Empowering AlphaFold2 for protein conformation selective drug discovery with AlphaFold2-RAVE

Xinyu Gu,[†] Akashnathan Aranganathan,[†,¶] and Pratyush Tiwary[*,†,‡,§]

[†]*Institute for Physical Science and Technology, University of Maryland, College Park, Maryland 20742, USA*

[‡]*Department of Chemistry and Biochemistry, University of Maryland, College Park 20742, USA*

[¶]*Biophysics Program, University of Maryland, College Park 20742, USA*

[§]*University of Maryland Institute for Health Computing, Bethesda, United States*

E-mail: ptiwary@umd.edu

Abstract

Small molecule drug design hinges on obtaining co-crystallized ligand-protein structures. Despite AlphaFold2's strides in protein native structure prediction, its focus on apo structures overlooks ligands and associated holo structures. Moreover, designing selective drugs often benefits from the targeting of diverse metastable conformations. Therefore, direct application of AlphaFold2 models in virtual screening and drug discovery remains tentative. Here, we demonstrate an AlphaFold2 based framework combined with all-atom enhanced sampling molecular dynamics and induced fit docking, named AF2RAVE-Glide, to conduct computational model based small molecule binding of metastable protein kinase conformations, initiated from protein sequences. We demonstrate the AF2RAVE-Glide workflow on three different protein kinases and their type I and II inhibitors, with special emphasis on binding of known type II kinase inhibitors which target the metastable classical DFG-out state. These states are not

easy to sample from AlphaFold2. Here we demonstrate how with AF2RAVE these metastable conformations can be sampled for different kinases with high enough accuracy to enable subsequent docking of known type II kinase inhibitors with more than 50% success rates across docking calculations. We believe the protocol should be deployable for other kinases and more proteins generally.

Introduction

Despite the groundbreaking impact of Alphafold2 (AF2)[1,2] on the computational prediction of the ligand-free protein native apo structures, it appears that the determination of high quality crystal or cryo-EM ligands bound holo structures remains irreplaceable in the field of structure based drug design. When ligands bind, residues within the protein pockets may adjust their side-chain rotamer configurations to optimize contacts with ligands, a phenomenon known as the induced-fit effect. Furthermore, thermodynamic fluctuations induce protein dynamics and structural flexibility, leading to rearrangements of side-chains and even large-scale backbone movements that can reveal cryptic pockets.[3] These metastable conformations and associated cryptic pockets can be stabilized upon binding to specific ligands. Moreover, due to the similarity of native structures among protein homologs, it has been widely believed that ligands targeting highly diverse metastable conformations should result in better selectivity.[4] It is thus highly desirable to account for metastable protein conformations or states, instead of investigating native states only. Several AF2-based techniques, including reduced multiple sequence alignment (rMSA) AF2 (or MSA subsampling AF2),[5–7] AF2-cluster[8] and AlphaFlow,[9] have been devised to generate distinct decoy structures from native states. However, the suitability of these decoys for subsequent docking and virtual screening remains uncertain. Moreover, accurately assigning Boltzmann weights to decoys produced by those methods lacking a direct physical interpretation is challenging. Such a Boltzmann ranking is critical simply because of the explosion in number of decoys that can be hallucinated from AlphaFold2 or future generative AI methods, now

including AlphaFold3.[10]

The AF2RAVE protocol integrates rMSA AF2 and the machine learning-based Reweighted Autoencoded Variational Bayes for Enhanced Sampling (RAVE) method[11–13] to systematically explore metastable states and accurately rank structures using Boltzmann weights.[14,15] Subsequently, traditional grid-based docking methods or recent generative diffusion models, like Diffdock[16] and DymanicBind[17] can be employed to dock ligands with a few top-ranked structures in metastable states, enabling further virtual screening on large ligand libraries. In this work, we chosen the industry-leading docking method Schrödinger Glide and Induced-fit docking (IFD)[18–20] as our primary docking method. By combining these two steps, we propose the AF2RAVE-Glide workflow (Fig 1) as an innovative approach for small molecule drug design, initiated from protein sequences.

In this work, we demonstrate this AF2RAVE-Glide workflow retrospectively on three different protein kinases and their type I and type II inhibitors. Protein kinases are involved in the regulation of various cellular pathways by catalyzing the hydrolyzing of ATP and transferring the phosphate group to substrate peptides/proteins. Dysfunctions of various kinases are know to cause human pathologies and cancers. The human genome contains about 500 protein kinases which share highly conserved structures in their catalytic ATP-binding pocket, due to the selection pressure towards functional catalysis. This poses significant challenges in developing selective small-molecule ATP-competitive kinase inhibitors, as they must effectively target the intended kinase while avoiding off-target interactions and associated side effects. Research efforts aimed at achieving selectivity have led to the development of 4 distinct types of kinase inhibitors,[21] including two types of ATP-competitive kinase inhibitors: type I inhibitors bind to the ATP-binding site adopting the active catalytic state, while type II inhibitors target the binding pockets adjacent to the ATP-binding site adopting the inactive state. These two states primarily differ in the configurations of their activation loop (A-loop), which is a flexible loop of approximately 20 residues. In the active state, the A-loop is "extended" to create a cleft for substrate peptides to bind, while in the inactive

state, it is collapsed or "folded" onto the protein surface, blocking substrate binding. Additionally, in the active state, the three residues 'Asp-Phe-Gly' (DFG motif) at the N-terminus of the A-loop bind to an ATP-binding Mg^{2+} ion, with the Asp side-chain pointing inward to the ATP-binding pocket, while in the inactive state, the Asp side-chain is flipped outward, and the DFG-motif adopts the DFG-out conformation.[22] Henceforth, we will refer to the type I inhibitor binding state as the active DFG-in state and designate the type II inhibitor binding state as the classical DFG-out state, as previously proposed in the literature[23–25] (demonstrated in Fig 2 A and B using Abl1 kinase as an example).

In this work, we investigated three kinases: (i) Abl1, which is targeted by the first clinically approved small-molecule kinase inhibitor imatinib for cancer therapy; (ii) DDR1, a structurally more flexible kinase which is identified as a promiscuous kinase targeted by chemically diverse inhibitors;[26] and (iii) Src kinase, another crucially important member of the tyrosine kinase family. In the following section, we applied AF2RAVE to enrich holo-like structures adopting the classical DFG-out state from the AF2-generated ensembles of DDR1 and Abl1 kinase, and further validated those holo-like structures by docking them with known type II kinase inhibitors. We initiated our investigation by conducting docking experiments involving both type I and type II inhibitors with the AF2 structures of these two kinases. We observed the incapacity of AF2 structures to effectively dock with type II inhibitors targeting the metastable classical DFG-out state. Subsequently, we employed rMSA AF2, an associative memory like process that generates diverse structure ensembles for both kinases[7,27], exploring their relatively improved but still limited potential to contain holo-like structures suitable for type II inhibitor binding. Ultimately, our study showcases AF2RAVE's effectiveness in enhancing the generation and selection of holo-like structures in metastable states by integrating AF2-based ensemble generation with physics-based methods.
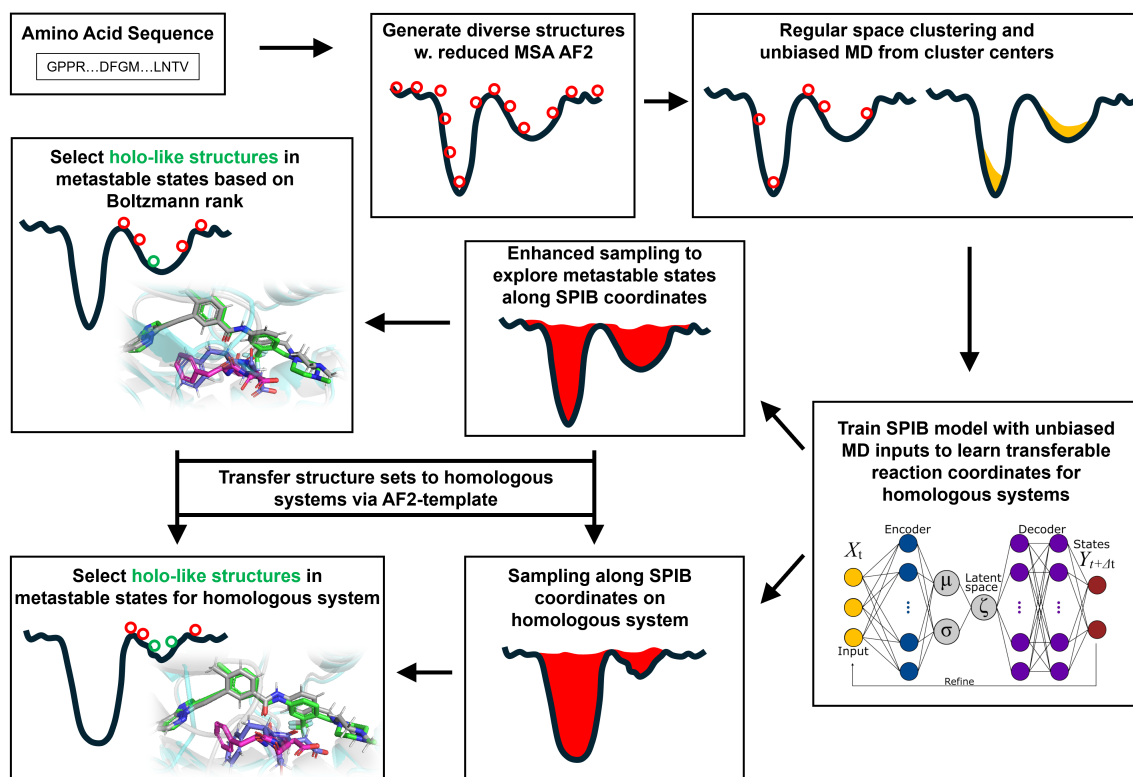
Figure 1: A schematic of the AF2RAVE-Glide workflow: (i) decoy structures generated by reduced MSA AF2, (ii) regular space clustering and unbiased MD simulations starting from cluster centers, (iii) State Predictive Information Bottleneck model (SPIB, a RAVE variant) to learn reaction coordinates from unbiased MD, (iv) enhanced sampling runs to calculate free energy landscape, (v) distinguish holo-like structures from decoys in metastable states based on Boltzmann rank and conduct Glide or Induced-fit Docking (IFD) on holo-like structures for ligands targeting metastable states. The decoy structure set and the learnt SPIB coordinates are transferable to homologous systems.

Results and discussion

AF2 structures fail to dock with type II kinase inhibitors

The utility of AF2-generated structures for structure-based drug discovery and virtual screening campaigns has been a subject of controversy and skeptical enthusiasm.[28–31] For proteins lacking crystal and cryo-EM structures but with homologous structures available in PDB, such as CDK20 kinase, AF2 demonstrates effectiveness as a homology modeling method for generating initial structures suitable for subsequent virtual screening.[32] Other AF2-based homology modeling method can bias AF2-generated structures towards user-selected template structures with specific druggable conformations.[33] However, a significant drop in the hits enrichment factor during virtual screening has been reported when employing AF2 structures as rigid receptors for docking, compared to using holo PDB structures. This occurs even in cases where the binding pockets of AF2 structures differ only slightly at 2 to 3 residues, from those of holo PDBs.[30,31] Given that ligands can induce slight relocation and side-chain rotation of pocket residues upon binding, it is important to note that AF2 does not account for this induced-fit effect, as it does not encode co-factors like ligands. Therefore, it appears necessary to perform ligand induced-fit modeling or relaxation on AF2 structures before engaging in any further structure-based drug design. Molecular dynamics (MD) simulations biased towards adjacent holo-template structure have proved effective in refining apo structures and improving their early enrichment performance.[34] It has also been demonstrated that AF2 structures can achieve comparable accuracy to crystal holo structures in Free Energy Perturbation (FEP) calculations, by superposing AF2 structures with crystal structures, grafting the co-crystallized ligands onto the AF2 structure, and optimizing the AF2 structure/ligand complex to account for subtle induced-fit effects.[35] AF2 structures, decorated with ligands from template ligand grafting method or known-hits docking method, and further refined by Schrödinger IFD-MD protocol, exhibit promising performance in early enrichment[36] and the prediction of novel ligand/protein complex structures.[37] Therefore, if

large backbone motion is not required, it appears feasible to refine the apo pockets in AF2 structures into holo-like pockets for structure based drug design. However, when AF2 structures exhibit significant steric clashes with holo ligands, especially in cases ligands targeting metastable states, direct application of out-of-the-box AF2 structures in docking methods for virtual screening and early enrichment may pose more challenges. The AF2 predicted kinase structures predominantly exhibit the DFG-in state, with over 95% of human kinases predicted in this conformation.[38,39] Significant A-loop motion and backbone flipping of the DFG motif are necessary to transition from the AF2-predicted DFG-in state to holo-like states, for type II ligands targeting the classical DFG-out state. As a result, AF2 structures of Abl1/DDR1 kinases exhibit superior performance when docking with type I inhibitors (achieving a minimum ligand RMSD of 2.14 Å) compared to docking with type II inhibitors using Glide XP (Fig 2 C&D). Even with Glide's Induced-fit Docking (Fig S16) and the highly side-chain clashes forgiving docking method DiffDock (Fig S20), AF2 structures struggle to dock with these metastable state-targeting ligands (type II inhibitors), with ligand RMSDs above 8 Å across all docking poses from all three docking methods.

