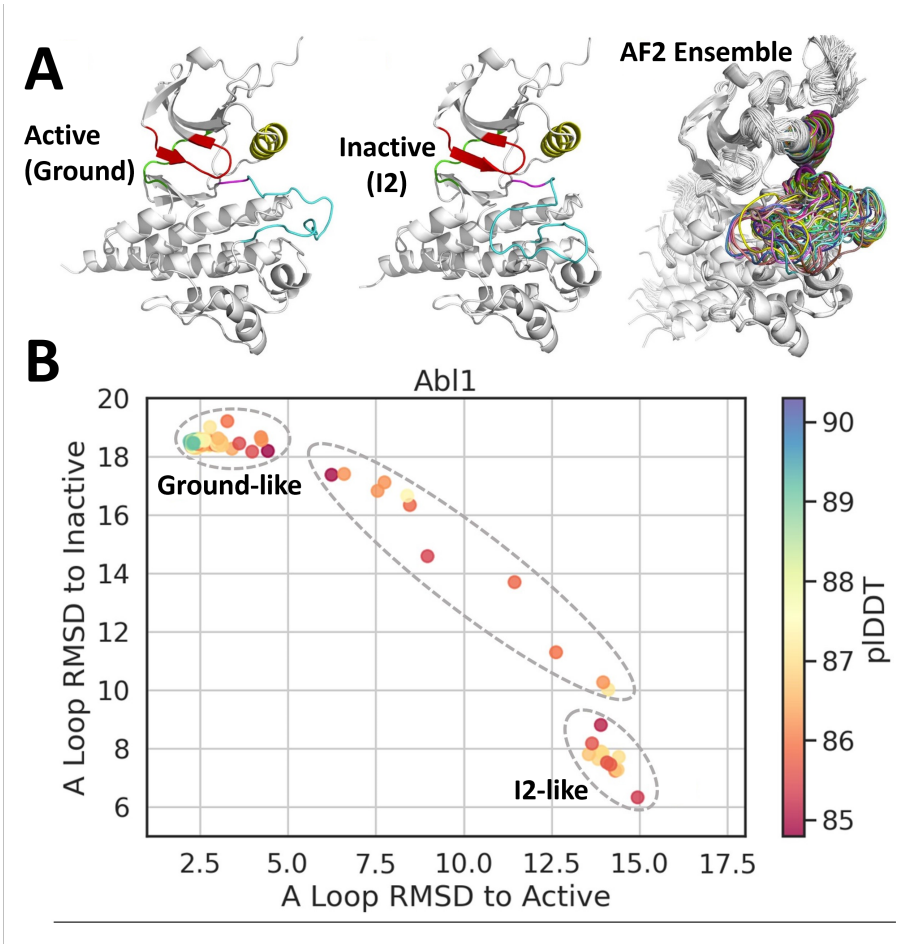

Figure 3: (A) Top: Three-dimensional models of the Abl1 kinase core in its active and inactive conformations. Structural elements that are relevant for kinase function are represented in different colors. Cyan: activation loop; red: phosphate-binding loop; yellow: activation helix; green: hinge; and pink: DFG motif. Models are taken from the PDB (6XR6 for active core, 6RXG for inactive core) [22]. Top Right: Ensemble of 160 three-dimensional models of the Abl1 kinase core as predicted by subsampled AF2, with the activation loop highlighted in different colors for each prediction. (B) Two-dimensional projection of 160 Abl1 kinase core predictions from subsampled AF2 plotted according to their backbone RMSDs relative to the Abl1 kinase core active (6XR6) and inactive states (6XRG). Data points are colored according to their average predicted local distance difference test score (pLDDT) as calculated by AF2, which is a metric of the confidence/accuracy of AF2 predictions [11]. Bottom Left: Cluster of predictions considered to be in the ground state. Bottom Right: Cluster of predictions considered to be in the inactive, I2 state.

The accuracy of our ensemble predictions was defined as their capacity to replicate the wild-type Abl1 kinase core conformations and their correct relative populations as validated by nuclear magnetic resonance experiments. Specifically, we sought a combination of parameters that led to an ensemble of predictions that met the following criteria: the ground state is the most frequent

prediction within the ensemble, the transition from the ground to I2 state is captured within the ensemble, and the I2 state is present in the ensemble more frequently than transition states. Importantly, we opted to examine the relative populations of the ground and I2 states because of the large backbone rearrangement involved in the transition between these conformations, which is more likely to be reproduced by AF2 than the comparatively small dihedral flips in the ground to I1 transition. We optimized the accuracy achieved as a function of the following parameters: *max_seq*, *extra_seq*, number of seeds, and number of recycles (see Table S1 for a complete list of tests and parameters). We evaluated the ensemble resulting from each parameter set by measuring the activation loop backbone RMSD relative to either the active kinase core (PDB 6XR6) or the I2 kinase core (PDB 6XRG) for each prediction. This decision is rooted in the fact that the activation loop is the structural element that changes the most (in terms of backbone motions) upon the transition from the ground to I2 state [22].

To encourage AF2 to generate a full ensemble of Abl1 conformations, we started by compiling an extensive MSA spanning over 600,000 sequences using the JackHMMR algorithm [31] on wild-type Abl1 kinase core (residues 235-497) sequences pulled from the Uniprot90 [32], Small BFD [33], and MGfny [34] databases. To increase the statistical power of our results, we then ran 32 predictions with independent seeds for each test, and enabled dropouts during inference to sample from the uncertainty of the models. All other parameters were left in their default settings (4 recycles per prediction, 5 models per seed, a total of 160 predictions per run, 3 independent runs with unique seeds, 480 predictions per test).

Consistent with previous observations, we found that merely changing *max_seq* and *extra_seq* was sufficient to obtain predictions that met all of the above criteria, with a *max_seq*:*extra_seq* ratio of 256:512 leading to the most diverse results in terms of activation loop conformations (see Figure 3). Importantly, the ensemble of activation loop conformations predicted by AF2 with the above parameter set is distributed across the ground state to I2 state transition in Abl1, with no predictions falling outside the boundaries of known activation loop conformations and no blatantly unphysical or misfolded predictions. As a further test to the claim that we are actually predicting conformations along a transition, we compared the ensemble of 160 subsampled AF2 Abl1 predictions to representative snapshots extracted from a Ground to I2 trajectory generated with enhanced-sampling MD simulations of apo Abl1 in solution. Specifics about the methodology used to generate this trajectory are described in Appendix 1 of the Supporting Information. Representative results from this comparison are illustrated in Figure 4 and the results of the entire analysis are illustrated in Figures S2-S5. Although we expected to observe a range of conformations, the coverage of the activation loop transition is remarkable and suggests the possibility of using AF2 to sample intermediate states and uncover pathways and mechanisms.

