Benchmarking Refined and Unrefined AlphaFold2 Structures for Hit Discovery

Yuqi Zhang[1], Marton Vass[2], Da Shi[1], Esam Abualrous[3], Jenny Chambers[4], Nikita Chopra[1], Chris Higgs[1], Koushik Kasavajhala[5], Hubert Li[1], Prajwal Nandekar[5], Hideyuki Sato[6], Edward B. Miller[4], Matt Repasky[7], Steven V. Jerome[1]*

[1] Schrödinger Inc. 10201 Wateridge Circle, Suite 220, San Diego, CA 92121, United States
[2] Schrödinger Technologies Limited, Nine Hills Road, Cambridge, CB2 1GE, United Kingdom
[3] Schrödinger GmbH, Glücksteinallee 25, 68159 Mannheim, Germany
[4] Schrödinger Inc., 1540 Broadway, 24th Floor, New York, New York 10036, United States
[5] Schrödinger India Private Limited, No: #147, 3rd Floor, Jawaharlal Nehru main road, Channasandra 5th stage Rajarajeswari Nagar, Above State Bank of India, Bengaluru, India, 560098
[6] Schrödinger KK, 13F Marunouchi Trust Tower North, 1-8-1 Marunouchi, Chiyoda-ku, Tokyo, 100-0005, Japan
[7] Schrödinger Inc., 101 SW Main Street, #1300, Portland, Oregon, 97239, United States

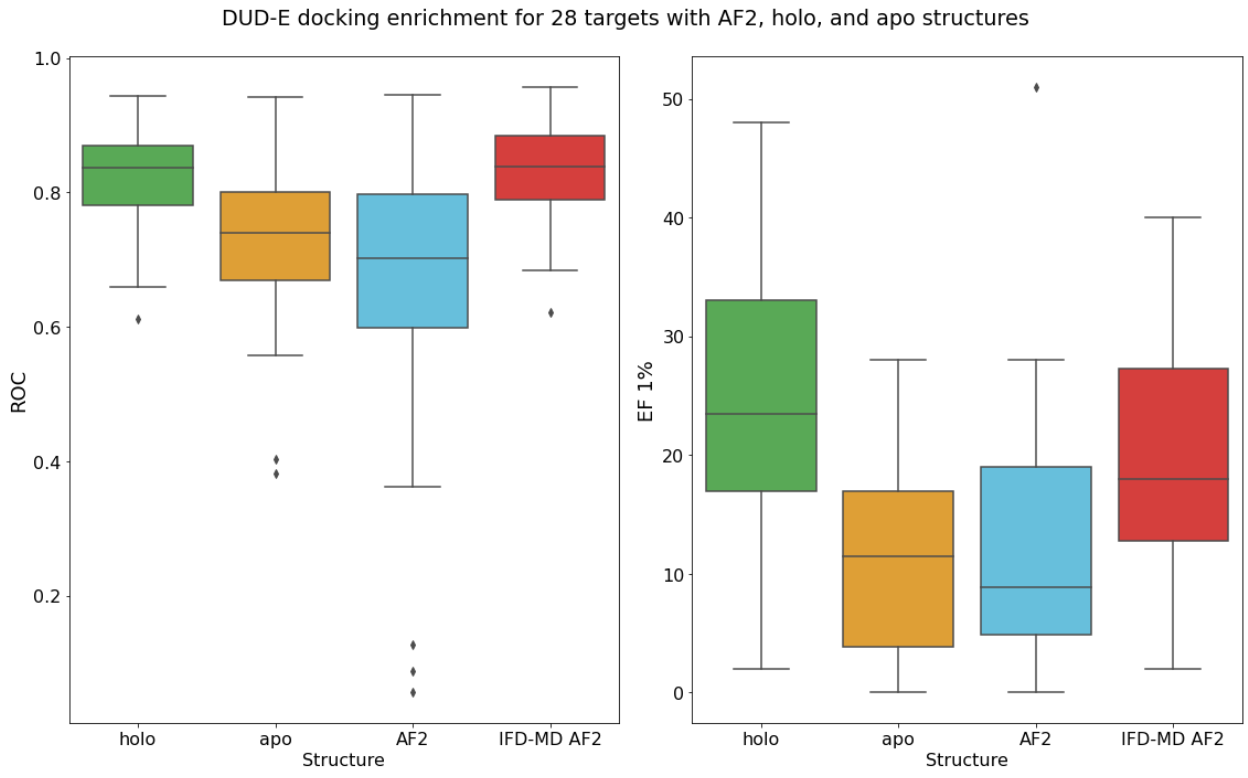
*Corresponding author: Steven Jerome: steven.jerome@schrodinger.com;