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

AF2-based methods can achieve structural diversity by introducing dropouts in MSA inputs stochastically (rMSA AF2) or in a clustering manner (AF2-cluster). Additionally, models modified from the AF2 framework, such as the flow-match generative model AlphaFlow,[9] have been developed to explore the diversity of conformational space. Similar protocols to rMSA AF2 have demonstrated the potential to address the induced-fit effect by generating diverse structures at the binding pocket, ranging from closed apo pockets to opened holo-like pockets.[40] Other investigations have also indicated that larger backbone motions, such as DFG-motif backbone flipping and A-loop movement in at least some protein kinases, can be captured by the rMSA AF2 ensemble, although the distributions of conformations deviate
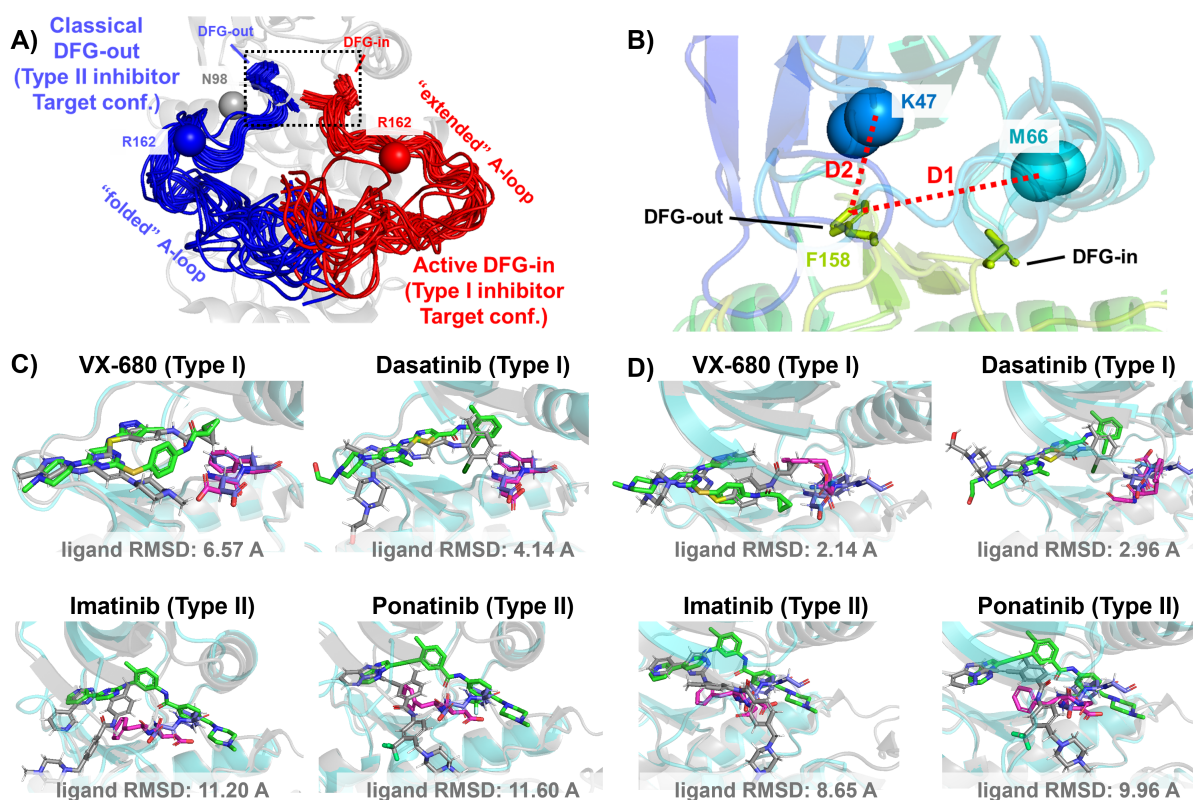
Figure 2: A)The NMR (Nuclear Magnetic Resonance) structures of Abl1 kinase overlay, comparing the activation loop (A-loop) in the active DFG-in state (red, PDB: 6XR6) with the classical DFG-out state (blue, PDB: 6XRG). Type I inhibitors target the active DFG-in state, where the DFG motif adopting the DFG-in configuration and the A-loop adopting the "extended" configuration, while type II inhibitors target the classical DFG-out state, where the DFG motif adopting the DFG-out configuration and the A-loop is "folded". The distance between CB atoms of residue N98 (grey bead) and residue R162 (red/blue bead) in Abl1 kinase serves as an order parameter here to illustrate the location of the A-loop. The dashed black block emphasizes the different configurations of the DFG motif in this two states. B) The Dunbrack definition for DFG motif configuration is employed here. The Dunbrack space is delineated by two order parameters: D1=dist(F158-CZ, M66-CA), D2=dist(F158-CZ, K47-CA). C) or D) The docking poses with the lowest ligand RMSD for 4 known kinase inhibitors targeting the Abl1 or DDR1 kinase AF2 structure, generated by Glide XP. Co-crystallized structures are shown as light-cyan cartoons (proteins), green sticks (ligand) and magenta sticks (DFG motif). Docking poses are shown as light-gray cartoons (proteins), gray sticks (ligand) and blue sticks (DFG motif). Comparing with type I inhibitors, AF2 structures of protein kinases fail to dock with type II inhibitors.

significantly from the correct Boltzmann distributions.[6,15]

In this section, we utilized rMSA AF2 to generate 1280 diverse structures for Abl1 kinase or DDR1 kinase: 640 for MSAs of depth 16 and 32. See Supplementary Material for information regarding Src kinase. Following a filtering step based on RMSD from corresponding AF2 structures, 1198 and 1147 structures remain for Abl1 and DDR1, respectively (Fig S1). As shown in Fig 3 A&B, for Abl1 kinase, only 4 out of 1198 structures have a folded A-loop, when using a distance cutoff of 15 Å between CB atoms of N98 and R162 in Abl1. However, for DDR1 kinase, 124 out of 1147 rMSA AF2 structures exhibit a folded A-loop, when using a salt bridge distance cutoff of 10 Å between the aligned residue pairs in DDR1 (E110 and R191). We then clustered the A-loop folded structures in the Dunbrack space. For Abl1, only two clusters were identified: one adopts the DFG-in state, and the other adopts the DFG-inter state, referring to an intermediate conformation during the backbone flipping from DFG-in to DFG-out, according to the Dunbrack definition. For DDR1, structures were divided into 5 clusters, with one cluster of size 15 being the closest to the classical DFG-out state.

We subsequently docked type II inhibitors, ponatinib and imatinib, to the most "classical DFG-out"-like cluster from rMSA AF2 ensembles (indicated as red circles in Fig 3 A&B), using IFD. Despite the A-loop being folded, the DFG-motif in structures from the Abl1 cluster is not strictly DFG-out. As expected, type II ligands fail to dock with those structures, with all IFD poses exhibiting ligand RMSDs above 9 Å (Fig S16 A, Fig 5 B). While AF2-based methods can indeed generate decoy structures that deviate from the native state, it remains uncertain whether these decoys correspond to metastable basins. Additionally, it's unclear whether these decoys includes structures that can represent the specific metastable states required for the intended types of drug design.

In contrast to Abl1, rMSA AF2 ensemble for DDR1 contains holo-like structure for type II inhibitors. One structure (which we label the IFD winner, shown as a red circle filled with green in Fig 3 B), out of the 15 in the DDR1 classical DFG-out cluster docked ponatinib with

a remarkably low RMSD of 0.89 Å using IFD (Fig 3 C). However, IFD poses from all the other structures exhibit large ligand RMSDs above 6 Å (Fig S16 B). Hence, an enrichment process to select holo-like structures from decoys becomes essential to ensure a practical number of pocket structures for ensemble docking and virtual screening.

We have also docked the IFD winner structure with another type II ligand, imatinib. In this case, the steric clashes are more significant, rendering the refinement for the induced-fit effect more challenging compared to the ponatinib scenario. We thus introduced an extra trimming step (see SI text for details) for the DFG-Phe residue which exhibit significant clashes with the holo-imatinib in the IFD winner structure (Fig S17 B). The IFD-trim protocol successfully achieves a minimum ligand RMSD of 1.04 Å for docking poses of the IFD winner structure with imatinib (Fig 3 C, Fig S17 C&D).

AF2RAVE on DDR1 enriches holo-like classical DFG-out structures in rMSA AF2 decoys

To assess the utility of physics-based methods in selecting holo-like structures from AF2-generated ensembles, we utilized a physics-based protocol, AF2RAVE[14] to explore the energy landscape of DDR1 kinase. We employed the identical set of collective variables (CVs) as in our previous study[15] to perform regular space clustering for the DDR1 rMSA AF2 ensemble. Two structures were selected from the clustering centers for each combination of DFG type (in, inter, or out) and A-loop position (folded or extended), resulting in a total of 12 initial structures (Fig S2). Subsequently, 50 ns unbiased MD simulations were conducted starting from each initial structure. The CVs extracted from all MD trajectories were input into the SPIB model with a linear encoder, to learn the reaction coordinates indicative of the slow motions linked to transitions between metastable states. The learnt SPIB reaction coordinates possess clear physical interpretations, with the first coordinate signifying the A-loop position and the second indicating the DFG-type (Fig 4 A&B). This alignment with physical features is expected as we deliberately selected diverse and representative initial structures
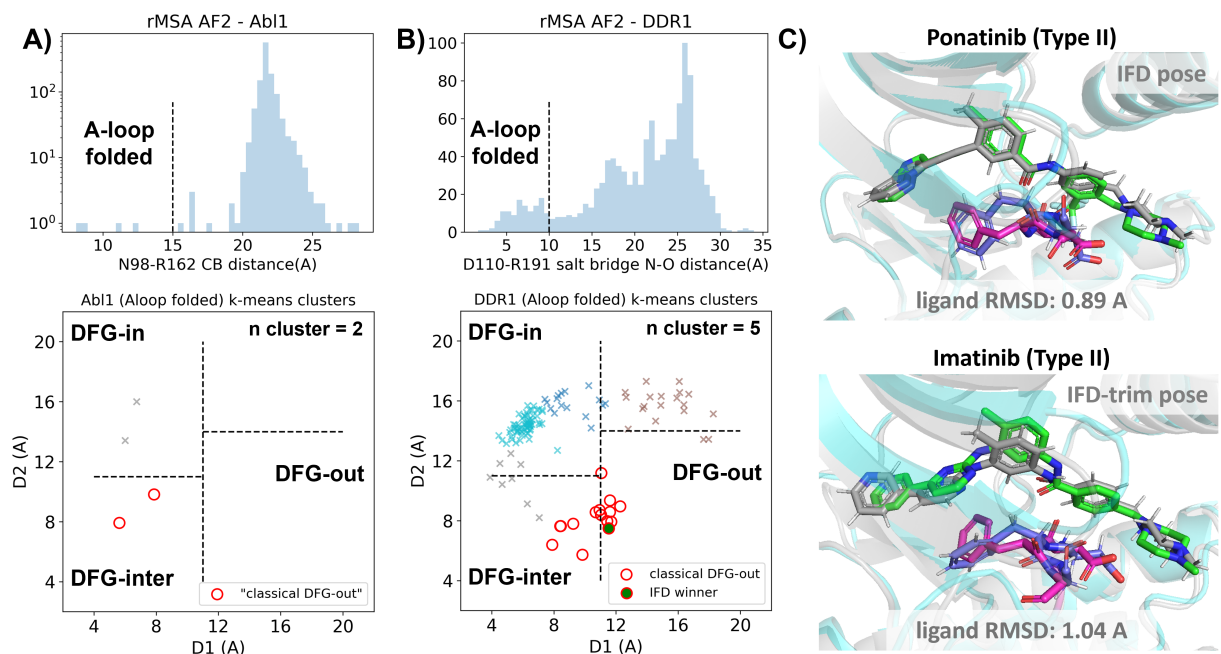
10

Figure 3: A) Upper panel: distribution of A-loop location for the reduced MSA AF2 structures of Abl1 kinase. 4 out of 1198 structures are A-loop folded. Lower panel: k-means clustering of 4 A-loop folded Abl1 structures in Dunbrack space, with the number of clusters (n cluster) set to 2. The structures in the cluster closest to the classical DFG-out state (red circles) fail to dock with type II inhibitors using induced-fit docking (IFD). B) Upper panel: distribution of A-loop location for the reduced MSA AF2 structures of DDR1 kinase. 124 out of 1147 structures are A-loop folded. Lower panel: k-means clustering of 124 A-loop folded DDR1 structures in Dunbrack space, with n cluster set to 5. Among the 15 structures in the cluster closest to the classical DFG-out state (red circles), one structure (IFD winner, highlighted by a red circle filled with green) demonstrates successful docking with type II inhibitors, showcasing a ligand RMSD < 2 Å, utilizing IFD or an extended-sampling version of IFD, IFD-trim. C) The docking poses with the lowest ligand RMSD for 2 type II kinase inhibitors targeting the DDR1 kinase IFD winner structure, generated by IFD or IFD-trim. The color code is the same as Fig. 2.

11

for unbiased MD simulations. Additionally, it's noteworthy that these reaction coordinates are transferable across various kinases and can effectively discern different metastable states.

Interestingly, the 15 classical DFG-out structures within the DDR1 rMSA AF2 ensemble are situated in a region of the latent space that were not thoroughly explored by the 12 unbiased MD trajectories. To address this gap, we employed enhanced sampling to sample along the SPIB-approximated reaction coordinates and compute the free energy profile inside the classical DFG-out basin. Considering both the flipping of the DFG motif and the overall motion of the large flexible A-loop, it might be impractical to sample direct back-and-forth transitions between various states using metadynamics. Therefore, we opted for umbrella sampling for its simplicity. The reliability of umbrella sampling hinges on two issues, first whether the latent space adequately represents the conformational space and second, whether there is sufficient overlap between different windows for efficient reweighting. Addressing the first challenge remains an ongoing endeavor in the dimensionality reduction research field, and we anticipate that our SPIB latent space is sufficient enough for our current purpose. The second challenge can be managed through careful setup of umbrella sampling windows and bias strength. Given our current setup, sampling the extensive motion of A-loop relocation remains challenging, showing insufficient overlap between the A-loop folded and extended regions (Fig S3). Consequently, the quantitative reliability of the absolute ΔG values between different states is limited, allowing us only to qualitatively assess the relative thermodynamic stability of the DFG-in versus the DFG-out basins. Nevertheless, the qualitative relative stability observed from umbrella sampling aligns with previous studies [15,26] (Fig S4). Furthermore, the local potential of mean force (PMF) surrounding each basin should provide quantitatively reliable insights, given their thorough sampling through umbrella sampling. Reassuringly, when we ranked the classical DFG-out structures within the DDR1 rMSA AF2 ensemble, using the local PMF values in the latent space, the IFD winner structure emerged among the top 2 structures with free energy relative to the minimum smaller than 1 kJ/mol, as illustrated in Fig 4 D.

Additionally, we used DiffDock to conduct docking experiments with type II ligands on the 15 classical DFG-out structures within the DDR1 rMSA AF2 ensemble. Despite poses from most structures surprisingly demonstrating very low ligand RMSD, typically below 2 Å (Fig S20), they exhibited significant steric clashes and low DiffDock confidence score due to Diffdock's disregard for side-chain configurations and clashes. Notably, upon ranking structures based on their AF2RAVE PMF and examining the corresponding poses with the lowest ligand RMSD, we observed that only the poses derived from the top 2 structures selected by AF2RAVE had Diffdock confidence scores higher than -1.5, surpassing the default threshold of the DiffDock model (Table S1). This implies the general advantage of structures selected by physics-based methods across various docking methods.