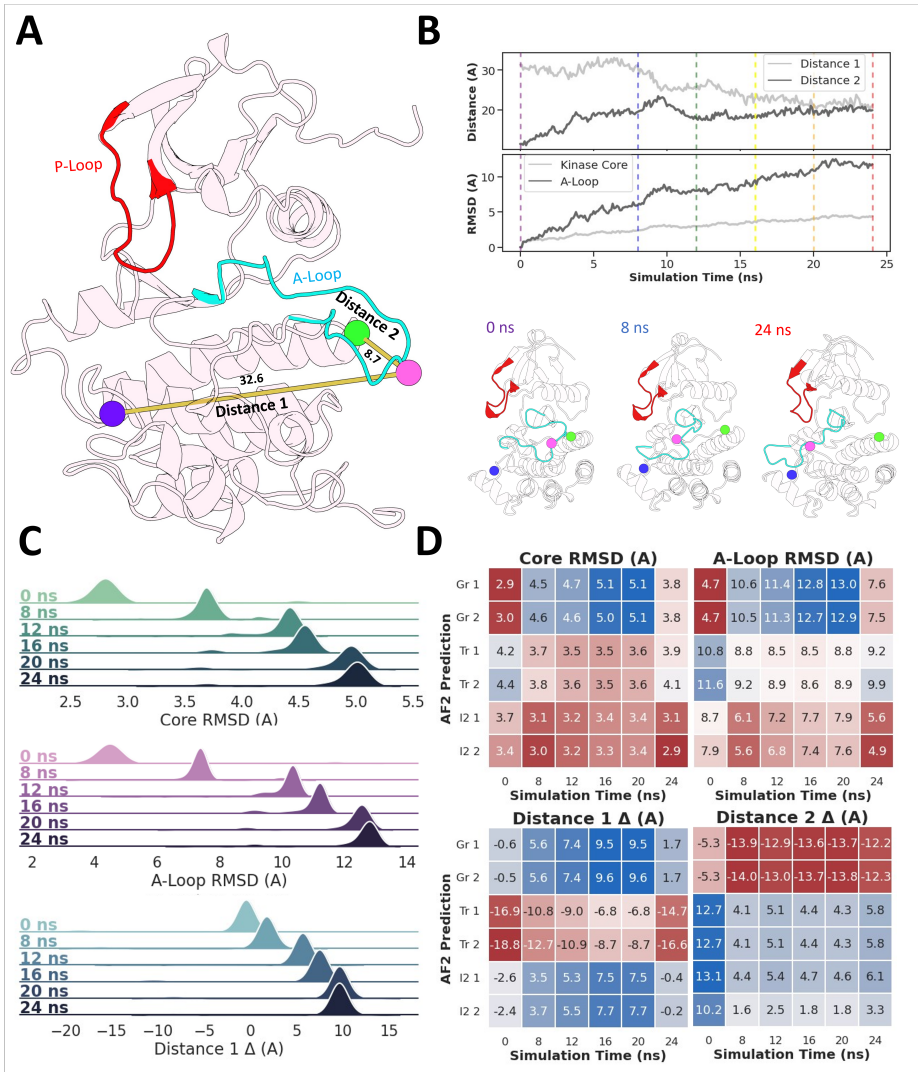

Figure 4: Comparison between the ground to I2 trajectory obtained using enhanced-sampling MD simulations of the Abl1 kinase core and representative AF2 predictions. (A) Structural elements relevant for Abl1 function and that are expected to shift significantly over the ground to I2 transition. (B) Evolution of four relevant structural observables across the trajectory containing the ground to I2 transition. Vertical lines indicate representative snapshots extracted from the trajectory for downstream comparisons with AF2 predictions (top). Three representative snapshots from the MD trajectory at 0 (ground), 8 (transition), and 24 (I2) ns, respectively (bottom). (C) Distribution of three observables relative to the MD snapshots for 160 subsampled AF2 predictions. Core and A-Loop RMSDs are defined as the backbone RMSDs of each AF2 prediction's kinase core (residues 242 to 459) or activation loop (residues 379 to 395) vs. the kinase core or activation loop backbone of the MD snapshot selected at each time point. Distance deltas are defined as the difference in atom pair distances between each AF2 prediction and each respective MD snapshot. Distance 1 corresponds to the distance between the backbone oxygens of E377 and L409, and Distance 2 corresponds to the distance between the backbone oxygens of L409 and G457. (D) Comparison between six representative AF2 predictions and the six MD snapshots described above.

In order to better quantify the effects of each parameter change, we binned each predicted structure into three classes based on the backbone RMSD of relevant structural elements (activation loop, phosphate-binding loop, and C helix) with respect to the backbone of these elements in the ground (active) state, as defined by the lowest-energy structure assignment in the NMR ensemble PDB 6XR6 [22]. Since the RMSD with respect to the ground state of the majority of predictions clustered within 3 Å, we classified predictions with RMSD values greater than 3.5 Å as "not in the ground state." Through this binning, we observed that the 256:512 and 512:1024 values for *max_seq* and *extra_seq* led to predictions in which the ground state is populated 80% and 85% of the time, respectively. Of note, NMR results suggest that the relative state population of the Abl1 kinase core's ground state in solution is 88%, which is in surprisingly good agreement with our AF2 predictions [22]. In contrast with the effects observed by changing MSA composition and length, increasing the number of seeds beyond 128 did not lead to significant changes in the state distribution, suggesting a degree of determinism in the prediction results, presumably stemming from AF2 training and information encoded in the co-evolutionary signal (see Figure 5).

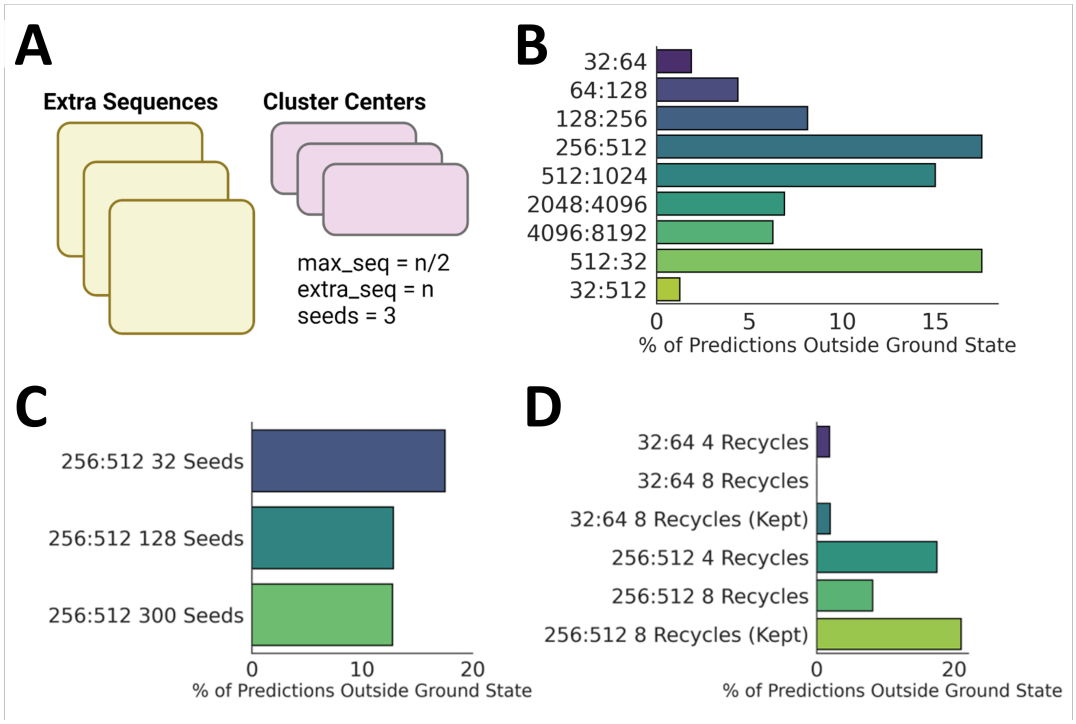

Figure 5: Percent of Abl1 kinase domain conformations predicted to fall outside of the ground state using AF2 based on different MSA clustering parameters (number of sequences selected as cluster centers, number of sequences sampled from the clusters, number of seeds used in the prediction, and number of recycles). (A) Summary of the MSA subsampling and clustering algorithm implemented in AF2. (B) Percent of Abl1 kinase core conformations predicted to fall outside of the ground state using subsampled AF2 based upon the number of seeds used and the amount of recycling performed, and whether recycling intermediates are kept. (C) Impact of changing the number of seeds (each seed corresponds to an independent AF2 prediction); and (D) impact of changing the number of recycles and if structures from recycled iterations are included in the analysis or discarded.