KEYWORDS

AlphaFold; Docking; Glide; Virtual screening; Induced-fit docking; Hit discovery.

ABSTRACT

The recently developed AlphaFold2 (AF2) algorithm predicts proteins' 3D structures from amino acid sequences. The open AlphaFold Protein Structure Database covers the complete human proteome. It shows great potential to provide structural information to enable and enhance existing and new drug discovery projects. Using an industry-leading molecular docking method (Glide), we benchmarked the virtual screening performance of 28 common drug targets each with an AF2 structure and known *holo and apo* structures from the DUD-E dataset. The AF2 structures show comparable early enrichment of known active compounds (avg. EF 1%: 13.16) to *apo* structures (avg. EF 1%: 11.56), while falling behind early enrichment of the *holo* structures (avg. EF 1%: 24.81). We also demonstrated that with the IFD-MD induced-fit docking approach, we can refine the AF2 structures using a known binding ligand to improve the performance in structure-based virtual screening (avg. EF 1%: 19.25). Thus, with proper preparation and refinement, AF2 structures show considerable promise for *in silico* hit identification.



1. INTRODUCTION

Computational methods play an increasingly important role in the hit discovery phase of the drug discovery process. Historically, these methods could be broadly categorized into two distinct classes: *structure-based* and *ligand-based* methods. The choice of method is constrained by the data available at the start of a new hit identification campaign. Structure-based approaches can be enabled with a single high quality 3D atomistic model of the protein of interest. These methods can typically leverage additional ligand information, when available, for model validation or parameterization, but such data is not generally required. Ligand-based approaches are enabled with one or more ligands with sufficient on-target potency and functional activity ('actives'). Features are extracted from the set of actives and can be used to identify virtual hit compounds. Where no previous ligands have been reported, structure-based approaches such as molecular docking allow for identification of hits, generation of putative binding modes, and insights into important interactions that can be leveraged to guide optimization of these hits. Docking-based studies have been used to great effect recently to identify hit compounds against COVID-19 protein targets[1–5] and have a long track record of identifying hit and tool compounds to progress projects[6,7]. For projects pursuing a best-in-class approach in a crowded competitive environment, structure-based approaches offer a way to explore new areas of chemical space, without being constrained by the space of known actives. The advent of ultra-large, make-on-demand libraries has opened new chemical space for molecular docking[6,8–11]. As such, structure-based hit identification methods are a critical component of the modern *in-silico* toolkit. Here, we focus on the structural enablement of physics-based molecular docking in drug discovery.

The domain of applicability of molecular docking has been increasing rapidly as new structures are reported, aided considerably in recent years by the emergence of high quality cryo-EM structures[12–14] and refinement protocols[11,15] According to one analysis[16] nearly 74% of the human proteome has at least one representative structure in the public domain. While this represents significant progress, there remain categories of targets, chiefly GPCRs, where coverage remains low. Structural enablement of "dark" regions of the proteome may evade current techniques or require considerable experimental expense over a long period of time. For such targets, validated computational models can facilitate successful hit discovery campaigns in place, or in advance of, an experimentally derived structure. Thankfully, the difficulty of this problem can be significantly reduced using template proteins, mined from public or structural private repositories. Structural elements of template sequences can be incorporated into the initial model, leaving only "gap-regions" or regions of sequence space without a nearby homologue, to be predicted using physics-based or data-based approaches. The number and size of gap regions will depend on the availability of suitable templates for the target. The selection of templates by an expert is a time-consuming process and must be repeated for each new target sequence.[17] The number of templates a