In summary, Boltzmann ranking from AF2RAVE effectively distinguishes holo-like structures from other decoys. Beginning with the DDR1 rMSA AF2 ensemble, AF2RAVE substantially enhances the likelihood of identifying the holo-like structure from 1 out of 15 to at least 1 out of 2 when using a PMF cutoff of 1 kJ/mol. We noted that although the PMF values and Boltzmann ranks may fluctuate with the setup of umbrella sampling, the enrichment effect of holo-like structures remains (Fig S7). Besides, we applied an alternative protocol to compute the PMF profile within the DDR1 classical DFG-out basin, by running unbiased MD simulations starting from the 15 classical DFG-out decoys in the DDR1 rMSA AF2 ensemble. However, this unbiased protocol poses a risk of failing to sample rare events if barriers exit within the region of interest. In this case, the mini-barriers inside the classical DFG-out basin are low enough to enable efficient sampling using 50 ns unbiased MD simulations. Consequently, the Boltzmann ranks derived from unbiased MD simulations also successfully enriched the single holo-like structure in the DDR1 rMSA AF2 ensemble to the top 5 of the structure set (Fig S11).
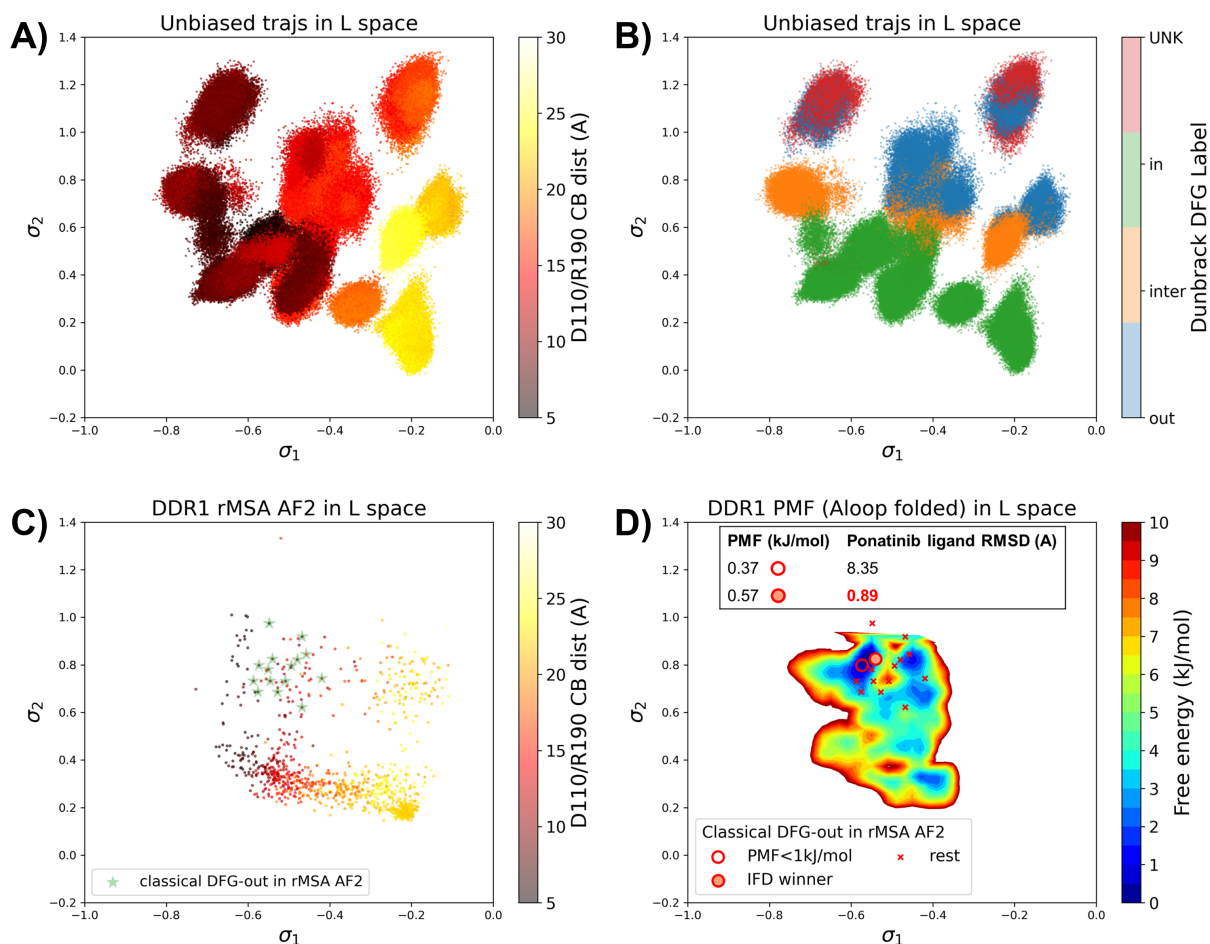
Figure 4: A or B) The unbiased MD trajectories of DDR1 are projected onto the learnt SPIB latent space. In plot A), the colors of sample points represent the A-loop location, while in plot B), they depict the Dunbrack DFG state. The first SPIB coordinate, σ_1, correlates with the A-loop location, and the second SPIB coordinate, σ_2, correlates with configuration of the DFG motif. C) The reduced MSA AF2 structures of DDR1 are projected onto the latent space. Sample points are color-coded based on the A-loop location. Light green stars highlight the 15 classical DFG-out structures selected based on prior information in Fig 3. D) Free energy profile in the A-loop folded region of the latent space, calculated from umbrella sampling simulations. The 15 classical DFG-out structures from reduced MSA AF2 are shown as red cross and circles (structures with free energy less than 1 kJ/mol). The IFD winner structure is emphasized using a red circle filled with red. The embedding table shows the lowest ligand RMSD in IFD poses of the rMSA AF2 structure with ponatinib. The IFD winner is among the 2 structures selected by AF2RAVE (PMF < 1 kJ/mol).

Transferable learning of holo-like structure for Abl1 and Src kinases with AF2RAVE from DDR1 templates

As mentioned earlier, our current Abl1 rMSA AF2 ensemble lacks any decoy structure in the classical DFG-out state. This is further illustrated in Fig 5 A, where the Abl1 rMSA AF2 ensemble is projected onto the same latent space learnt during the DDR1 AF2RAVE protocol. Hence, it's necessary to prepare the Abl1 decoy set adopting the classical DFG-out state before we can rank them using Boltzmann weights derived from physics-based methods and conduct further docking for selected holo-like structures.

There are several approaches to generate the Abl1 decoy sets. One can conduct enhanced sampling on the latent space starting from Abl1 rMSA AF2 structures to reach the classical DFG-out basin. Subsequently, MD structures from this basin can serve as templates for asking AF2 to generate crystal-like structures in classical DFG-out state. However, for simplicity, we opted to use the 15 classical DFG-out structures from the DDR1 rMSA AF2 ensemble directly as templates and employed an AF2-based homology modeling protocol, referred to as AF2-template (tAF2 for short, detailed protocol can be found in the SI text), to generate a decoy set comprising 30 Abl1 structures, as illustrated by the green stars in Fig 5 A.

Compared to the AF2 and rMSA AF2 structures, the performance of IFD on type II inhibitors shows a significant improvement when using the 30 tAF2 decoy structures of Abl1. The lowest ligand RMSD achieved is 2.74 Å for imatinib and 0.78 Å for ponatinib (Fig 5 B). However, only 4 structures out of the 30 decoys are capable of docking with type II inhibitors with ligand RMSD < 3 Å, and all other structures produce IFD poses with ligand RMSD > 6 Å (Fig S18).

We then investigated whether physics-based methods could enrich the 4 "IFD winner" structures from the 30 tAF2 decoys to a practical number of holo-like candidates. To explore the local classical DFG-out basin, we conducted 50 ns unbiased MD simulations starting from each tAF2 decoy structure and combined the trajectories to obtain the final Boltzmann dis-

tribution. Remarkably, the PMF for the Abl1 classical DFG-out basin calculated from the unbiased protocol (Fig 5 C) showed the enrichment of all 4 holo-like "IFD winner" structures among the top 8 tAF2 structures with PMF value < 1 kJ/mol. The PMF values of the top 8 tAF2 structures and the lowest ligand RMSD from the corresponding structures are presented in Fig 5 D. We also employed umbrella sampling for the Abl1 kinase by transferring setups and initial structures using the AF2-template from DDR1 umbrella sampling. However, we noted a significant occurrence of αC-helix breakage in the Abl1 umbrella sampling trajectories compared to DDR1 (Fig S5). After excluding windows with broken αC-helix, the Abl1 PMF derived from umbrella sampling ultimately enriched all 4 "IFD winner" structures to the top 6 among the 30 tAF2 structures (Fig S6). In Supplementary Information, we also provide results for Src kinase which are in the same quality as for Abl1 kinase.

Discussion

Through our retrospective analysis, we have thus demonstrated that the default AlphaFold2 models are ineffective for docking ligands targeting metastable protein kinase conformations. While AF2-based methods can be coaxed into generating diverse structures, they still struggle to produce reliable accuracy decoys for metastable conformations since the AF2 ensembles do not follow Boltzmann distribution. This failure is evident in the inability to generate an Abl1 AF2-ensemble containing holo-like structures for type II inhibitors. To further investigate whether this limitation is common among AF2-based methods, including the rMSA AF2 method we employed earlier, we tested another AF2-based approach, AF2-cluster.[8] The AF2-cluster ensemble of Abl1 comprises more A-loop folded structures (21 out of 197) compared to our Abl1 rMSA AF2 ensemble (4 out of 1198). However, similar to rMSA AF2, there are still no decoys in the classical DFG-out state, and all A-loop folded structures are located far from the PMF basin of the classical DFG-out state in the latent space (Fig S14 C). This indicates that AF2-cluster, like rMSA AF2, also fails to generate Abl1 metastable states effectively. Interestingly, we observed comparable structural diver-
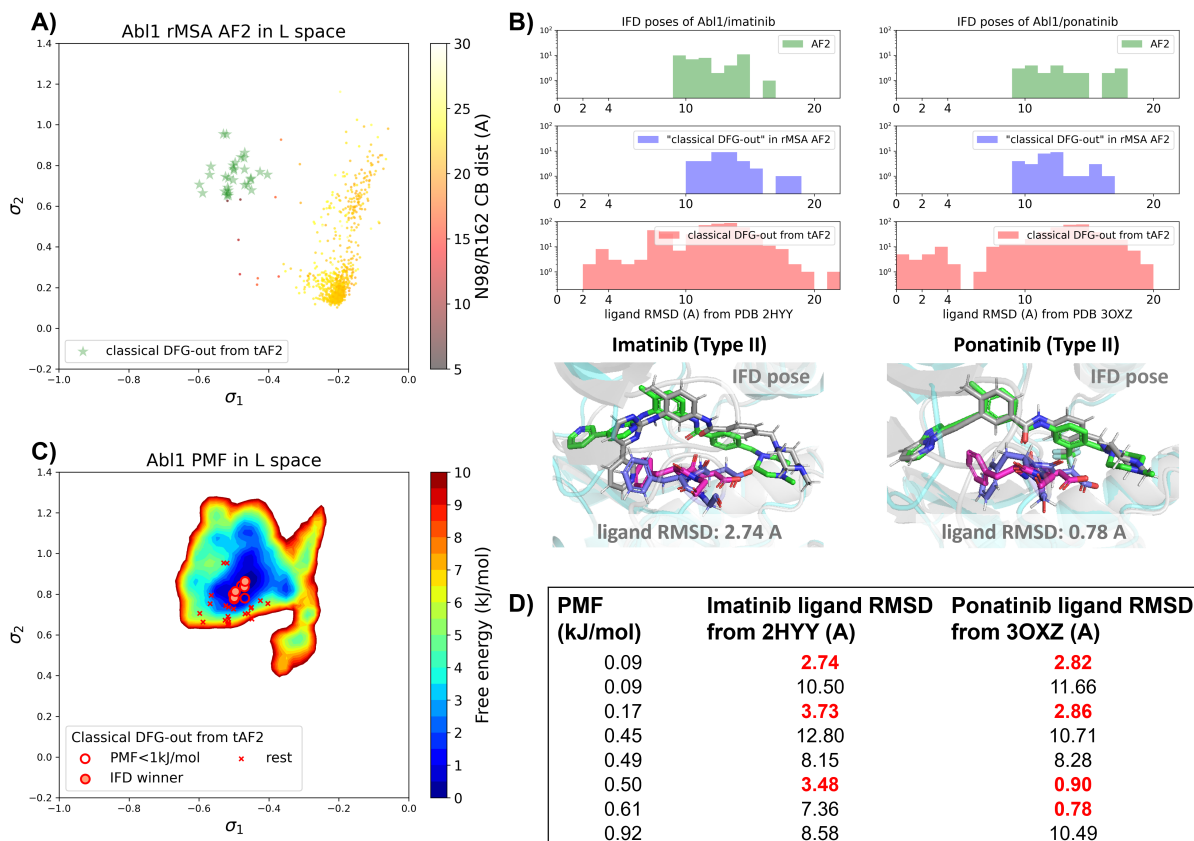
Figure 5: A) The reduced MSA AF2 structures of Abl1 are projected onto the latent space. Sample points are color-coded based on the A-loop location. Light green stars highlight the 30 AF2-template Abl1 structures modelled from the 15 DDR1 classical DFG-out structures. B) Upper panel: the distribution of ligand RMSD for the IFD poses of Abl1 structures and two type II ligands. Lower panel: IFD poses with the lowest ligand RMSD for Abl1 AF2-template structures and two type II ligands. The color code is the same as Fig. 2. C) Free energy profile in the latent space, calculated from unbiased MD simulations. The 30 Abl1 classical DFG-out structures from AF2-template are shown as red cross and circles (structures with free energy less than 1 kJ/mol). The IFD winner structures are emphasized using red circles filled with red. D) The table shows the lowest ligand RMSD in IFD poses of the AF2-template structures with two type II inhibitors. All the four IFD winners are among the 8 structures selected by AF2RAVE (PMF < 1 kJ/mol).

sity in sampling the A-loop folded configurations within the AF2-cluster ensembles for DDR1 and Abl1 (Fig S14 C&D). While for the rMSA AF2 method, the promiscuous kinase DDR1 ensemble exhibit superior structural diversity compared to the Abl1 kinase. This enhanced diversity leads to the identification of one dockable structure for type II inhibitors among decoys in DDR1 rMSA AF2 ensemble. Through the application of a homology modeling method, AF2-template, we demonstrated that the classical DFG-out decoys in the DDR1 rMSA AF2 ensemble can be transferred to Abl1 kinase. Furthermore, we tested an additional kinase, Src, for which non-native state decoys are reported to be even more challenging to produce using AF2 subsampling methods than Abl1 kinase.[6] We have also verified that rMSA AF2 and AF2-cluster struggle in producing distinct decoys from native structure of Src kinase (Fig S1 & Fig S12). Besides, neither the rMSA AF2 nor the AF2-cluster ensembles for Src kinase adequately sample the classical DFG-out basin in the latent space (Fig S13). Remarkably, AF2-template, as a homolgy modelling method, can easily produce the classical DFG-out structure of Src kinase, using the top 2 AF2RAVE-picked DDR1 classical DFG-out structures as templates (Fig S9). The IFD poses from the tAF2 structure of Src kinase shown a minimal ligand RMSD of 2.82 Å with imatinib (Fig S19).