Interestingly, predictions with the *max_seq* and *extra_seq* parameters of 512 and 8, respectively, led to results that are similar to those of the 512:1024 test. Similarly, changing *max_seq* and *extra_seq* to 8:1024 led to results that closely resemble those from the 8:16 test. These results suggest that the *max_seq* parameter is the principal driver of alternative state predictions. This is unsurprising considering the different roles played by each parameter: the MSA of length defined by the *max_seq* argument and formed by the sequences randomly selected as cluster centers is passed to the expensive row/column attention Evoformer track, while the MSA of length *extra_seq* skips it. Due to the increased computational effort needed for featurization and attention, we expect AF2 to distill significantly more coevolutionary signal from the MSA of length *max_seq*, thus changes to *max_seq* will exert greater influence than changes to *extra_seq*.

Finally, we also tested the hypothesis that changing the number of recycles (*n_recycles*) per seed could lead to changes in predicted state distributions by doubling the number of recycles. Interestingly, increasing the number of recycles significantly increases the population of the ground state, suggesting that the recycling stage plays a role in AF2's propensity to generate different conformations. Considering all of the above, we defined our target-specific parameters for all subsequent kinase predictions as follows: *max_seq*: 256, *extra_seq*: 512, *n_recycles*: 4, *n_models*: 5, *n_seeds*: 96. Considering its significant impact on the distribution of predictions, the optimization of the *max_seq* parameter is paramount for successfully obtaining conformational ensembles when running AF2. While 256 cluster centers (defined by *max_seq* = 256) works for Abl1, significantly smaller values are likely to be required for protein systems with less available sequence data.

Predicting the Conformations of Key Members of a Kinase Evolutionary Line Using Subsampled AF2

Given our success predicting the relative populations of the Abl1 kinase core ground and I2 states, we next studied AF2's potential for predicting state distributions without the need for downstream MD simulations. As a basic sanity check, we tested if the Abl1 prediction results were actually the product of AF2 decoding coevolutionary signal pertaining to relative state populations, or just a fortuitous coincidence. Accordingly, we used the same subsampled AF2 protocol outlined in the previous section to predict the relative state populations of the wild-type Src kinase core, which is known to occupy the ground (active) state significantly more frequently than Abl1 [25] (see Figure 6), making it an attractive control case for this test. If our hypothesis regarding the potential of subsampled AF2 is indeed correct, we expect that the method will output significantly more predictions of ground state Src than it did for ground state Abl1, mirroring the experimentally-confirmed relative state distribution differences between the two kinases. Accordingly, we built a large MSA for the Src kinase core (residues 235-497) sequence using the same procedure as described for Abl1, and ran our implementation of subsampled AF2 with it as an input. We then measured the relative population of the Src kinase core ground and I2 states.

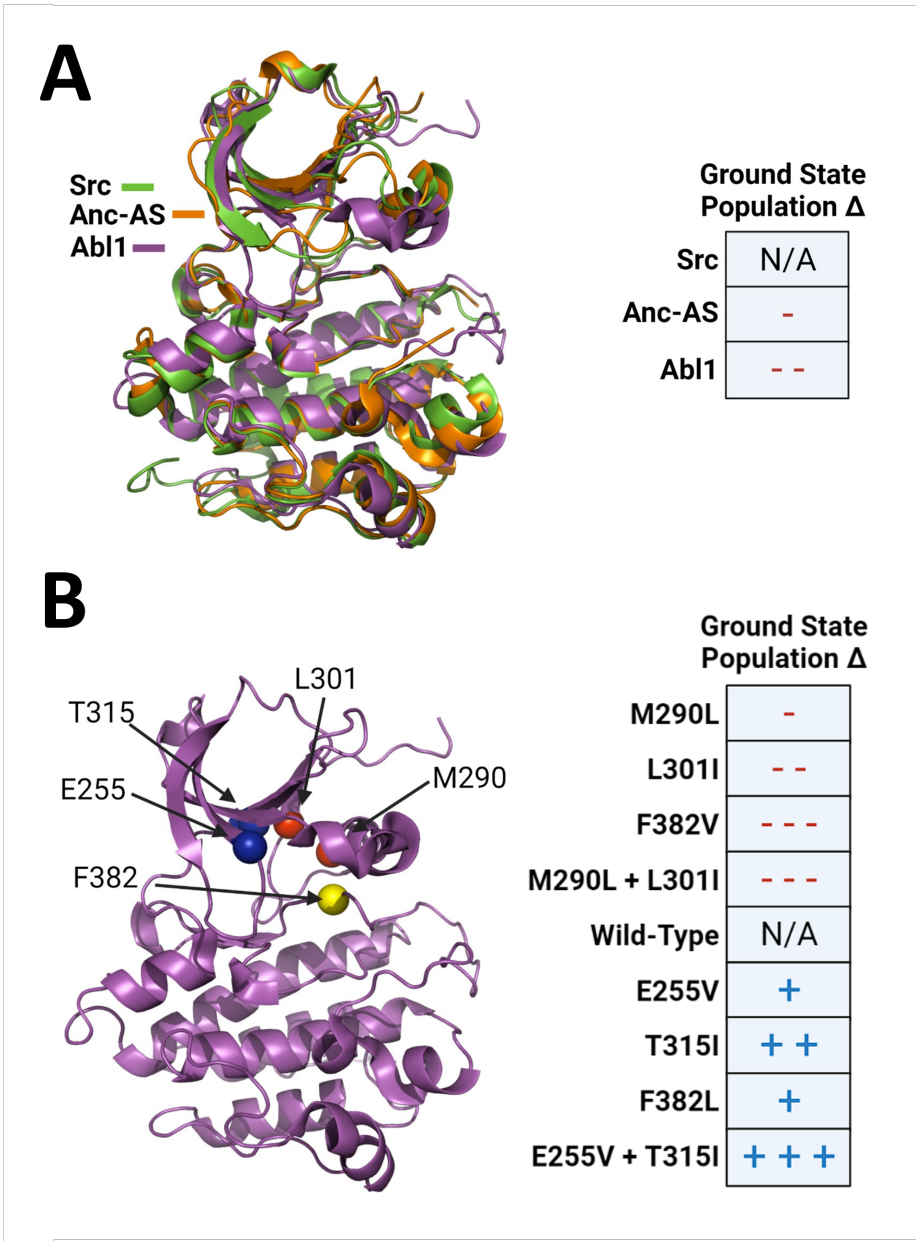

Figure 6: Summary of experimental observations regarding the relative state populations of kinase cores along Abl1's evolutionary line and in the Abl1 allelic series studied here. (A) Abl1, Anc-AS, and Src are part of the same evolutionary line and share significant sequence and structural homology, but Abl1 occupies the ground state significantly less frequently than Src, and Anc-AS occupies it slightly less than Src, suggesting that Abl1 evolution has directed it to be more flexible than Src [25]. (B) A series of point mutations in wild-type Abl1 are known to increase or decrease the relative population of the enzyme's active (ground) state [22]. Residues with mutations known to increase the population of the ground state are shown as blue spheres in the Abl1 structure on the left side, while those with mutations known to decrease the ground state population are shown as red spheres. The yellow sphere denotes phenylalanine 382, part of the DFG motif, which can be mutated to valine to increase the ground state population, or to leucine to reduce it. While the effects of the E255V + T315I mutation on the ground state population were not directly reported in the literature, we used its cumulative effect on kinase activity and Imatinib binding [35] to infer an increase in the ground state population.