human expert can reasonably handle is another limitation of this approach. Skolnick and coworkers suggested[18] in 2005, that the problem of protein structure prediction could be solved by the structural information in the PDB at the time – one simply needed an algorithm capable of identifying the optimal protein fold from among all possible templates and the optimal alignment to the target. This algorithm would need to be able to consider many more potential templates than a human expert could reasonably handle and produce a capable scoring model for evaluating potential alignments.

A deep learning-based approach could, in principle, efficiently make use of all structural information available to solve the problem. Rather than predict each new structure as an entirely separate problem, a single pass of training could optimize parameters assignment for thousands of model predictions simultaneously. DeepMind's AlphaFold2 (AF2)[19] provided powerful validation of this approach, setting a new standard for protein structure prediction in the biennial CASP14 competition. AF2 combines multiple sequence alignments (MSA) with a novel attention mechanism, called EvoFormer, to focus the training on portions of the sequence. These components are connected to a structure prediction module that considers rotations and translations of individual residues and a custom loss function that places special emphasis on the correct orientation of residues in order to capture important interatomic interactions. The authors used the fully trained AF2 network to make predictions on a set of ~350,000 proteins, which were subsequently made available free of charge to researchers. The success of AF2 has generated considerable excitement in the community as evidenced by the nearly 2,000 citations garnered by the 2021 Nature publication. We expect iterations of the methodology and competing approaches to improve the performance of deep-learning based approaches. Baker and coworkers, inspired by AlphaFold2, proposed a "three-track" method that more tightly integrated sequence (1D), distance-matrix (2D) and cartesian coordinates (3D) information during training.[20] The authors note that the performance of the method is currently limited by available computational power. It remains for the community to validate the use of these deep-learning derived structures in drug discovery applications. In this work, we explore the potential of AF2 specifically to enable "dark" protein targets for structure-based virtual screening.

Rigid receptor docking protocols (e.g. GOLD, Glide, AutoDock Vina)[21–23] perform essentially no sampling or refinement of the binding site structure during the docking calculation, ignoring induced fit effects. We have argued previously that these limitations can be addressed by docking into an ensemble of protein conformations[24,25]. AF2 predictions are essentially *apo* structures of the complete protein sequence. As there is no explicit information included of ligands or cofactors, these structures often have unstructured regions of low predicted confidence occluding potential binding sites. AF2 offers no way to use a bioactive conformation of a known binder to induce an alternative conformation of the protein, creating an opening for induced fit docking protocols. Im and coworkers recently demonstrated modest

improvements in virtual screening performance using an initial template-based alignment combined with restrained MD in order to refine *apo* structures[25]. This approach increased enrichment from EF 1% of 3.5 to 6.2 on a 40 target subset of DUD-E, a well known dataset for benchmarking molecular docking programs[26]. This compares with an EF 1% of 10.0 for *holo* structures in the same study. Performance is likely to be limited by the availability of an adjacent *holo* template in the PDB.