Table 1: Comparing the IFD performance of various structure generation methods for docking type II kinase inhibitors

Source-protein (# of structures)	Lowest imatinib ligand RMSD (Å)	Lowest ponatinib ligand RMSD (Å)	Ratio of structs. w. ligand RMSD < 3 Å
AF2-Abl1 (1)	9.22	9.4	0/1
AF2-DDR1 (1)	9.24	9.33	0/1
rMSA AF2-Abl1 (2)	10.14	9.11	0/2
rMSA AF2-DDR1 (15)	**1.04***	**0.89**	1/15
tAF2-Abl1 (30)	**2.74**	**0.78**	4/30
AF2RAVE-DDR1 (2)	**1.04***	**0.89**	**1/2**
AF2RAVE-Abl1 (8)	**2.74**	**0.78**	**4/8**

* result from IFD-trim

With the rapid expansion of chemical space, virtual screening on libraries containing billions of diverse molecules becomes enticing for novel drug discovery.[41] Therefore, the enrichment of candidate holo-like structures emerges as a necessary step, offering significant

benefits in terms of computational efficiency and feasibility. As summarized in Table 1, unlike the non-dockable AF2 structures for type II inhibitors, the diverse rMSA AF2 ensemble (referred to as rAF2 in Table 1 for brevity) shows potential in generating holo-like structures within a large set of decoys. However, it's only upon AF2RAVE ranking and selection that the ratio of holo-like structures in selected structure set increases to a plausible value of 50%, facilitating further virtual screening on computational models of protein pockets.

Conclusion

AlphaFold2 has arguably revolutionized protein structure prediction, but it remains to be constructively demonstrated if it can be reliably used for drug discovery purposes, especially involving non-native protein conformations. In this work we have demonstrated through retrospective studies on kinase inhibitors that a combination of AlphaFold2, statistical mechanics based enhanced sampling and induced fit docking can be deployed for such calculations. Specifically, we have utilized the AF2RAVE protocol by inputting the sequences of the DDR1 kinases, along with two additional pieces of prior information: a pairwise distance cutoff for evaluating A-loop positions and the Dunbrack definition of DFG-type. We then employed Glide Induced-fit Docking (IFD) to assess our AF2RAVE-generated computational models for type II kinase inhibitor binding pockets. This AF2RAVE-Glide workflow yielded holo-like structure candidates with a 50% successful docking rate for known type II inhibitors. Notably, the holo-like structures in metastable state and the latent space constructed from AF2RAVE of DDR1 are transferable to other kinases. This includes the challenging cases[6] of Abl1 and Src kinases, wherein we showed that SPIB and sampling performed for DDR1 allowed generating classical DFG-out structures for both Abl1 and Src kinases. This severely reduces the computational cost for retraining SPIB to learn low-dimensional latent space for different kinases.

For the development of novel drugs targeting general proteins, all metastable states iden-

tified by AF2RAVE should be explored in subsequent docking experiments. Integration of algorithms capable of predicting ligand binding sites on protein surfaces, such as the Graph Attention Site Prediction (GrASP) model,[42] is then essential before utilizing AF2RAVE-selected structures in docking, thus expanding the workflow to AF2RAVE-GrASP-Glide. Additionally, the inclusion of free energy perturbation (FEP) calculations for front-runner ligands to evaluate the actual ligand binding affinity can further enhance this workflow.

The integration of AF2-based and physics-based methods presents a promising approach toward the development of a mature workflow for computer-aided drug design. AF2-based methods are capable of producing ensembles with structural diversity, which aids physics-based methods in better sampling and exploring the energy landscape of proteins. Additionally, AF2-generated structures can serve as crystal-like decoys, free from distortion that may occur in biased simulations. Physics-based methods play a crucial role in accurately assigning Boltzmann weights and ranking decoy structures to guide the enrichment of holo-like structures, essential for virtual screening on large libraries. This collaborative approach leverages the strengths of both methodologies, leading to enhanced efficiency and efficacy in drug discovery efforts.

We conclude this manuscript by making a final comment on the horizons opened by Generative AI methods, including those involving diffusion models[10,43] and any such future frameworks. These approaches make it possible to easily hypothesize regions of the conformational space underlying arbitrarily complex molecules of life, that then could serve as a starting point to launch more careful investigations. However, these predictions without careful *in situ* or *a posteriori* testing through advanced simulations or experiments, are only predictions and could just be hallucinations. Having the capability to quickly generate numerous - thousands or more - such structural hypotheses is what made AlphaFold2 so crucial to this current work. We believe that a deep integration of the hypothesis creation possibilities of Generative AI with careful Molecular Dynamics,[44,45] experiments or other forms of rapid testing is one way these methods will truly facilitate new and reliable scientific

discoveries.

Acknowledgement

Notes

The authors declare the following competing financial interest(s): P.T. is a consultant to Schrodinger, Inc. and is on their Scientific Advisory Board.

Supporting Information Available

Detailed description of methods used and further analysis of systems and sampling can be found in the supplement.

References

(1) others,, et al. Nature **2021**, 596, 583–589.

(2) Mirdita, M.; Schütze, K.; Moriwaki, Y.; Heo, L.; Ovchinnikov, S.; Steinegger, M. Nature methods **2022**, 19, 679–682.

(3) Amaro, R. E. Biophysical Journal **2019**, 116, 753–754.

(4) Davis, M. I.; Hunt, J. P.; Herrgard, S.; Ciceri, P.; Wodicka, L. M.; Pallares, G.; Hocker, M.; Treiber, D. K.; Zarrinkar, P. P. Nature biotechnology **2011**, 29, 1046–1051.

(5) Del Alamo, D.; Sala, D.; Mchaourab, H. S.; Meiler, J. Elife **2022**, 11, e75751.

(6) Monteiro da Silva, G.; Cui, J. Y.; Dalgarno, D. C.; Lisi, G. P.; Rubenstein, B. M. Nature Communications **2024**, 15, 2464.

(7) Porter, L. L.; Chakravarty, D.; Schafer, J. W.; Chen, E. A. bioRxiv **2023**, 2023–11.

(8) Wayment-Steele, H. K.; Ojoawo, A.; Otten, R.; Apitz, J. M.; Pitsawong, W.; Hömberger, M.; Ovchinnikov, S.; Colwell, L.; Kern, D. Nature **2024**, 625, 832–839.

(9) Jing, B.; Berger, B.; Jaakkola, T. arXiv preprint arXiv:2402.04845 **2024**,

(10) others,, et al. Nature **2024**, 1–3.

(11) Wang, Y.; Ribeiro, J. M. L.; Tiwary, P. Nature communications **2019**, 10, 3573.

(12) Wang, D.; Tiwary, P. The Journal of Chemical Physics **2021**, 154.

(13) Mehdi, S.; Smith, Z.; Herron, L.; Zou, Z.; Tiwary, P. Annual Review of Physical Chemistry **2024**,

(14) Vani, B. P.; Aranganathan, A.; Wang, D.; Tiwary, P. Journal of chemical theory and computation **2023**, 19, 4351–4354.

(15) Vani, B. P.; Aranganathan, A.; Tiwary, P. Journal of Chemical Information and Modeling **2023**,

(16) Corso, G.; Stärk, H.; Jing, B.; Barzilay, R.; Jaakkola, T. arXiv preprint arXiv:2210.01776 **2022**,

(17) Lu, W.; Zhang, J.; Huang, W.; Zhang, Z.; Jia, X.; Wang, Z.; Shi, L.; Li, C.; Wolynes, P. G.; Zheng, S. Nature Communications **2024**, 15, 1071.

(18) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Journal of medicinal chemistry **2004**, 47, 1750–1759.

(19) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Journal of medicinal chemistry **2006**, 49, 6177–6196.

(20) Sherman, W.; Day, T.; Jacobson, M. P.; Friesner, R. A.; Farid, R. Journal of medicinal chemistry **2006**, 49, 534–553.

(21) Müller, S.; Chaikuad, A.; Gray, N. S.; Knapp, S. Nature chemical biology **2015**, 11, 818–821.

(22) Modi, V.; Dunbrack Jr, R. L. Proceedings of the National Academy of Sciences **2019**, 116, 6818–6827.

(23) Gizzio, J.; Thakur, A.; Haldane, A.; Levy, R. M. Elife **2022**, 11, e83368.

(24) Thakur, A.; Gizzio, J.; Levy, R. M. The Journal of Physical Chemistry B **2024**,

(25) Gizzio, J.; Thakur, A.; Haldane, A.; Levy, R. bioRxiv **2024**, 2024–03.

(26) Hanson, S. M.; Georghiou, G.; Thakur, M. K.; Miller, W. T.; Rest, J. S.; Chodera, J. D.; Seeliger, M. A. Cell chemical biology **2019**, 26, 390–399.

(27) Roney, J. P.; Ovchinnikov, S. Physical Review Letters **2022**, 129.

(28) others,, et al. bioRxiv **2023**, 2023–12.

(29) Holcomb, M.; Chang, Y.-T.; Goodsell, D. S.; Forli, S. Protein Science **2023**, 32, e4530.

(30) Scardino, V.; Di Filippo, J.; Cavasotto, C. **2023**,

(31) Díaz-Rovira, A. M.; Martín, H.; Beuming, T.; Díaz, L.; Guallar, V.; Ray, S. S. Journal of Chemical Information and Modeling **2023**, 63, 1668–1674.

(32) others,, et al. Chemical Science **2023**, 14, 1443–1452.

(33) Sala, D.; Hildebrand, P. W.; Meiler, J. Frontiers in Molecular Biosciences **2023**, 10, 1121962.

(34) Guterres, H.; Park, S.-J.; Jiang, W.; Im, W. Journal of chemical information and modeling **2020**, 61, 535–546.

(35) Beuming, T.; Martín, H.; Díaz-Rovira, A. M.; Díaz, L.; Guallar, V.; Ray, S. S. Journal of Chemical Information and Modeling **2022**, 62, 4351–4360.

(36) others,, et al. This content is a preprint and has not been peer-reviewed

(37) Coskun, D.; Lihan, M.; Rodrigues, J. P.; Vass, M.; Robinson, D.; Friesner, R. A.; Miller, E. B. Journal of Chemical Theory and Computation **2023**, 20, 477–489.

(38) Modi, V.; Dunbrack Jr, R. L. Nucleic Acids Research **2022**, 50, D654–D664.

(39) Al-Masri, C.; Trozzi, F.; Lin, S.-H.; Tran, O.; Sahni, N.; Patek, M.; Cichonska, A.; Ravikumar, B.; Rahman, R. Bioinformatics Advances **2023**, 3, vbad129.

(40) Meller, A.; Bhakat, S.; Solieva, S.; Bowman, G. R. Journal of Chemical Theory and Computation **2023**, 19, 4355–4363.

(41) others,, et al. Nature **2019**, 566, 224–229.

(42) Smith, Z.; Strobel, M.; Vani, B. P.; Tiwary, P. Journal of Chemical Information and Modeling **0**, 0, null, PMID: 38453912.

(43) Wang, Y.; Herron, L.; Tiwary, P. Proceedings of the National Academy of Sciences **2022**, 119, e2203656119.

(44) Herron, L.; Mondal, K.; Schneekloth, J. S.; Tiwary, P. arXiv preprint arXiv:2308.14885 **2023**,

(45) others,, et al. Nature Machine Intelligence **2024**, 1–10.

Supplementary Information: Empowering AlphaFold2 for protein conformation selective drug discovery with AlphaFold2-RAVE

Xinyu Gu,[†] Akashnathan Aranganathan,[†,¶] and Pratyush Tiwary[∗,†,‡,§]

[†]*Institute for Physical Science and Technology, University of Maryland, College Park, Maryland 20742, USA*

[‡]*Department of Chemistry and Biochemistry, University of Maryland, College Park 20742, USA*

[¶]*Biophysics Program, University of Maryland, College Park 20742, USA*

[§]*University of Maryland Institute for Health Computing, Bethesda, United States*

E-mail: ptiwary@umd.edu

rMSA AF2 ensemble generation
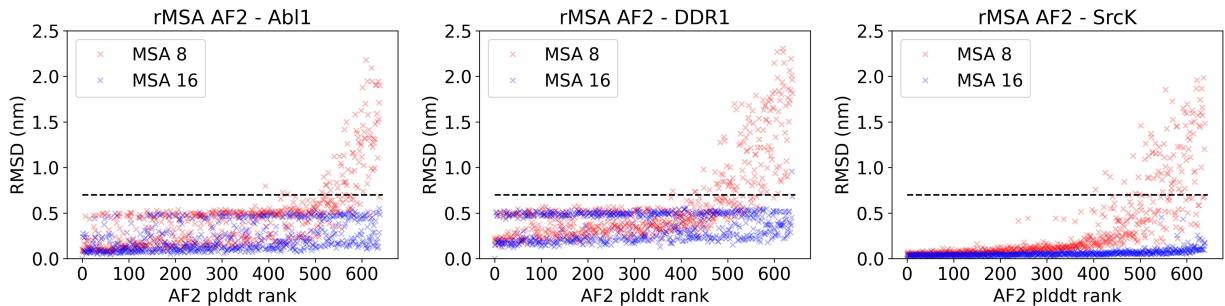
Figure S1: The AF2 pLDDT rank is plotted against the CA RMSDs from the AF2 structure (the one with the highest pLDDT) for each structure in the rMSA AF2 ensemble for Abl1, DDR1 or Src kinase. A RMSD cutoff of 7 Å (dashed black line) is applied to filter out unphysical structures with large RMSD from the native structure. Each rMSA AF2 ensemble is consist of 1280 structures, 640 for MSAs of depth 8:16 (red) and 16:32 (blue), separately.

AF2RAVE protocol on DDR1

Regular space clustering on the rMSA AF2 ensemble

We used the same 14 collective variables (CVs) for regular space clustering as in the previous AF2RAVE work on kinases.[1] These CVs are pariwise distances, selected to describe the kinase conformations around the ATP-binding pocket and the A-loop.
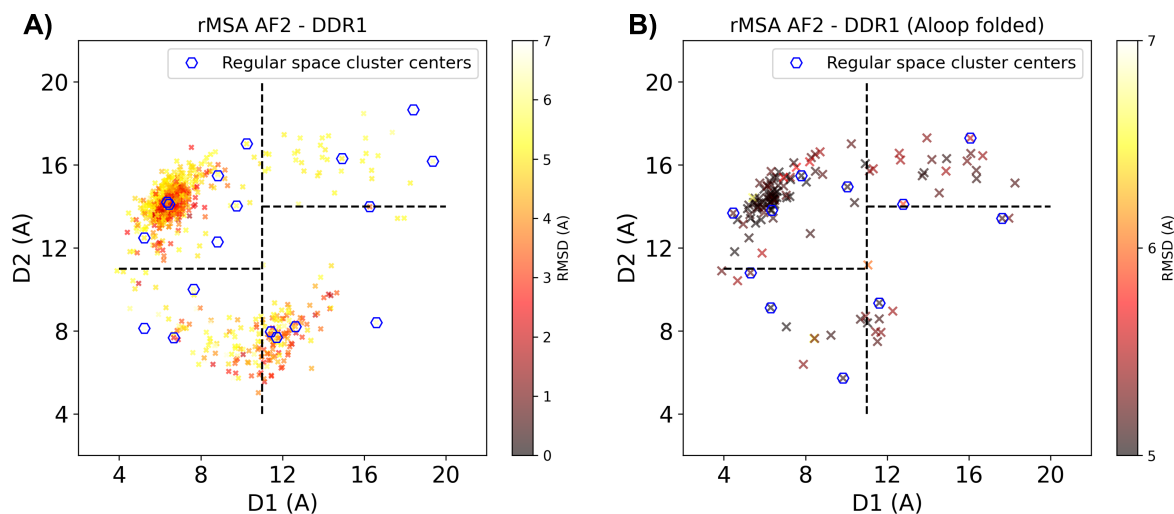
Figure S2: A) the rMSA AF2 ensemble for DDR1 is projected in the Dunbrack space. Sample points are color-coded based on the CA RMSD from the AF2 structure with the highest pLDDT. Regular space cluster centers are marked by blue hexagons. For each DFG-type (in, inter or out), top two cluster centers with the lowest CA RMSD are selected as AF2RAVE initial structures. B) To take account of the underrepresented A-loop folded configurations, an extra regular space clustering is conducted only for the A-loop folded structures in the rMSA AF2 ensemble. The color code, notation and the way to select initial structures are the same as plot A. Combining AF2RAVE initial structures from both plot A&B, there are 12 initial structures in total.