Crucially, the vast majority of Src kinase core predictions from subsampled AF2 were found to be in the ground state, with a predicted relative state population of 97% compared to 89% for Abl1, as summarized in Figure 7. Interestingly, none of the Src predictions were found to be in the I2 state, although the enzyme is known to infrequently occupy this conformation. This suggests a resolution limitation in using AF2 to predict relative state populations: conformations with very low occupancy such as I2 in Src might be missed by the algorithm in its current implementation. A potential cause of this limitation is the fact that AF2's prediction models were extensively trained on structures from the Protein Data Bank (PDB), in which most of the structures of Src and its orthologs are in the ground state [36].

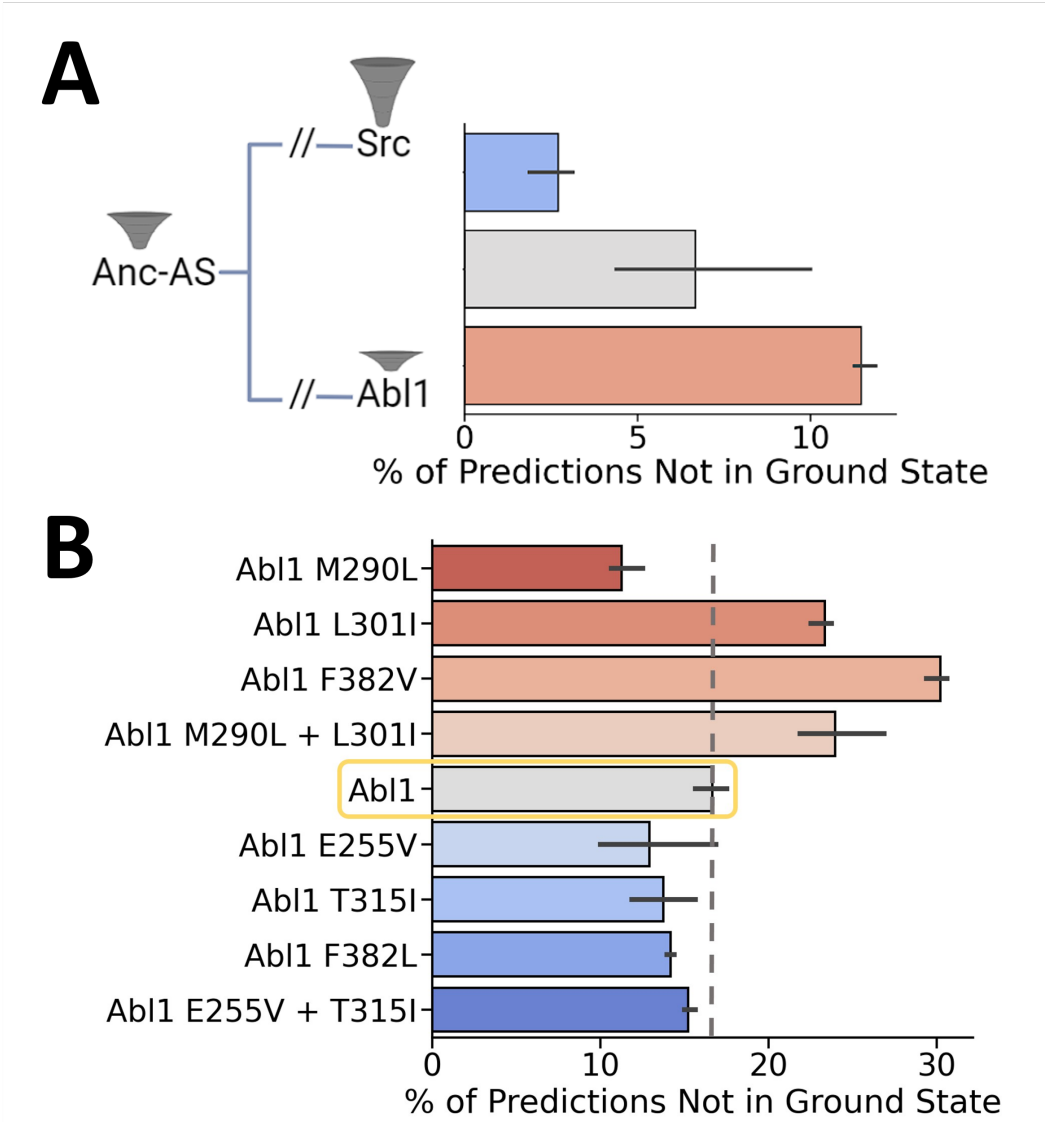

Figure 7: Subsampled AF2 predictions of the percent of conformations not in the ground state for proteins along the Src to Abl1 evolutionary pathway and Abl1 resistance-causing mutations. (A) Representation of the evolutionary history connecting Abl1, Anc-As, and Src, and the relative population of the active/ground state for each kinase core [25]. Right: Percentage of predictions that fall outside of the ground state for each test case. (B) Results of subsampled AF2 for the Abl1 kinase domain allelic series. Data are colored based on the expected relative state populations of each mutation: red-tinted bars represent mutations that are known to decrease the ground state population, while blue-tinted bars represent mutations with the opposite effect [22]. Error bars were calculated from three independent sets of predictions, each with 32 unique seeds and 5 models, totaling 160 predictions per replicate.

Despite resolution caveats, subsampled AF2 correctly predicted the stark contrast in conformational distributions between the Abl1 and Src kinase cores, lending credence to its promise as a high-throughput method for predicting relative state populations. To further probe this capacity,

we applied subsampled AF2 to make predictions for the Anc-AS kinase core (residues 1-263), an Abl1 ancestor with a known conformational state distribution and dynamics [25], and compared the results to the Abl1 and Src cases. The sequences of Abl1, Anc-AS, and Src used to generate the MSAs and used as target sequences for AF2 are summarized in Figure S1. While there are other Abl1 ancestors that could be used in this test, there are experimental results for Anc-AS, including a deposited structure in the Protein Databank (4UEU), justifying its choice. For the subsampled AF2 predictions to be considered accurate, the relative population of the ground state in the Anc-AS predictions should be in-between the populations of the same state for the Abl1 and Src predictions, as observed in experimental results. Once again, subsampled AF2 correctly replicated experiments (Figure 7), as Anc-AS was predicted to be in the ground state 93% of the time, in-between the frequencies predicted for Src (97%) and Abl1 (89%).

Predicting the Conformations of a Kinase Allelic Series with AF2

While the results we obtained for the Abl1 to Src evolutionary line are promising, an even more impactful application would be using it to predict how such populations change across an allelic series with single substitution resolution. This is because many point mutations in proteins are thought to lead to different phenotypes - such as drug resistance - by changing conformational landscapes and relative state populations.

To measure the capacity of subsampled AF2 to fill this niche, we repeated our predictions using a series of Abl1 single and double mutants with well-characterized and significant effects on the relative populations of the ground and I2 states, and contrasted the results with those obtained from the wild-type prediction. Specifically, we tested the method on four mutations that are expected to decrease the population of the ground state (M290L, L301I, F382V, M290L + L301I), and four mutations that are expected to increase it (E255V, T315I, F382L, E255V + T315I). The mutations tested, their locations in Abl1, and their expected effects on ground state populations are summarized in Figure 6 (see Table S2 for the full list of mutations tested and their expected outcomes in the relative state populations of the Abl1 kinase core). Importantly, the length of the kinase core sequence used as input for this test (229-515) differs slightly from the previous (235-497) so as to match the length of the constructs tested in the literature. This difference caused a slight variation in the wild-type Abl1 ground state predictions (84% vs. 89%).