In this work, we assess the docking performance of Glide across a range of targets using the DUD-E dataset. We compare enrichment performance with truncated AF2 structures to experimental *apo* and *holo* structures. As seen in Table 1, native Glide SP delivers considerably higher enrichment over the DUD-E dataset using *holo* structures as compared to the enrichments reported in the study of Im and coworkers[25], where AutoDock[21] was used to generate the docked poses. Next, we use our recently released induced-fit docking program (IFD-MD)[26,27] to induce a more *holo*-like structure suitable for docking using the *holo* ligand provided in the DUD-E dataset and repeat the docking calculations on this structure. IFD-MD is a rigorous, template-free physics-based protocol that combines traditional docking calculations, pharmacophore-based analysis of the active site, detailed analysis of the active site water network using WScore, side chain sampling, and MD-based refinement, to capture induced fit effects upon ligand binding[27]. IFD-MD structures have been validated for use in free energy calculations across a wide range of targets[27,28]. The current IFD-MD protocol was targeted toward systems where cross-docking can succeed without the need for large backbone motions. For applications where large backbone motions are needed, the authors of the IFD-MD publication suggest that further development of the method to include accurate prediction of loop-regions would be necessary to broaden the domain of the method to these cases. Here, we assess the potential to induce more *holo*-like structures for molecular docking from the AF2 *apo* predictions using IFD-MD. Given the limitations noted above, this approach is likely to be successful only if the protein backbone in the active site is essentially in a correct conformation. Our preliminary findings are promising, showing performance of the IFD-MD refined AF2 structure coming close to the *holo* structure. Docking into an ensemble composed of the top 5 poses from IFD-MD refined AF2 structures, we see a roughly 45% improvement in early enrichment, as measured by BEDROC (α=160.9). No fitting or reparameterization of the IFD-MD protocol was required. It is also important to acknowledge the limitations of this work. All the targets in the DUD-E dataset have representative structures deposited in the PDB prior to the training data of the AF2 model used to generate the structure predictions. In a future work, we will investigate this approach with additional targets that would not have been included in the training of the AlphaFold model. Nonetheless, these preliminary results show great promise for the use of AF2 to structurally enable additional targets for molecular docking campaigns[26,27].

2. METHODS

We used a set of well validated methods from the Schrödinger Suite[29] (2021-4 release) for this benchmark study. To reduce any potential bias introduced by different settings in each method, we chose to use the default settings for most of the applications unless otherwise specified. It's important to note that this sets a baseline for early enrichment that is likely to be lower than that achievable by judicious use of experimental observations or expert intuition in the docking experiment.

2.1 Dataset

The DUD-E set is a dataset widely used for benchmarking virtual screening methods. The DUD-E set has 102 targets from diverse protein classes including kinases, nuclear receptors, proteases and several other protein classes. For each target, DUD-E provides a set of diverse active ligands and a set of decoy ligands within a similar chemical space[26]. In the DUD-E set there are 40 targets from the original DUD set with an assigned *apo* structure. Among the 40 targets, there are two HIV viral protein targets where there is no native prediction in the AlphaFold Database as of April, 2022. Among the remaining 38 targets, there are 6 targets (*ALDR, CAH2*, *COMT, DEF*, *DYR*, and *PA2GA*) with cofactors in the binding site of the associated *holo* structure in the dataset, which can be crucial for protein-ligand interactions. Three targets contain metals, two have organic cofactors, and one has both. AF2 structures do not have any metal atoms or cofactors modeled in the native predictions, thus we analyzed these 6 targets separately. *CASP3* is a target with a covalently bound inhibitor and is beyond the scope of this study. We also remove three other targets, *CSF1R*, *KIT,* and *RENI*, from the main analysis as there are severe clashes between the *holo* ligand and the AF2 structures in the binding site. After this filtering, 28 targets from the DUD-E dataset were the focus of our analysis. Please refer to the Supplementary Information Table S1 for a full list of the targets investigated.

2.2 Structure preparation

For each target, we downloaded the DUD-E specified *holo* and *apo* structures from the PDB[30] and the AF2 structure from the AlphaFold Protein Structure Database[31]. The AF2 model predicts a structure for the complete sequence of a genome and the resulting structure may exist in a different bio-unit composition as compared to the relevant *holo* or *apo* structures from the PDB. Fig. 1 shows the example of *BRAF*, where the AF2 structure contains almost three times more amino acids than the *holo* structure. More importantly, the excess sequence can interfere with the binding site, either by preventing docking or by introducing artificial interactions between the protein and docked ligands.