Unbiased MD and SPIB

50 ns unbiased MD simulation was run for each AF2RAVE initial structure of DDR1 from Fig S2. All the systems in this paper were parameterized with the Amber99SB*-ILDN force field[2] with the TIP3P water model.[3] The standard deviations of the 14 CVs are calculated after concatenating all the 12 unbiased MD trajectories. 8 CVs with standard deviations larger than 0.25 of the maximum standard deviation remain as the input features of the SPIB model. We conducted a parameter screening of the SPIB time lag, ranging from 1 ns to 40 ns with intervals of 1 ns. Eventually, we selected a time lag of 16 ns based on the performance of SPIB coordinates in representing physical features, including DFG-type and A-loop position.

PMF calculations from umbrella sampling

2D umbrella sampling is conducted along the two learnet SPIB coordinates, employing 11x11 windows. The bias potential equilibrium points are uniformly distributed in the SPIB latent space, with σ_1 ranging from -0.8 to -0.1 and σ_2 ranging from 0 to 0.8. The strength of the bias potential is set to 1000 $kJ/mol/nm^2$. Each window originates from the structure closest to the window's equilibrium point in Euclidean distance within the latent space among the 12 AF2RAVE initial structures and lasts for 100 ns (with the first 10 ns discarded in PMF calculation). For Abl1 kinase, we noticed a substantial proportion of the αC helix breaking in the umbrella sampling trajectories (Fig S9). This finding aligns with earlier enhanced sampling investigations on DDR1 using metadynamics,[1] where the authors imposed restraints to prevent αC helix breakage. In our study, we opted to exclude all umbrella sampling windows where the ratio of broken αC helix exceeded 20% for the Abl1 PMF calculation (Fig S6). The WHAM algorithm is applied to reweight the biased trajectories and compute the final PMF.

We must acknowledge the limitations of the way we assigned PMF values to AF-generated candidate holo structures. First of all, the free energy profiles are derived from MD simulations, and the PMF values directly correspond to the MD structures. We assumed that the latent space adequately represents the conformational changes of protein pocket within specific metastable states. Therefore, we project AF2-generated structures into the latent space and directly assign PMF values of the MD structures to AF-generated structures if they fall into the same bins in the latent space. Additionally, while the enrichment of holo structures in the top Boltzmann-ranked structures persists, the absolute PMF values and Boltzmann ranks of proper holo structures may fluctuate with the umbrella sampling setups, as depicted in Fig S7. Theoretically, the number of umbrella sampling windows and the simulation length should be sufficiently large for PMF convergence. However, there is always a trade-off between PMF accuracy and computational costs, so we opted to stick with the current setups.
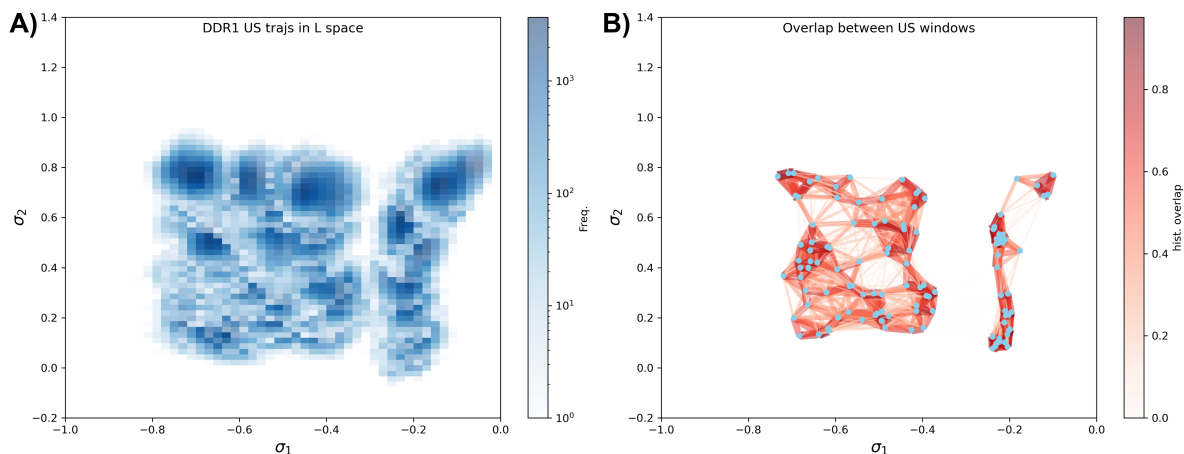
Figure S3: A) Distributions from different umbrella sampling windows in the latent space. B) The distribution overlap graph for all the umbrella sampling windows. The mean value of each distribution is shown as blue dots. Each distribution's 2D histogram is flattened into 1D vectors, and the cosine similarity between two distributions is then indicated by the width and color of the edge connecting the respective dots. Windows from the A-loop folded region are not overlapped well with the windows from the A-loop extended region, while windows inside the A-loop folded region (the left part of the graph) are well connected and are used for the local PMF calculation in Figure 4D.
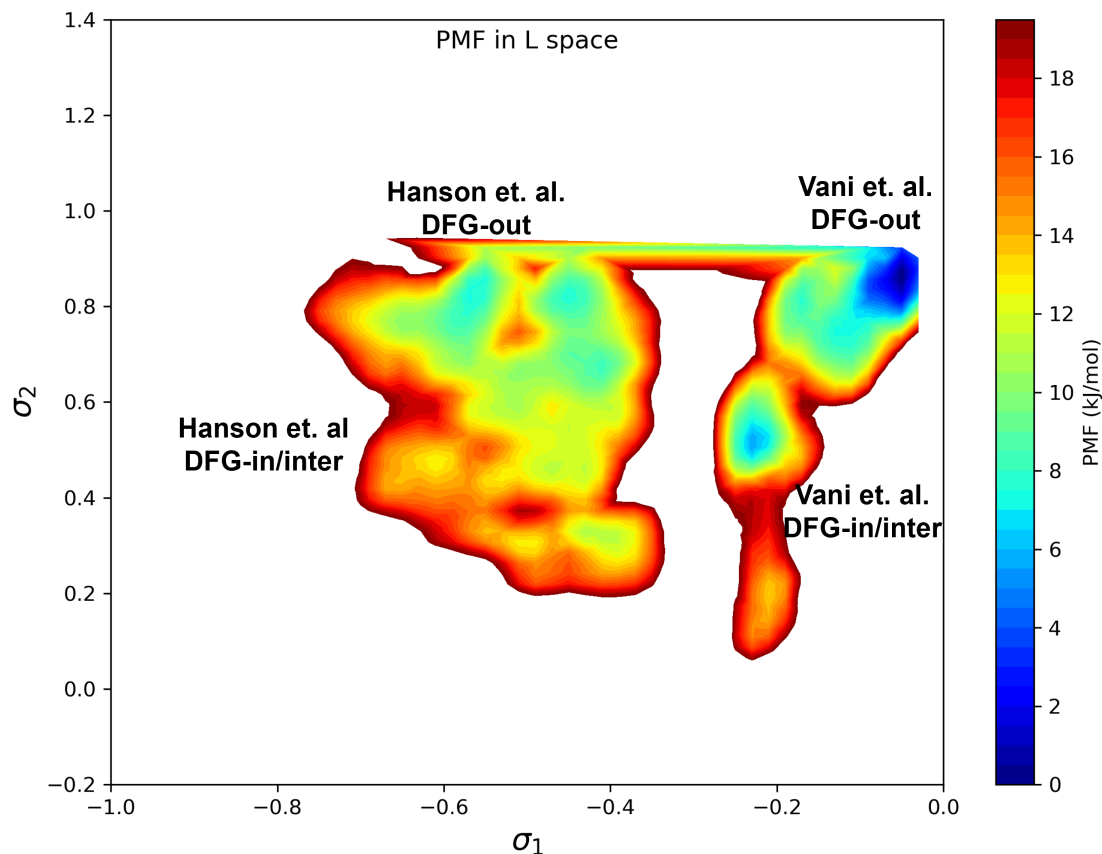
Figure S4: DDR1 PMF calculated with all the umbrella sampling windows. Hanson et al.[4] found the A-loop folded DFG-out state to be more stable than the A-loop folded DFG-in/inter state for DDR1; Vani et al.[1] reported that the A-loop extended DFG-out state is more stable than the A-loop extended DFG-in/inter state for DDR1. Although our umbrella sampling setup is not sufficient to sample the A-loop movement, the observed relative stability corresponds with the findings of Hanson et al. and Vani et al.
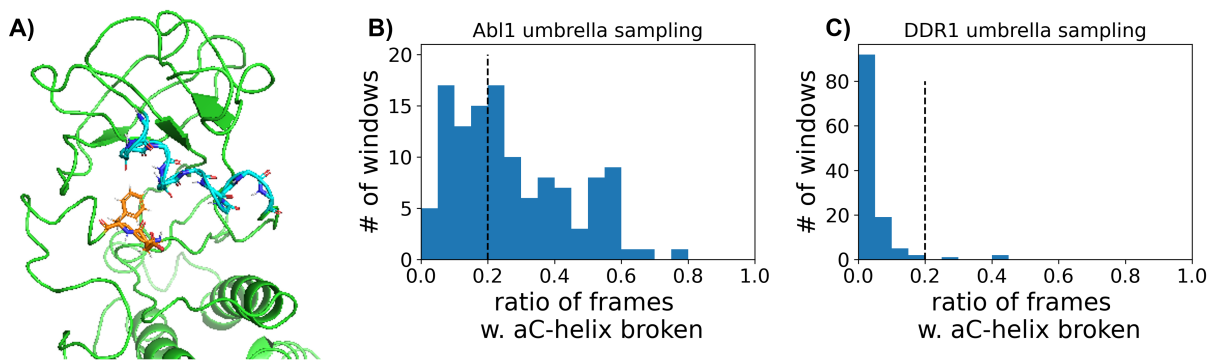
Figure S5: A) one representative frame with αC helix broken in Abl1 umbrella sampling trajectories. The backbone of the αC helix is shown with cyan sticks, while the DFG motif is shown as orange sticks. B) or C) The distribution of the ratios of frames with αC helix broken in each umbrella sampling window for Abl1 or DDR1.
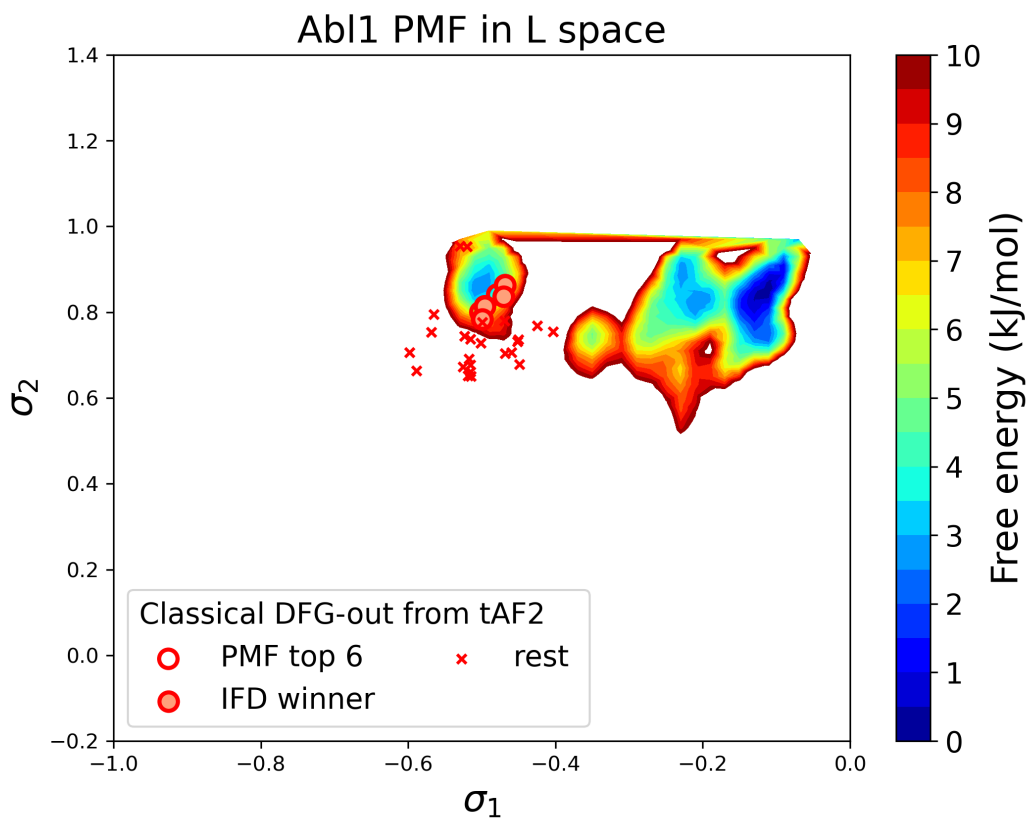
Figure S6: Abl1 PMF calculated from umbrella sampling after discarding windows with αC helix broken. The four holo-like structures (IFD winners) are enriched to the top six based on PMF values.
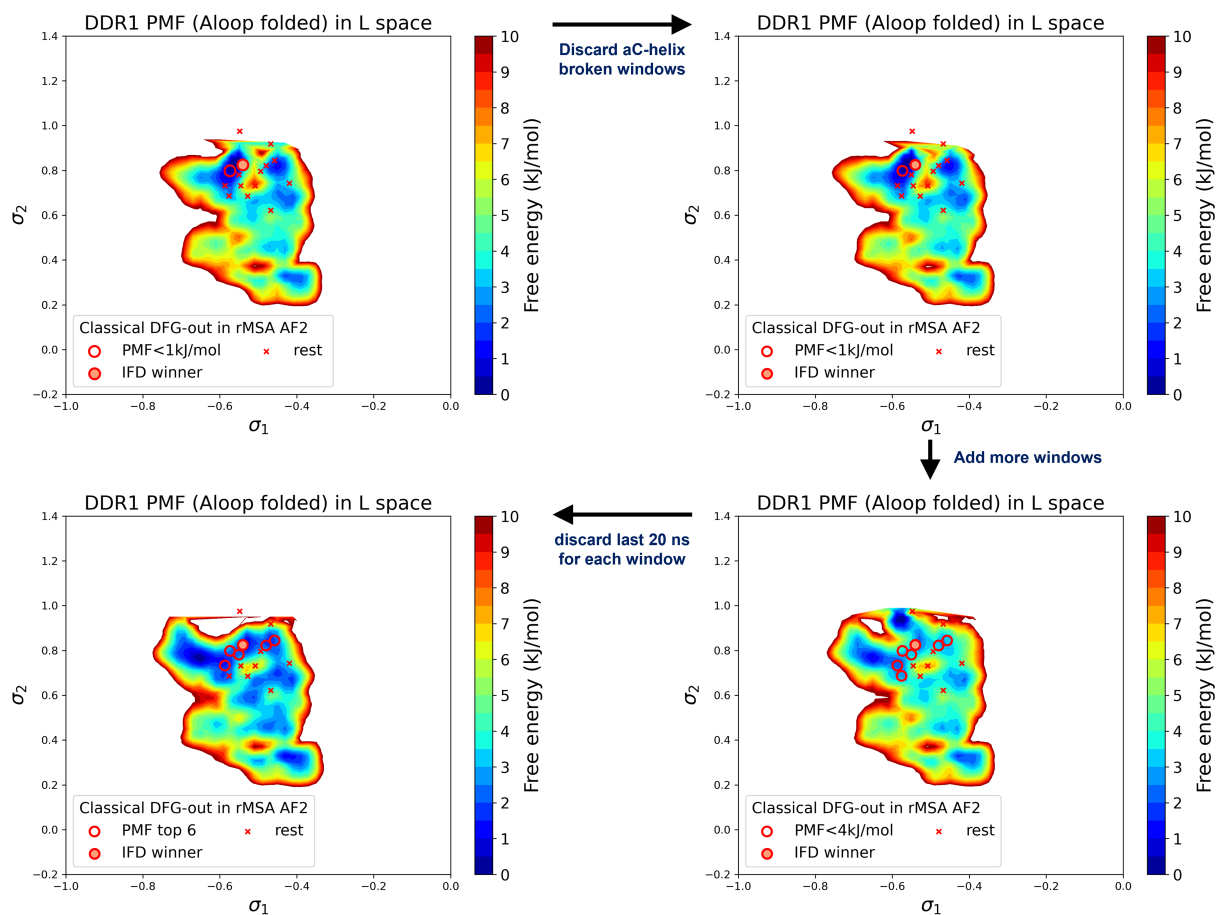
Figure S7: PMF values and Boltzmann ranks of candidate structures fluctuate with the selection of the umbrella sampling windows and the simulation length of umbrella sampling trajectories, demonstrated with the DDR1 system.