AF2 predictions for this allelic series are summarized in Figure 7. Strikingly, subsampled AF2 correctly predicted a change in the relative state population and its direction in over 80% of the tested cases. Although promising, these results are not without significant caveats. First: the M290L predictions are inaccurate. Specifically, the effects of the mutation on the ground state population are predicted to be the opposite of those seen in experiments. Second: the prediction accuracy only applies in a qualitative sense, as double mutations that are known to significantly increase the ground state population such as M290L + L301I are predicted by subsampled AF2 to increase it only slightly more than single mutations such as M301I, which is known to cause a more subtle increase. We believe that this inaccuracy is a direct consequence of the incorrect M290L prediction, which should also result in the underestimation of the effects of double mutations including M290L. Furthermore, the statistical significance of the results is reduced for the benchmarks where the mutations are known to reduce ground-state populations. We hypothesize that AF2 performed better at predicting decreases in the ground state population because its predictions reflect an abundance of Abl1 I2 structures in its training set.

Predicting GMCSF Conformations with AF2

Considering the success of our Abl1 predictions, we sought to test if the accuracy of these predictions was contingent on the wealth of kinase sequence data, or if we could obtain similar results with significantly less sequence data. To do so, we repeated our prediction workflow but using the sequence of the human granulocyte-macrophage colony-stimulating factor (GMCSF). GMCSF is a 14 kDa monomeric glycoprotein that plays a central role in innate immunity, stimulating a variety of cells in response to pathogens [37]. In contrast to that for kinases, literature regarding GMCSF's structure and dynamics is sparse, and sequence data for homologs or orthologs is orders of magnitude less abundant. Crucially, the MSA built for the wild-type human GMCSF sequence is only 112 sequences long, while that for human Abl1 kinase is over 600,000 sequences using the same parameters and methods (see Figure 8). This stark contrast represents a useful opportunity to study how the accuracy of subsampled AF2 is modulated by the availability of sequence data.

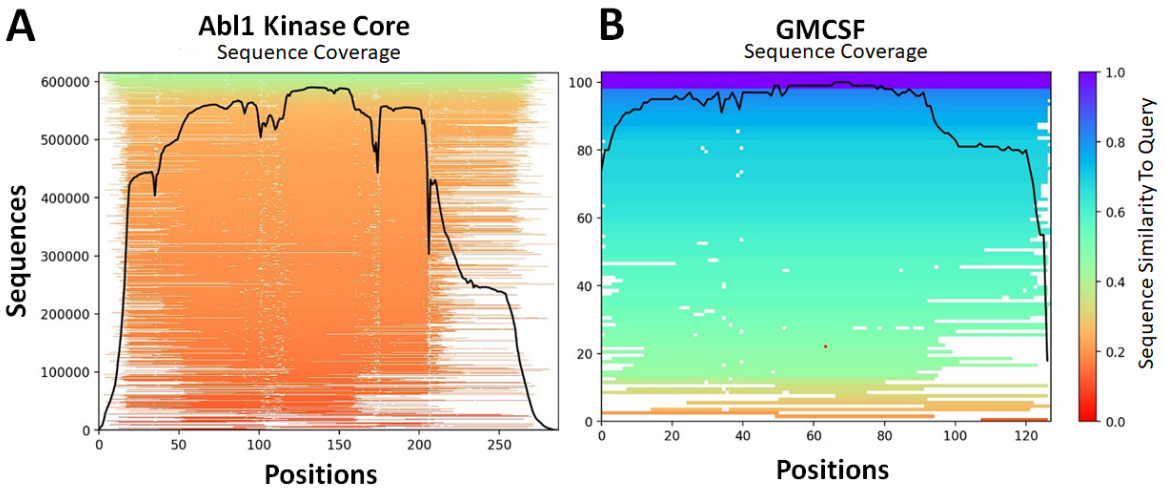

Figure 8: Comparison between MSA length and per-position coverage for (A) Abl1 and (B) GMCSF multiple sequence alignments used as input in AF2.

Previous NMR results have shown that the dynamics of GMCSF's N-terminal helix A are involved in the binding of heparin and other charged molecules [38]. Specifically, it is thought that the relative position of helix A with respect to the center of mass of the protein is flexible, usually being tightly packed against helix C through π-π interactions among histidines 15, 83, and 87. Experimental results suggest that this closed configuration is the most stable GMCSF conformation when packed as a crystal [39]. Thus, for clarity, this closed conformation will be henceforth referred to as the ground state. It is thought that the π-π interactions are eventually weakened either through intrinsic breathing motions and resultant interactions with the solvent or via induction by other molecules, and helix A is thought to move away from the GMCSF core. This opening motion should expose part of the GMCSF core to solvent and creates a V-shaped groove that is thought to be the binding site of heparin and other immune system modulators [38]. This binding-competent GMCSF state will be referred to as the open conformation. Importantly, GMCSF binds heparin much more strongly in an acidic environment, suggesting that helix A dynamics are protonation-dependent [38].

Further NMR experiments (Figure S6) have shown that point mutations in the aforementioned histidine triad lead to significant chemical shift perturbations, similar to those caused by reductions in pH [38], hinting at increased occupancy of the open state or other unknown conformations. Importantly, the amplitudes of the changes to the backbone dynamics caused by these mutations vary significantly depending on which histidine of the triad is mutated. Specifically, mutating H15 or H83 leads to more pronounced changes in chemical shifts than mutating H87 (see Figure 9). This is expected as H15 and H83 are significantly more buried than H87, meaning that mutations at these positions are likely to be harder for the GMCSF backbone to accommodate. Conversely, changes in the mostly solvent-exposed H87 usually lead to modest shifts in backbone dynamics, suggesting that mutations in this codon are more readily tolerated. Finally, mutations H83Y and H83R lead to larger scale conformational changes (inferred from broadened peaks) than any other mutation in the test set, and considerably more than the H83N mutation, suggesting that specific substitutions at each position induce more significant changes in the GMCSF backbone dynamics (see Figure S6 for a detailed analysis of all mutant CSPs and broadened peaks). These mutations are therefore useful for benchmarking as they represent a tiered challenge of predicting which specific amino acid substitution will lead to the largest changes in GMCSF structure.

Accordingly, we sought to test if our subsampled AF2 method was capable of predicting the expected changes in the backbone dynamics of each mutant with respect to the flexibility of helix A and other structural elements in GMCSF. Our predictions were considered accurate if they matched NMR experiments we performed along two axes: if they predicted mutations to H15 and H83 to provoke larger changes in the distribution of conformations than mutations to H87; and if they correctly predicted the most significant mutations in H83 or H87 as evidenced by CSPs or broadened peaks. After building the master MSA using the wild-type human GMCSF sequence as a query and the JackHMMR method [31], we determined the max_seq and *extra_seq* parameters that led to the greatest diversity of GMCSF conformations while still maintaining the ground state as the most frequent state. As shown in Figure S7, *max_seq*=4 and *extra_seq*=8 were found to maximize diversity. Employing these parameters while keeping all others the same from the Abl1 tests, we used AF2 to predict wild-type GMCSF structures. We then repeated every previous step minus the parameter optimization for each GMCSF mutant and quantified the ground state populations from the resulting set of predictions.