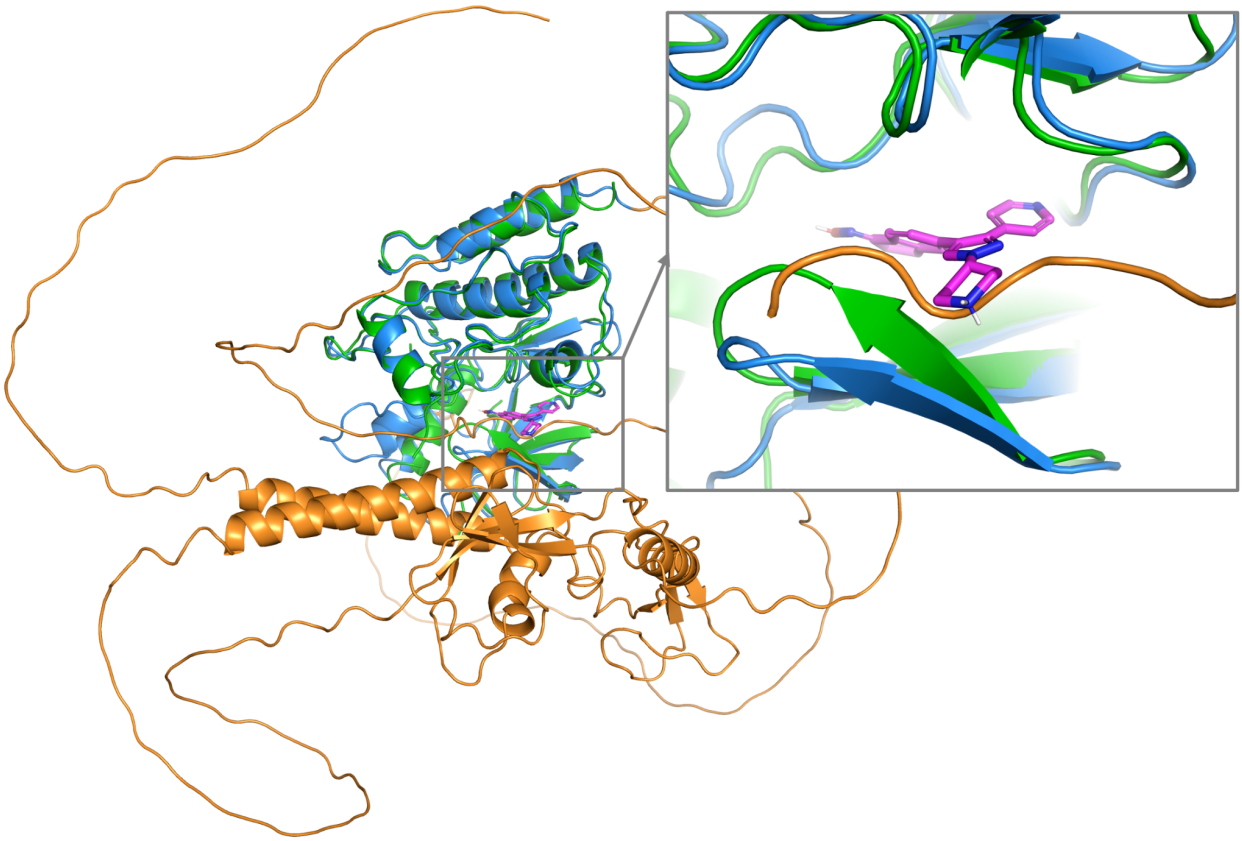


Figure 1. AF2 predicted *BRAF* structure (blue and orange ribbon) aligned to the *holo* structure (green ribbon, PDB: 3D4Q). The ligand in the relevant binding site is shown in magenta. The AF2 structure has a sequence length of 766, while the *holo* structure has 264 amino acids in the structure. The extra amino acids (orange ribbon) from the AF2 structure severely block the binding site, preventing Glide from obtaining reasonable docked poses. The orange region of the AF2 structure was truncated to recover the binding site.