AF2-template

The AF2-based homology modeling protocol, AF2-template is implemented by uploading desired template structure to the colabfold[5] and deactivating the evoformer module. For each template, 5 AF2-template models are generated, with the last 3 structures exhibiting low pLDDT discarded.
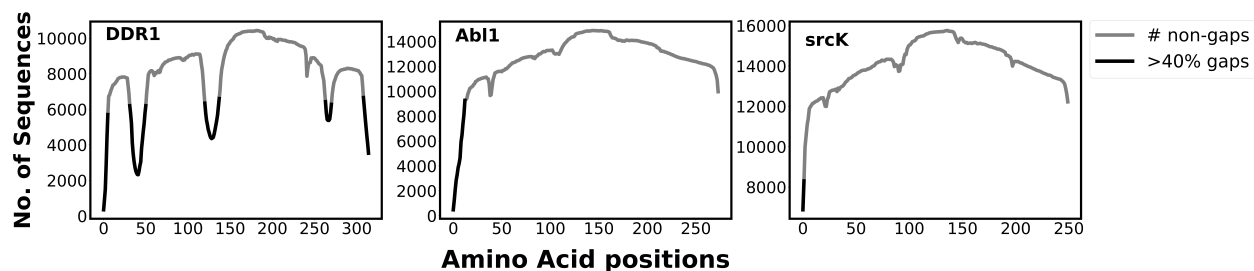
Figure S8: The plot illustrates the number of gaps in the multiple sequence alignment (MSA) generated by mmseq2 (using Colabfold[5]) for different kinases. The non-gap count describes the coverage of each position in the MSA. The presence of residue positions with gap counts higher than 40 per cent of the total sequence in DDR1 implies that it has fewer conserved regions than abl1 kinase and src kinase. This characteristic of DDR1 MSA enables the rMSA AF2 protocol to generate multiple conformations for DDR1, including the classical DFG-out conformation, by initializing it at various states. However, the highly conserved nature of abl1 and src makes it challenging for the rMSA AF2 to initialize at a state that can lead to a classical DFGout conformation. Therefore, we used the AlphaFold template protocol to overcome this initialization issue with rMSA AF2.
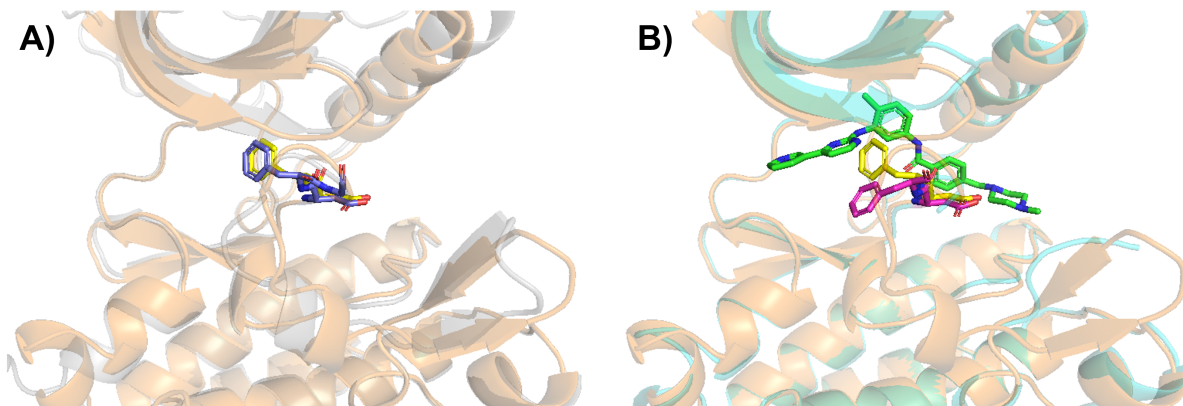
Figure S9: A) the AF2-template structure for Src kinase is superimposed with its template structure (classical DFG-out in DDR1 rMSA AF2 ensemble, IFD winner). The tAF2 structure of Src is shown as light-orange cartoon (protein) and yellow sticks (DFG motif), while DDR1 template is shown as light-gray cartoon (protein) and blue sticks (DFG motif). B) the AF2-template structure for Src kinase is again superimposed with Src/imatinib co-crystallized structure (PDB 2OIQ). Crystal structure is shown as light-cyan cartoon (protein), green sticks (ligand) and magenta sticks (DFG motif).

9

PMF calculations from unbiased MD

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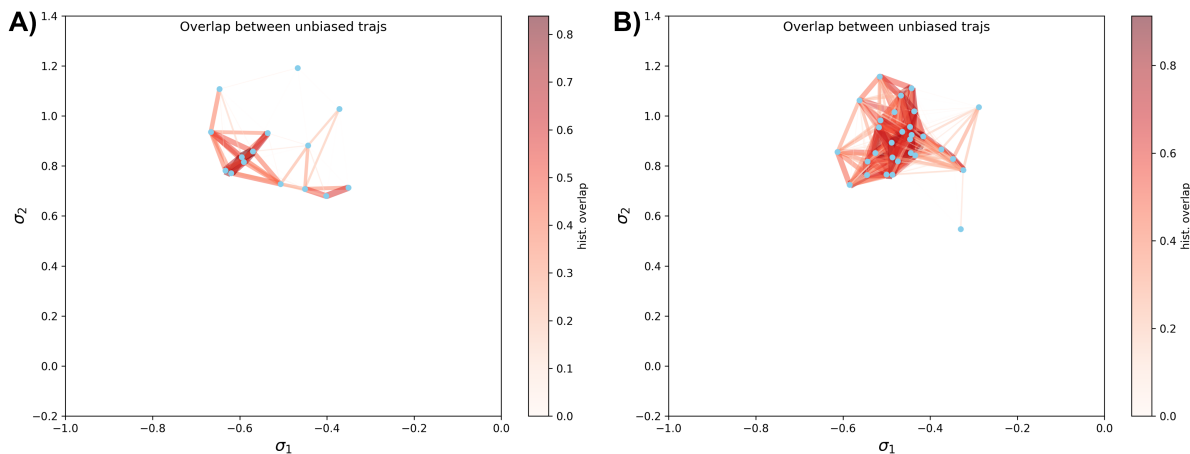

Figure S10: A) The distribution overlap graph for all the unbiased MD trajectories starting from 15 classical DFG-out structures in DDR1 rMSA AF2 ensemble. B) The distribution overlap graph for all the unbiased MD trajectories starting from 30 Abl1 tAF2 structures in classical DFG-out state. The color-code is the same as Figure S3
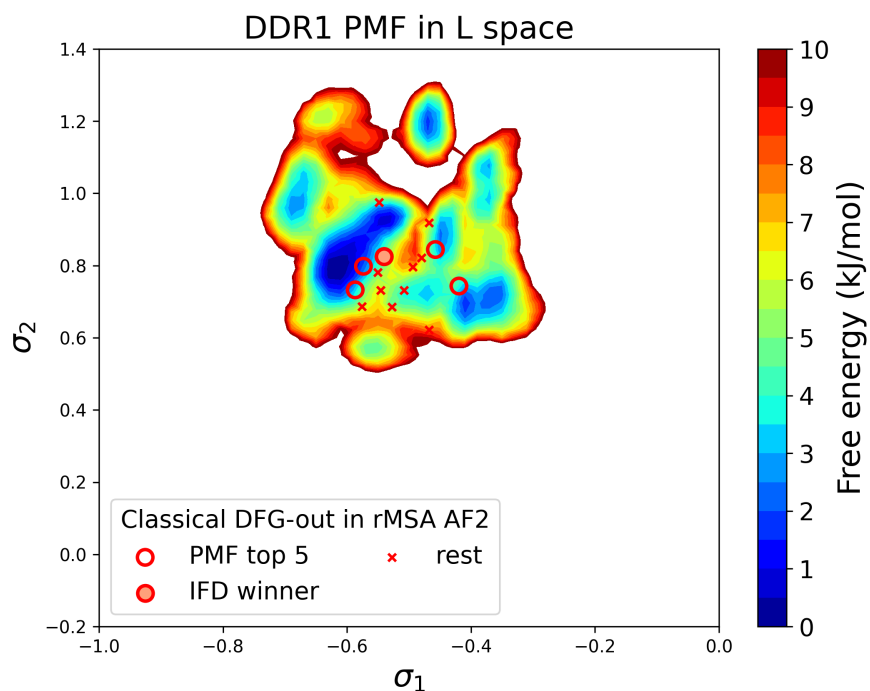
Figure S11: Free energy profile for DDR1 in the latent space, calculated from unbiased MD simulations. The 15 DDR1 classical DFG-out structures in rMSA AF2 are shown as red cross and circles (top 5 structures ranked by free energy values). The IFD winner structure is emphasized using a red circle filled with red.

AF2-cluster

AF2-cluster for DDR1 and Abl1 were run with default setups.[6]
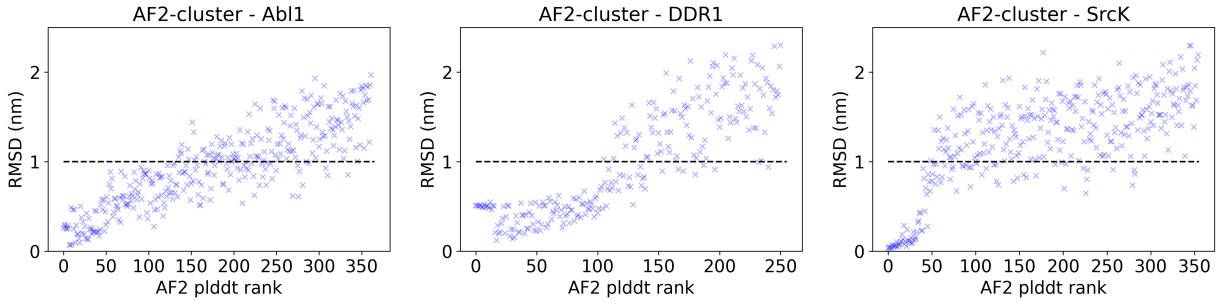
Figure S12: The AF2 pLDDT rank is plotted against the CA RMSDs from the AF2 structure for each structure in the AF2-cluster ensemble for Abl1, DDR1 or SrcK. A RMSD cutoff of 10 Å (dashed black line) is applied to filter out unphysical structures with large RMSD from the native structure. After the RMSD filter, 197 out of 362 structures remain for Abl1, 134 out of 251 structures remain for DDR1, and 93 out of 355 structures remains for SrcK.
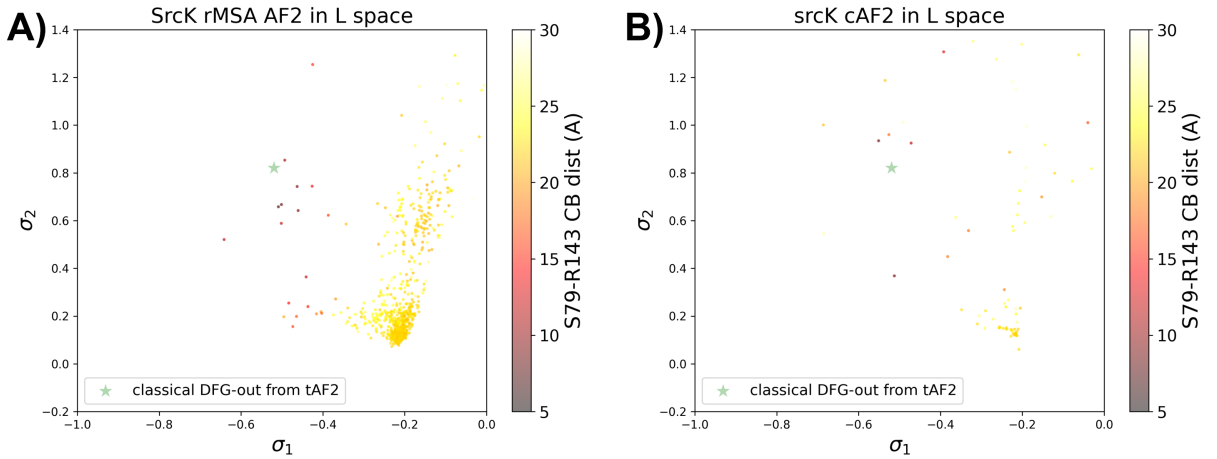