To assess how mutations affected the relative populations of different GMCSF conformations, we measured the RMSD of specific backbone atomic positions of each predicted GMCSF structure with respect to the AF2 prediction that was most similar to the wild-type crystallographic reference (PDB 1CSG) [39]. The regions mentioned above are described in Figure 9 and correspond to GMCSF elements known to show significant perturbations in NMR experiments. The results of this analysis for the two regions with the most significant changes in our NMR results are described in Figure 9. Additionally, we also measured the distance between the alpha carbons of H15 and H83 as well as the overall backbone RMSD with respect to the ground state reference for each prediction with the goal of identifying predictions that led to unexpected conformations or partially or completely unfolded structures. These additional measurements as well as three examples of unexpected structures are illustrated in Figure S8.

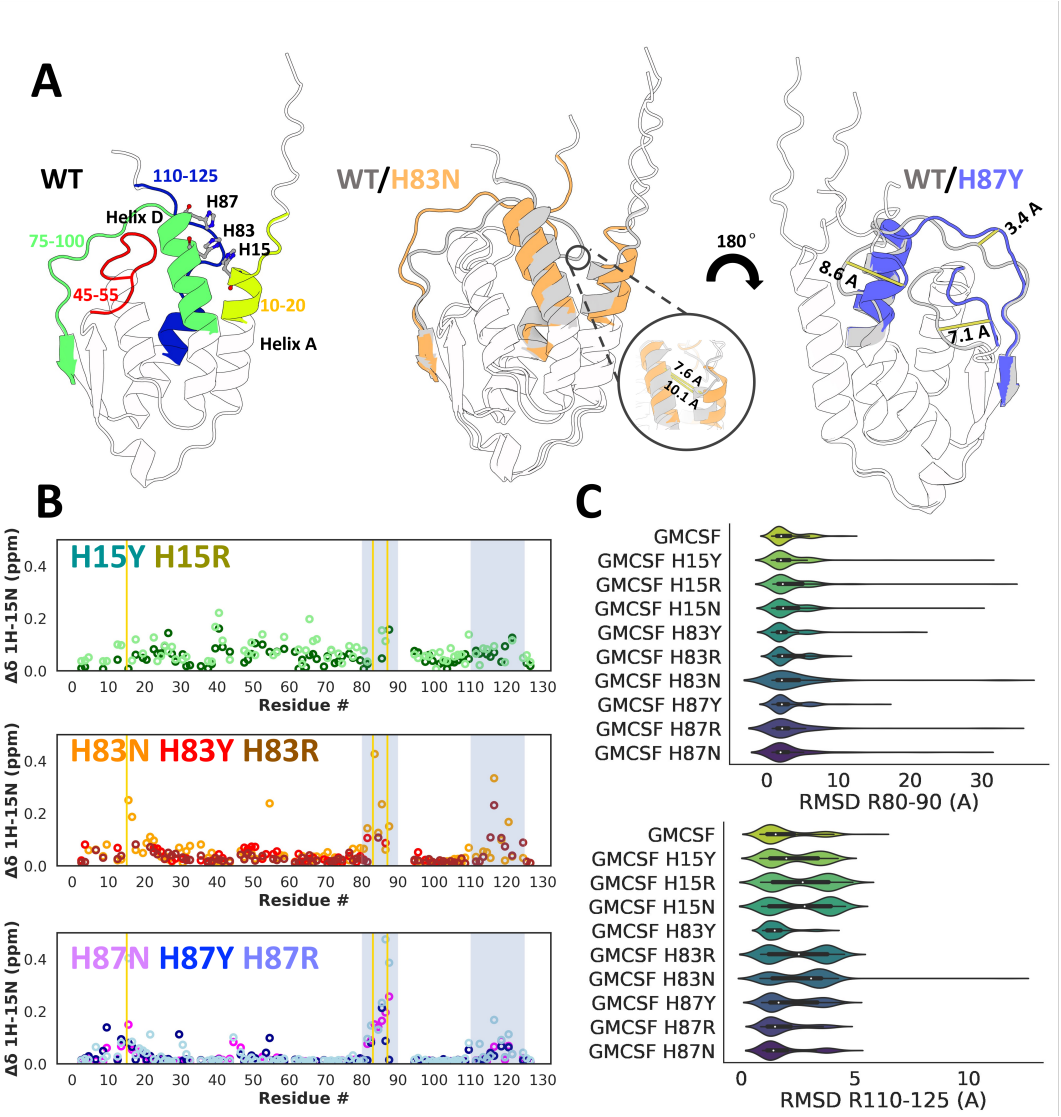

Figure 9: Subsampled AF2 results for GMCSF mutations. (A) Left: Annotated wild-type GMCSF structure in the ground (closed) conformation as predicted by AF2. Each colored element represents a region of putative high mobility according to chemical shift perturbations (CSPs) in the histidine triad mutants. Middle: Superposition of GMCSF in the ground conformation and a putative open conformation whose population is enriched in the AF2 predictions for H83 and H87 mutants, especially H83N. The inset compares distances between codon positions 15 and 83 in both conformations. Right: Superposition of GMCSF in the ground conformation and a tilted helix D conformation whose population is enriched in the AF2 predictions for H87 mutants (especially H87Y). Distances between flexible elements for both conformations are shown as yellow bars. (B) Per-residue chemical shift perturbations for GMCSF mutants, separated by codon position. Gold vertical bars represent the three sites of mutation, and silver shaded areas correspond to the residue indices plotted on the x-axis in C. (C) Distributions of root mean square deviations of atomic positions for the backbone atoms of residues 83-100 (top) and 110-125 (bottom) superimposed upon the GMCSF wild-type ground state structure. The distributions span 480 independent predictions (32 unique seeds * 5 models * 3 replicates.)

Despite the paucity of sequence data, our subsampled AF2 method correctly identified residues H15 and H83 as the most sensitive to mutation, as the range of the distribution of RMSDs of residues 80-90 and 110-125 is significantly larger for most of the mutations tested at both of these sites. Conversely and in-line with the experimental results, mutations to H87 led to significant changes in the distribution of the residue 80-90 RMSDs, but to comparatively modest changes for the RMSDs of residues 110-125. In addition to accurately predicting the differences in amplitude of backbone dynamics between mutants H15/83 and H87, subsampled AF2 correctly estimated the significant impact of mutations H83R and H83N on C-terminal dynamics, while also accurately predicting H83N, H83Y, and all three H87 mutants to have a large impact on the R80-90 RMSD distribution.

As with the Abl1 allelic series example, our method did not achieve perfect accuracy. Specifically, it failed to predict the significant changes in GMCSF's N-terminus dynamics associated with the H83N mutation (Figure S9). Additionally, our method failed at replicating the significant changes in dynamics of residues 80-90 for the H83Y mutations, which should be of larger amplitude than those for H83N and closer to those for H83R. Similarly, we expect the H87Y mutation to induce a larger amplitude of conformational changes in the region composed of residues 80-90 than the H87N mutation due to the enrichment of broadened peaks in that area in the H87Y NMR results, but AF2 predicts H87Y to be significantly less impactful than H87N in that context. In summary, AF2 performed exceptionally well at predicting subtle conformational changes in specific loops for most of our test cases but failed at replicating the comparatively larger expected effects of mutation H83Y.