Figure S13: The projection of A) the rMSA AF2 ensemble or B) the AF2-cluster ensemble on the AF2RAVE latent space for SrcK. The classical DFG-out SrcK structure generated from AF2-template in Fig S9 is shown as the green star. The color-code shows the A-loop location.
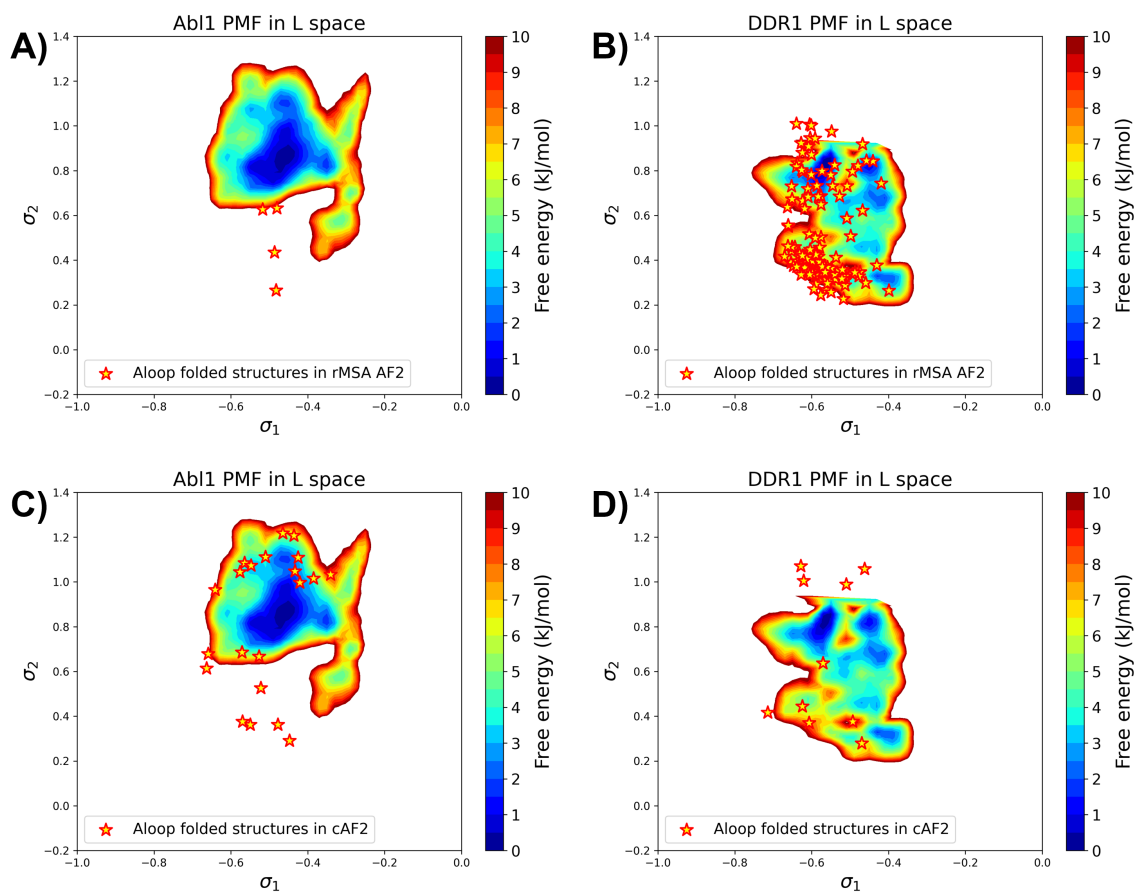
12

Figure S14: The projection of A-loop folded structures from the rMSA AF2 ensemble or the AF2-cluster ensemble on the AF2RAVE PMF for Abl1 or DDR1.
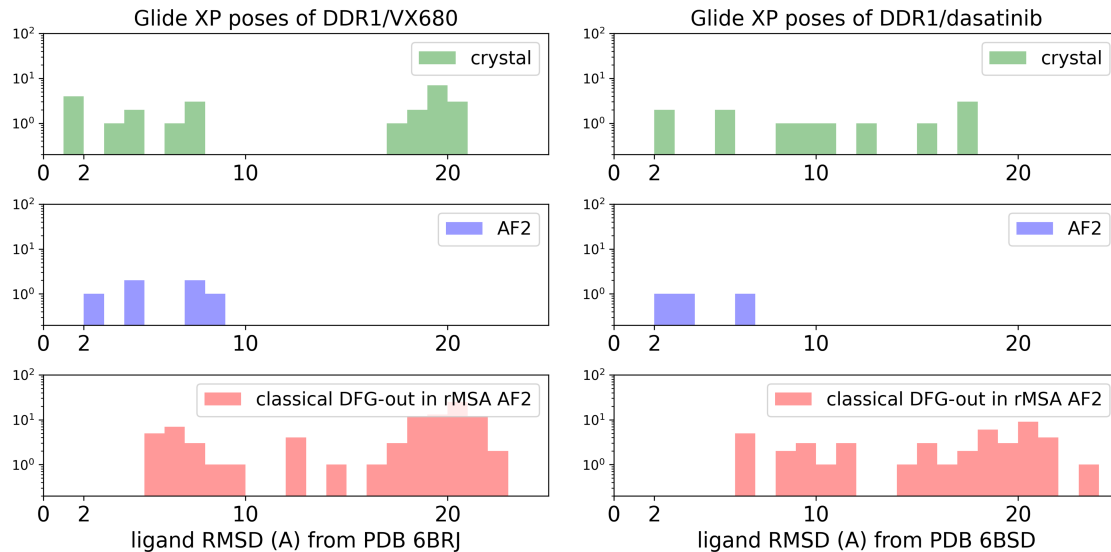
Docking details

All the input structure for our docking experiments were relaxed in solution first with a MD energy minimization step.

Glide XP

Glide XP docking experiments in this work were run with default setups in the Maestro software.
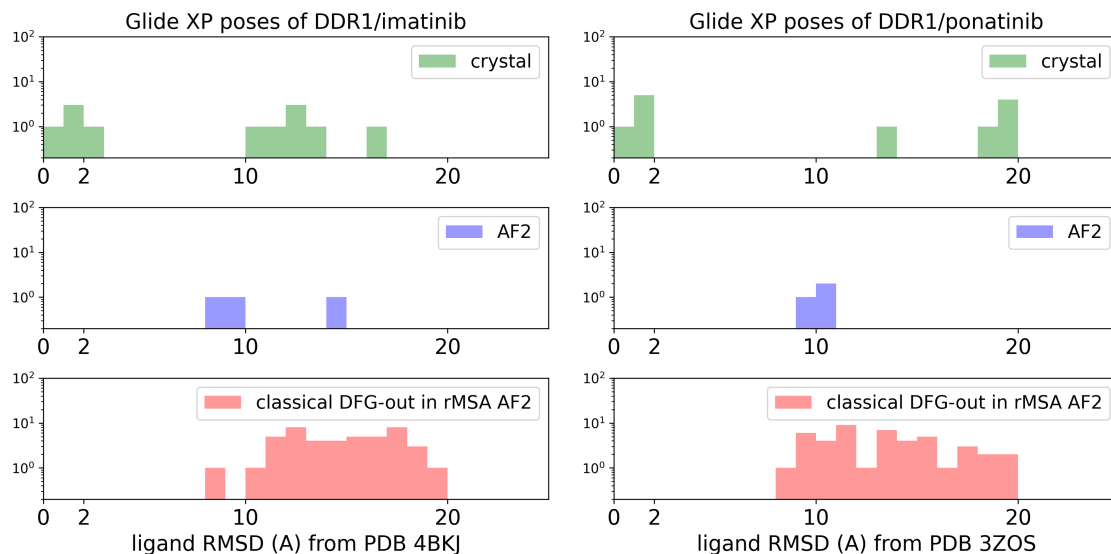
Figure S15: The distributions of ligand RMSDs for Glide XP docking poses of DDR1 and type I/type II inhibitors (upper/lower panel). Results from cross-docking against 4 crystal holo structures, docking against the AF structure, and docking against 15 classical DFG-out structure in rMSA AF2 ensemble are shown as green, blue, and red, separately.

IFD

For Induced-fit Docking, we used the Glide XP for initial docking, followed by Prime relaxation and final Glide XP docking. Parameters remain default in Maestro IFD.

Given Ponatinib's backbone feaftures, notably its lengthy and slender C-C triple bond, it exhibits reduced sensitivity to steric clashes, resulting in successful docking with the DDR1 IFD winner structure at a ligand RMSD of 0.89 Å. Conversely, the IFD winner structure struggles to accurately dock with imatinib using IFD (Fig S16 B). We then employed an extended-sampling IFD approach by initially trimming the DFG-Phe residue from the IFD winner structure. The trimmed residue is temporarily mutated to alanine during initial docking and later restored in subsequent Prime relaxation and final docking steps. Here, in the IFD winner structure, we manually chose the DFG-Phe residue which is significantly hindered by holo-imatinib (Fig S17 B). For generic systems lacking groundtruth information, broader screening of single residue trimming for the protein pocket may be necessary for this extended-sampling IFD method. In essence, it's a trade-off between the quality of holo-like structure to dock with and the accuracy/complexity of the docking method.
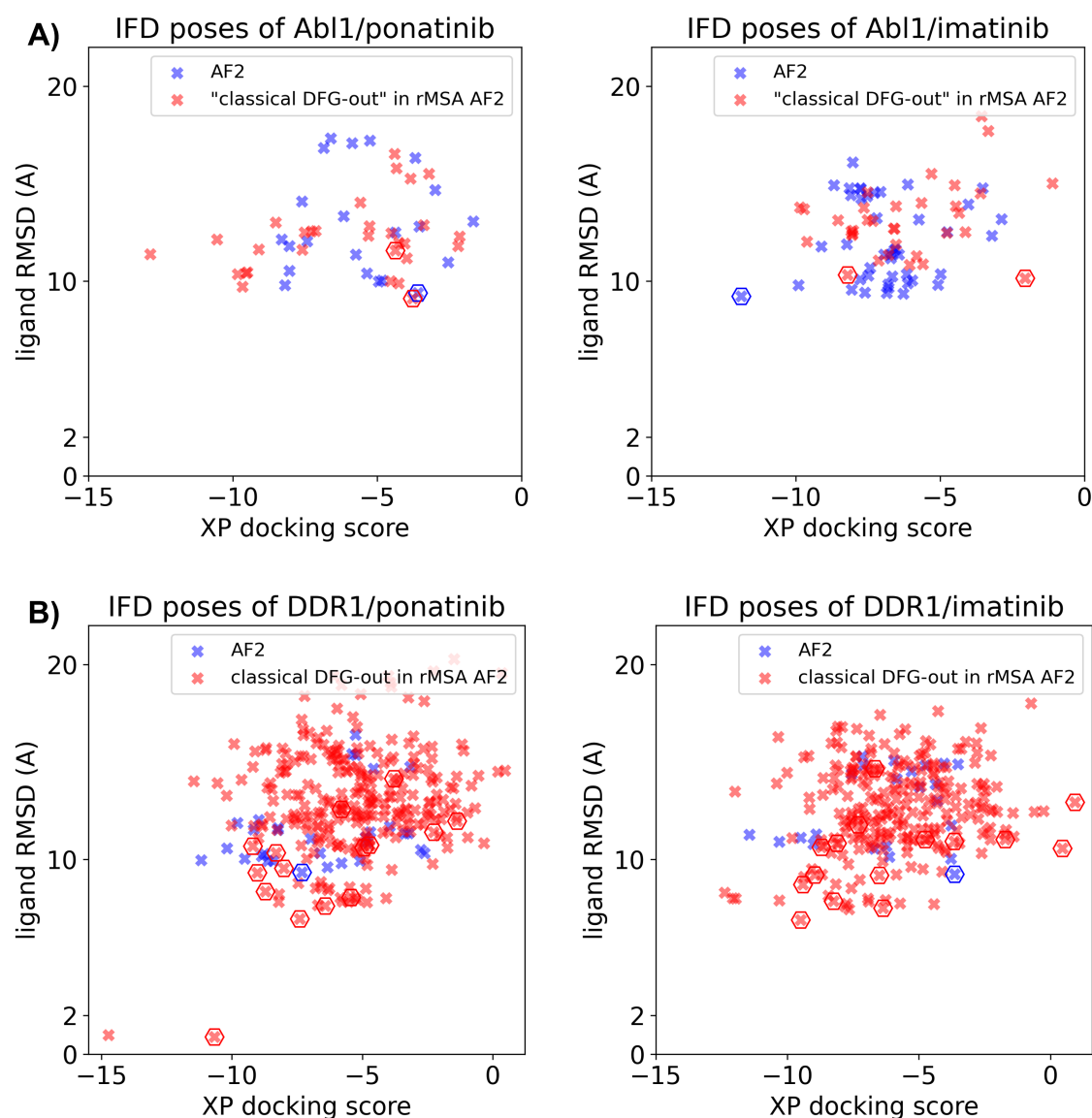
Figure S16: Ligand RMSDs are plotted against the docking scores for the IFD docking poses of type II inhibitors (ponatinib and imatinib) against AF2 structure (blue) or classical DFG-out structures in rMSA AF2 ensembles (red). A) IFD docking results for Abl1. B) IFD docking results for DDR1. The pose with the lowest ligand RMSD from each input structure is marked by hexagon.
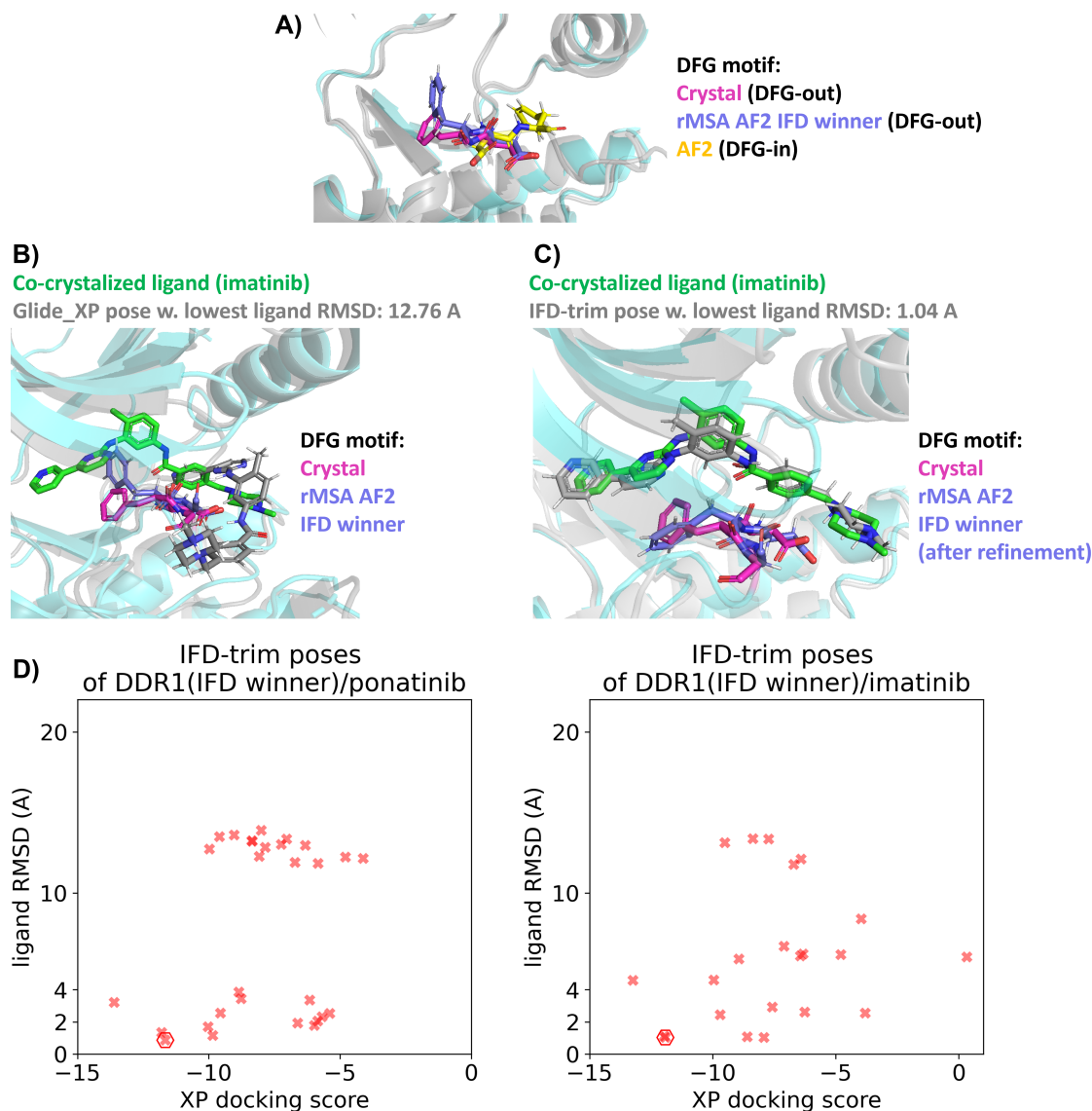
Figure S17: A) Comparison of the DFG motif for DDR1 in its co-crystalized structure with imatinib (PDB 4BKJ), its IFD winner structure and its AF2 structure. B&C) In the IFD winner structure, the Phe residue in the DFG-motif requires rotation to prevent steric clashes with imatinib. proteins from crystal structure are shown as cyan cartoon, while all the other proteins are shown as grey cartoon. D) Ligand RMSDs are plotted against the docking scores for the IFD-trim docking poses of type II inhibitors (ponatinib and imatinib) against the IFD winner structure in DDR1 rMSA AF2 ensembles. The pose with the lowest ligand RMSD is marked by hexagon.
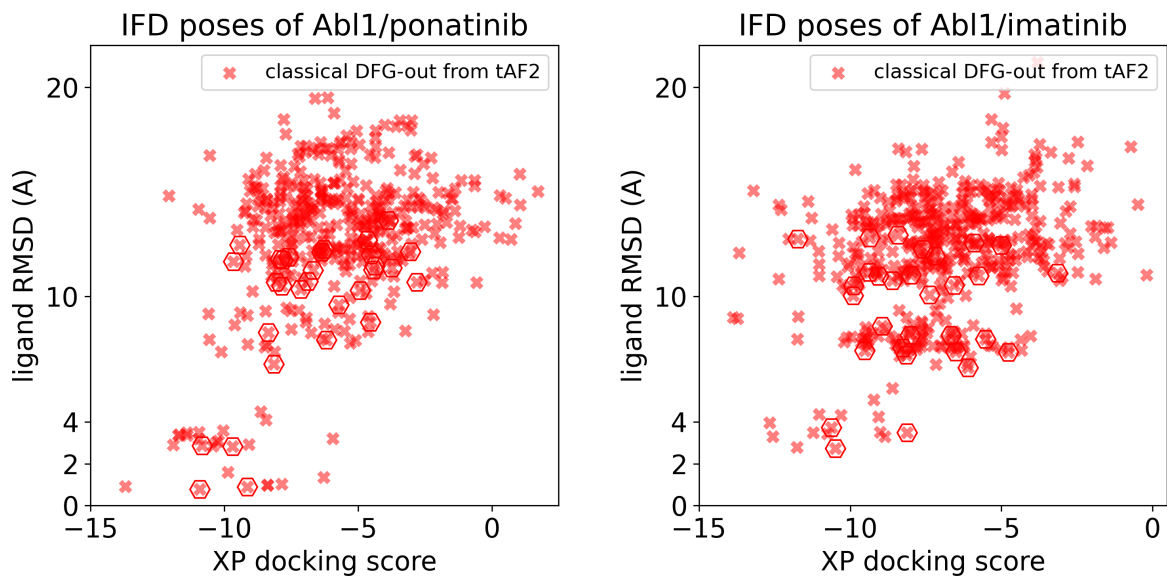
Figure S18: Ligand RMSDs are plotted against the docking scores for the IFD docking poses of type II inhibitors (ponatinib and imatinib) against Abl1 tAF2 structures. The pose with the lowest ligand RMSD from each input structure is marked by hexagon.
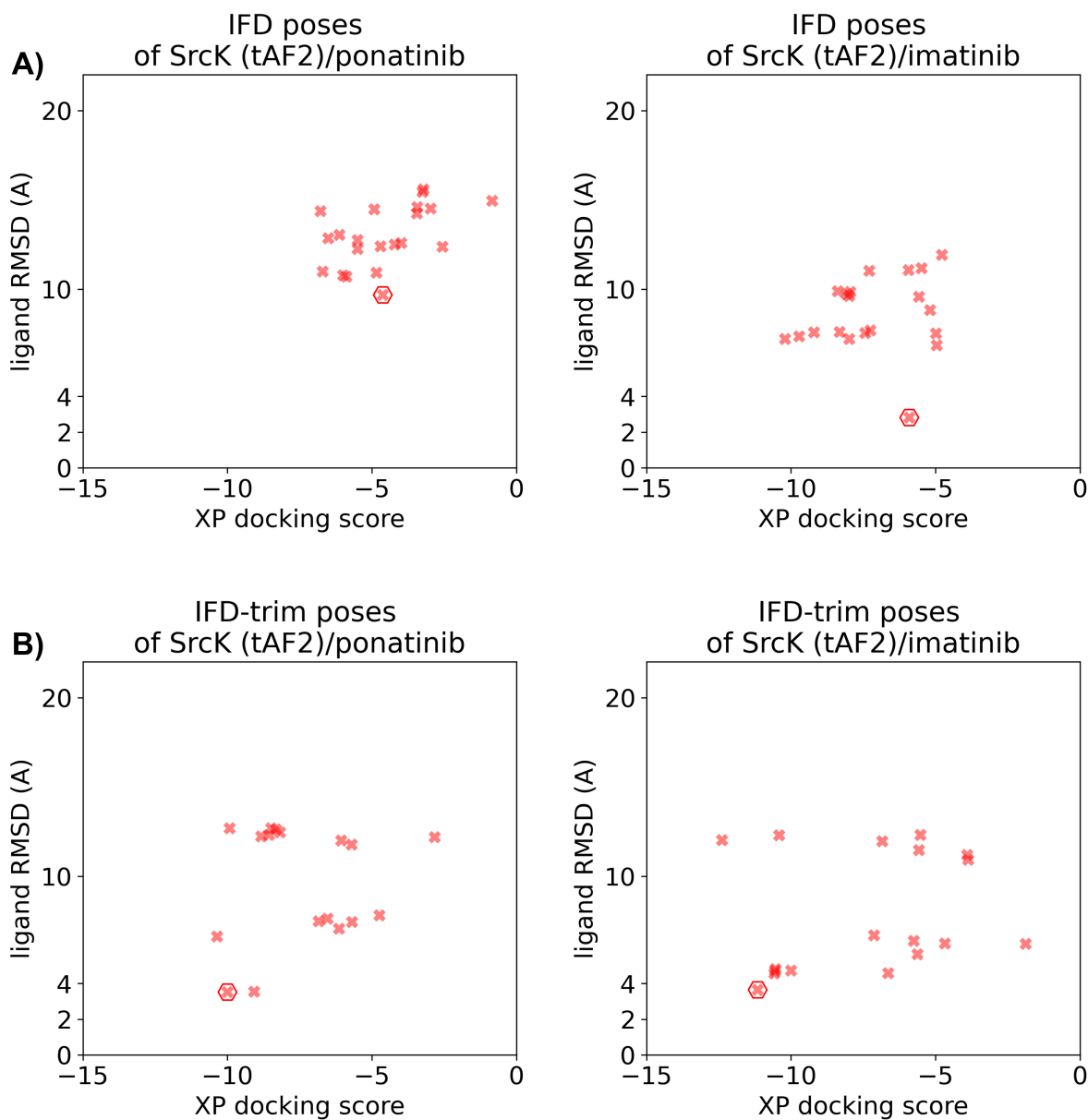
Figure S19: Ligand RMSDs are plotted against the docking scores for the IFD/IFD-trim docking poses of type II inhibitors (ponatinib and imatinib) against the SrcK tAF2 structure. The pose with the lowest ligand RMSD from each input structure is marked by hexagon.

Diffdock performace on classical DFG-out DDR1

DiffDock docking experiments in this work were run in the webserver with default setups in the version before 3/8/2024 (https://huggingface.co/spaces/simonduerr/diffdock).[7]
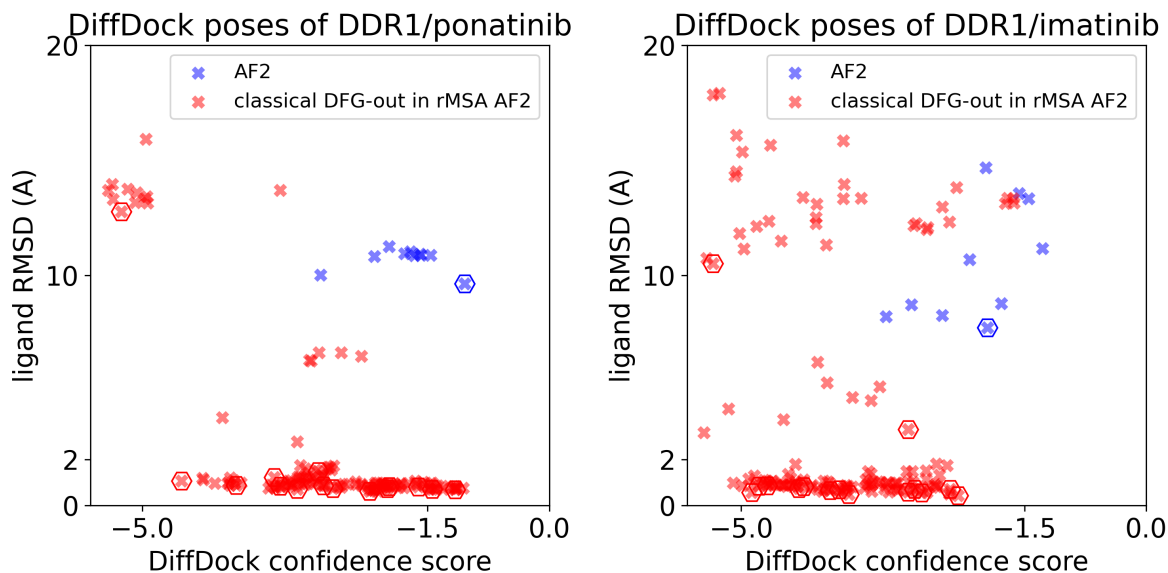
Figure S20: Ligand RMSDs are plotted against the DiffDock confidence scores for the Diff-Dock poses of type II inhibitors (ponatinib and imatinib) against DDR1 AF2 structure (blue) or the classical DFG-out structures in DDR1 rMSA AF2 ensemble (red). The pose with the lowest ligand RMSD from each input structure is marked by hexagon.

Table S1: Confidence score for the DiffDock pose aligns with AF2RAVE pmf values. The DiffDock confidence score of the pose with the lowest ligand RMSD (marked in red/bold) from each classical DFG-out structure in DDR1 rMSA AF2 ensemble is compared with the AF2RAVE pmf value for corresponding strutcure (marked in red/bold).

AF2RAVE pmf (kJ/mol)	Lowest ponatinib ligand RMSD (Å)	DiffDock conf_score	Lowest imatinib ligand RMSD (Å)	DiffDock conf_score
0.37	0.71	**-1.44**	0.55	-2.77
0.57	0.70	**-1.16**	0.43	-2.32
1.13	1.48	-2.84	3.31	-2.94
1.45	0.88	-3.85		-1000
1.62	0.82	-1.61	0.70	-2.42
2.07		-1000		-1000
2.69	0.62	-2.21	0.59	-2.94
3.82	0.86	-2.78	0.83	-4.23
4.02	0.71	-3.1	0.75	-4.29
4.21	0.82	-1.97	0.70	-2.86
4.39	12.77	-5.26	10.52	-5.35
4.70	0.85	-3.31	0.66	-3.9
7.60	1.22	-3.38	0.86	-4.68
N/A	0.75	-1.99	0.49	-3.67
N/A	0.75	-2.66	0.69	-3.79

20

References

(1) Vani, B. P.; Aranganathan, A.; Tiwary, P. Journal of Chemical Information and Modeling **2023**,

(2) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Proteins: Structure, Function, and Bioinformatics **2010**, 78, 1950–1958.

(3) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. The Journal of chemical physics **1983**, 79, 926–935.

(4) Hanson, S. M.; Georghiou, G.; Thakur, M. K.; Miller, W. T.; Rest, J. S.; Chodera, J. D.; Seeliger, M. A. Cell chemical biology **2019**, 26, 390–399.

(5) Mirdita, M.; Schütze, K.; Moriwaki, Y.; Heo, L.; Ovchinnikov, S.; Steinegger, M. Nature methods **2022**, 19, 679–682.

(6) Wayment-Steele, H. K.; Ojoawo, A.; Otten, R.; Apitz, J. M.; Pitsawong, W.; Hömberger, M.; Ovchinnikov, S.; Colwell, L.; Kern, D. Nature **2024**, 625, 832–839.

(7) Corso, G.; Stärk, H.; Jing, B.; Barzilay, R.; Jaakkola, T. arXiv preprint arXiv:2210.01776 **2022**,