Beyond these limitations, we also observed a set of novel conformations that are significantly different than both the ground and open states. Clustering the results by the structural features described in Figure S9 reveals that one alternative conformation in particular is significantly enriched, especially in predictions of the H83 mutants. In this alternative conformation, henceforth dubbed A1, the C helix has switched places with the B helix, placing H83 and H15 more than 10 Å away (Figure S8). In this state, helix B occupies the groove to which heparin is thought to bind, which could be a mechanism for self-inhibition. Although further NMR experiments and simulations are necessary to confirm if this in fact a metastable GMCSF conformation with physiological functions, the fact that it has structural rearrangements with amplitudes that seem to be comparable to those of the H83 mutations (as evidenced by the H3 NMR chemical shift perturbations and broadened peaks [Figure S6]) and that this conformation was predicted more often for H83 mutants highlights the promise of using subsampled AF2 to help understand NMR results and derive novel hypotheses or mechanisms.

Finally, many of the alternative conformations not discussed above were found to be partially unfolded. Although the frequency of these was extremely low compared to predictions binned to the ground, open, or A1 states, the fact that they existed at all was surprising as running subsampled AF2 for the Abl1 test case led to no unfolded predictions whatsoever. One plausible hypothesis beyond the destabilizing effects of the mutations that may explain this observation is that the substantially shorter length of the GMCSF MSA relative to Abl1's drastically increases the uncertainty in AF2's predictions leading to a wider range of predictions.

Discussion

In this work, we successfully modified AF2's inputs and parameters to predict the conformational ensembles and relative state populations of two proteins, Abl1 kinase and human granulocyte-macrophase colony-stimulating factor (GMCSF), that have very different amounts of known sequence homology. In studying these proteins, we focused on how well our subsampled version of AF2 can reproduce the effects of evolution and mutations on relative state populations.

Our subsampled AF2 implementation predicted Abl1, a kinase for which there is an abundance of sequence homology, to occupy its ground state nine times more frequently than its I2 state, consistent with NMR observations. More importantly, subsampled AF2 correctly predicted the relative state populations of two evolutionarily-related kinases, Anc-AS and Src kinase, leading to the correct correlation between their evolutionary distances and relative state populations. Using single and double Abl1 mutants with known effects on the relative state populations of Abl1's ground and I2 states, we moreover found that AF2 yielded surprisingly accurate state populations even for single mutations. This is best evidenced in the results for the F382 substitutions: phenylalanine 382 is a codon known to slightly increase the population of Abl1's ground state if mutated to leucine, or significantly reduce it if mutated to valine, an observation that is accurately replicated by our prediction method. Furthermore, an unexpected but remarkable feature illustrated by all of the Abl1 test cases is the capacity of subsampled AF to predict intermediate conformations spanning the transition from the ground state to I2 which closely match intermediate conformations obtained from more costly MD simulations.

Finally, we also obtained overwhelmingly accurate backbone dynamics predictions for GMCSF. Despite the lack of GMCSF sequence data which led to an input MSA of fewer than 120 sequences (versus more than 600,000 for Abl1), predictions from subsampled AF2 were distributed with frequencies that matched the expected relative state populations for most GMCSF variants. These results suggest that AF's prediction engine is robust enough to decode population signals from relatively scarce data.

The results we have obtained for the two test cases discussed above confirm AF2's potential for predicting conformational ensembles and, more importantly, demonstrate entirely novel applications of AF. In particular, we show that optimization of subsampling parameters allows users to accurately predict relative state populations and how they change in response to mutations with single-substitution resolution. This feature is a significant advance over the previous state of the art, as it facilitates high-throughput applications such as the design of fold-switching proteins, inference of mechanisms of acquired drug resistance, and reweighting of binding affinity predictions. Additionally, our workflow generates conformational intermediates, which has direct implications for discovering drugs that bind alternative conformations and for improving our understanding of biophysics in general. In addition to these immediate practical applications, the more fundamental observation that AF2 is decoding information regarding conformational distributions from sequence data alone points to many other potential unforeseen uses of AF2 that can result in further methodological advances and discoveries.

This said, our AF2 pipeline is not perfect; our workflow inaccurately predicted the M290L mutation to significantly decrease the ground state population when that mutation is known experimentally to have the opposite effect. Interestingly, AF2 predicted the double mutation (M290L+L301I) to increase the ground state population more than the L301I mutation alone. As AF2 correctly predicted the relative state populations of Src, which differs from Abl1 by dozens of mutations, this

suggests a potential more general trend that AF2 is more accurate in predicting the effect of multiple mutations than that of a single mutation. Furthermore, our pipeline also struggled to correctly predict all of the structural elements expected to differ in each conformation. Specifically, while AF2 predicted an ensemble of activation loop conformations, a few of the inactive-like predictions (activation loop closed) contained structural elements that are typically thought to belong to the active state, such as the position of the C-helix and the dihedral angles of the D381 and F382 side-chains. Moreover, even when our pipeline predicts a change in a structural element from its configuration in the ground (active) state, the amplitude of that change may not correspond to that seen experimentally. For example, the side-chain dihedral angle of the D381 residue in Abl1 ranges from -130 (active state) to 40 (inactive state) degrees in structures in the Protein Data Bank (PDB) [22], but ranges from -130 to -90 degrees in AF2 predictions. Last but not least, while our method is capable of predicting whether mutations will increase or decrease state populations, it remains to be shown whether it is capable of directly predicting Boltzmann ratios of states in a quantitative fashion.

Despite these limitations, we believe that our results showcase promising, unexpected applications of AF2. Using the high-throughput pipeline we developed or one inspired by it could save significant time when filtering large allelic series to identify mutations with significant impacts on conformation. It could also be used as a prediction engine for classifying arrays of drug-resistant mutations based on their shared effects on the stability of a given conformation, thus facilitating polypharmacology. Finally, as demonstrated by AF2's Abl1 activation loop predictions, our method could also be useful for identifying previously unknown, potentially metastable states of known proteins. While it remains to understand exactly how AF2 is gathering and interpreting signals about state populations from sequence data, we hope that our work will motivate many future investigations.

Methods

Protein Structure Visualization

To visualize the predicted structures and trajectories and calculate descriptors such as distances between atoms, RMSD to reference, dihedral angles, etc., we used PyMol (version 2.4.1) (Schrodinger LLC, 2020).

Multiple Sequence Alignments

The JackHMMR algorithm [31] was used to generate multiple sequence alignments (MSAs) for each protein of interest by querying sequences from the UniProt90 [32], Mgnify [34], and BFD [33] databases.

Structure Prediction

We used AlphaFold 2 [11] to predict protein structures with the MSAs generated in step 1 by subsampling and clustering the MSAs and using random sequences as cluster centers to generate a distribution of conformational states.

Molecular Dynamics Simulations

Enhanced-sampling molecular dynamics simulations were conducted on the proteins described using the WESTPA2 [40] implementation within the OpenMM molecular simulation engine [41]. For an extensive description of the MD methodology employed in this study, please refer to the Supplementary Materials.

Miscellaneous and Data Visualization

Data plotting and visualization were performed using seaborn (version 0.11.1). Data were plotted as mean ± confidence intervals.

Acknowledgements

The authors thank Jacob Rosenstein, F. Marty Ytreberg, Jagdish S. Patel, Winston Y. Li, Haibo Li, and Kyle Lam for insightful conversations and support. G.M.d.S. and B.M.R. were supported in part by the National Science Foundation under Grant No. 2027108. J.C. and G.L.'s contributions were supported by NIH Grant R01 GM144451. The computational aspects of this research were conducted using computational resources and services at the Center for Computation and Visualization, Brown University.

Author Contributions

G.M.d.S. performed the simulations and analyzed the sequence and AlphaFold 2 data. J.C. performed NMR experiments and contributed the NMR data. D.D., G.L., and B.M.R. provided direction and oversight. G.M.d.S. and B.M.R. drafted the manuscript. All authors edited the manuscript.

Competing Interests

There are no competing interests.

Data Availability

The datasets from this study are available from the authors upon request.

References

1. Robertson, A. D. & Murphy, K. P. Protein Structure and the Energetics of Protein Stability. *Chemical Reviews* **97.** PMID: 11851450, 1251–1268. eprint: `https://doi.org/10.1021/cr960383c` (1997).
2. Bai, X.-c., McMullan, G. & Scheres, S. H. How cryo-EM is revolutionizing structural biology. *Trends in Biochemical Sciences* **40,** 49–57 (Jan. 2015).

3. Su, X.-D. *et al.* Protein crystallography from the perspective of technology developments. *Crystallography Reviews* **21,** 122–153 (Dec. 2014).

4. Wüthrich, K. Protein structure determination in solution by NMR spectroscopy. *Journal of Biological Chemistry* **265,** 22059–22062 (Dec. 1990).

5. Slabinski, L. *et al.* The challenge of protein structure determination-lessons from structural genomics. *Protein Science* **16,** 2472–2482 (Nov. 2007).

6. Leach, A. & Thomas, P. in *Comprehensive Medicinal Chemistry III* 120–144 (Elsevier, 2017).

7. Ołdziej, S. *et al.* Physics-based protein-structure prediction using a hierarchical protocol based on the UNRES force field: Assessment in two blind tests. *Proceedings of the National Academy of Sciences* **102,** 7547–7552 (May 2005).

8. Jothi, A. Principles, Challenges and Advances in ab initio Protein Structure Prediction. *Protein and Peptide Letters* **19,** 1194–1204 (Sept. 2012).

9. Torrisi, M., Pollastri, G. & Le, Q. Deep learning methods in protein structure prediction. *Computational and Structural Biotechnology Journal* **18,** 1301–1310 (2020).

10. AlQuraishi, M. Machine learning in protein structure prediction. *Current Opinion in Chemical Biology* **65,** 1–8 (Dec. 2021).

11. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596,** 583–589 (Aug. 2021).

12. Tunyasuvunakool, K. *et al.* Highly accurate protein structure prediction for the human proteome. *Nature* **596,** 590–596 (Aug. 2021).

13. Baek, M. *et al.* Efficient and accurate prediction of protein structure using RoseTTAFold2 (May 2023).

14. Lin, Z. *et al.* Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* **379,** 1123–1130 (Mar. 2023).

15. Roney, J. P. & Ovchinnikov, S. State-of-the-Art Estimation of Protein Model Accuracy Using AlphaFold. *Physical Review Letters* **129** (Nov. 2022).

16. Callaway, E. What's next for AlphaFold and the AI protein-folding revolution. *Nature* **604,** 234–238 (Apr. 2022).

17. Chakravarty, D. & Porter, L. L. AlphaFold 2 fails to predict protein fold switching. *Protein Science* **31** (May 2022).

18. Porter, L. L. & Looger, L. L. Extant fold-switching proteins are widespread. *Proceedings of the National Academy of Sciences* **115,** 5968–5973 (May 2018).

19. Bryan, P. N. & Orban, J. Proteins that switch folds. *Current Opinion in Structural Biology* **20,** 482–488 (Aug. 2010).

20. Kim, A. K. & Porter, L. L. Functional and Regulatory Roles of Fold-Switching Proteins. *Structure* **29,** 6–14 (Jan. 2021).

21. Borkakoti, N. & Thornton, J. M. AlphaFold2 protein structure prediction: Implications for drug discovery. *Current Opinion in Structural Biology* **78,** 102526 (Feb. 2023).

22. Xie, T., Saleh, T., Rossi, P. & Kalodimos, C. G. Conformational states dynamically populated by a kinase determine its function. *Science* **370** (Oct. 2020).

23. Michielssens, S., de Groot, B. L. & Grubmüller, H. Binding Affinities Controlled by Shifting Conformational Equilibria: Opportunities and Limitations. *Biophysical Journal* **108,** 2585–2590 (May 2015).

24. Iqbal, N. & Iqbal, N. Imatinib: A Breakthrough of Targeted Therapy in Cancer. *Chemotherapy Research and Practice* **2014,** 1–9 (May 2014).

25. Wilson, C. *et al.* Using ancient protein kinases to unravel a modern cancer drug's mechanism. *Science* **347,** 882–886 (Feb. 2015).

26. Wayment-Steele, H. K., Ovchinnikov, S., Colwell, L. & Kern, D. Prediction of multiple conformational states by combining sequence clustering with AlphaFold2 (Oct. 2022).

27. Vani, B. P., Aranganathan, A., Wang, D. & Tiwary, P. AlphaFold2-RAVE: From Sequence to Boltzmann Ranking. *Journal of Chemical Theory and Computation* (May 2023).

28. Meller, A., Bhakat, S., Solieva, S. & Bowman, G. R. Accelerating Cryptic Pocket Discovery Using AlphaFold. *Journal of Chemical Theory and Computation* (Mar. 2023).

29. Stein, R. A. & Mchaourab, H. S. SPEACH_AF: Sampling protein ensembles and conformational heterogeneity with Alphafold2. *PLOS Computational Biology* **18** (ed Deane, C. M.) e1010483 (Aug. 2022).

30. Del Alamo, D., Sala, D., Mchaourab, H. S. & Meiler, J. Sampling alternative conformational states of transporters and receptors with AlphaFold2. *eLife* **11** (eds Robertson, J. L., Swartz, K. J. & Robertson, J. L.) Publisher: eLife Sciences Publications, Ltd, e75751 (Mar. 2022).

31. Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research* **39,** W29–W37 (May 2011).

32. Suzek, B. E., Wang, Y., Huang, H., McGarvey, P. B. & and, C. H. W. UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics* **31,** 926–932 (Nov. 2014).

33. Steinegger, M. & Söding, J. Clustering huge protein sequence sets in linear time. *Nature Communications* **9** (June 2018).

34. Richardson, L. *et al.* MGnify: the microbiome sequence data analysis resource in 2023. *Nucleic Acids Research* **51,** D753–D759 (Dec. 2022).

35. Hoemberger, M., Pitsawong, W. & Kern, D. Cumulative mechanism of several major imatinib-resistant mutations in Abl kinase. *Proceedings of the National Academy of Sciences* **117,** 19221–19227 (July 2020).

36. Modi, V. & Dunbrack, R. L. Kincore: a web resource for structural classification of protein kinases and their inhibitors. *Nucleic Acids Research* **50,** D654–D664 (Oct. 2021).

37. Lee, K. M., Achuthan, A. A. & Hamilton, J. A. GM-CSF: A Promising Target in Inflammation and Autoimmunity. *ImmunoTargets and Therapy* **Volume 9,** 225–240 (Oct. 2020).

38. Cui, J. Y. *et al.* Mapping the Structural and Dynamic Determinants of pH-Sensitive Heparin Binding to Granulocyte Macrophage Colony Stimulating Factor. *Biochemistry* **59,** 3541–3553 (Sept. 2020).

39. Walter, M. R. *et al.* Three-dimensional structure of recombinant human granulocyte-macrophage colony-stimulating factor. *Journal of Molecular Biology* **224,** 1075–1085 (Apr. 1992).

40. Bogetti, A. T. *et al.* A Suite of Advanced Tutorials for the WESTPA 2.0 Rare-Events Sampling Software [Article v2.0]. *Living Journal of Computational Molecular Science* **5** (2022).

41. Eastman, P. *et al.* OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLOS Computational Biology* **13** (ed Gentleman, R.) e1005659 (July 2017).