To recover the AF2 binding site blocked by the excess amino acids, we performed a sequence alignment between the AF2 structure and the *holo* structure. Then, we removed the excess amino acids within the AF2 structure before the first and after the last amino acid in the sequence of the associated *holo* structure. After the removal of these amino acids, we recovered a reasonably sized binding site for *BRAF*, *EGFR*, *IGF1R*, *ITAL*, *RXRA,* and *THRB*. Three targets removed from the main analysis, *CSF1R*, *RENI,* and *KIT*, still had additional loops occupying the binding sites, which are not present in the *holo* structure. These AF2 structures after removal of excess amino acids are referred to as the "AF2 truncated structures".

The *holo*, *apo*, AF2 full, and AF2 truncated structures were prepared with the Protein Preparation Wizard[28,32] from the Schrödinger Suite. The protein protonation state is assigned with PROPKA[33,34] at pH 7.4. DUD-E actives and decoys were processed using the LigPrep program of Schrödinger Suite starting from a SMILES representation of each molecule, using the following settings: (1) Tautomers were generated using Schrödinger's Epik with a target pH of 7.0 +/- 1.0 and (2) a maximum of 32 stereoisomers were generated for each molecule.

For the cofactor containing targets (*ALDR, CAH2*, *COMT, DEF*, *DYR*, and *PA2GA*), the cofactors were transposed from the *holo* structure to the truncated AF2 structure after alignment of the two protein structures. Severe steric clashes were resolved by selecting a different rotamer of the clashing residue. For *CAH2* the metal binding histidines were in the HID94, HID96, HIE119 tautomer states, and for *DEF* the metal binding histidines were in the HID133 and HID137 tautomeric states, and Cys91 was deprotonated. Other residues were unchanged and metal coordinating waters were not added. Investigation of the accuracy of *de novo* cofactor placement methods was outside of the scope of this study. The manually transposed structures were used directly in docking to evaluate the effect of the presence of the cofactors on VS performance of the AF2 structures, and were also used as starting structures for IFD-MD calculations. The IFD-MD output was used for docking either with the cofactors present (to compare to the *holo* structure) or with the cofactors removed (to compare with the truncated AF2 structure). For the *CAH2* target, metal binding states in ligand preparation were added and the grid size was increased due to a very small reference ligand in the *holo* structure.

2.3 IFD-MD refinement of the AF2 structures

To refine and optimize the AF2 structures, we used the recently developed IFD-MD program in the Schrödinger Suite. IFD-MD combines shape-based ligand alignment with QSAR atom typing [35], rigid receptor docking (Glide[23] and WScore[24]), and energy-guided protein structure refinement (Prime[36]) with explicit solvent molecular dynamics simulations to address the induced-fit effect[27]. IFD-MD requires a protein-ligand complex as the template. In this study, we investigated using different templates as described in section 3.3.

With the refined poses obtained from IFD-MD, we removed the docked ligand and solvents and followed the steps described in section 2.5 to generate the grid and perform the docking. We also evaluated the VS enrichments using an ensemble docking approach. For each target, we generated 5 docking grids using the receptor structures from the top 5 IFD-MD refined poses. The actives and decoys are docked into each of the 5 grids and later combined where only the pose with the best docking score is kept for each ligand.

2.4 Binding site analysis

We used SiteMap[37] in the Schrödinger Suite to evaluate and compare the binding site quality of the *apo*, *holo*, AF2 full structure, AF2 truncated structure, and IFD-MD structure of each target. For each target, the five structures were first aligned based on the whole protein structure. The alignment was refined by the binding site residues which were defined by the co-crystalized ligand in the *holo* structure. The refined alignment was used to have a consistent definition of the binding site. Next, we analyzed the binding site properties (volume and SiteScore properties) of each structure with SiteMap[37,38] using the ligand in the *holo* structure as the reference ligand and a 5 Å distance threshold. For structures where multiple sites were identified in the 5 Å box around the ligand, the site with the largest *SiteScore* was selected. SiteScore was calculated via Equation 1 below, where *n* is the number of site points, *e* is the enclosure score, and *p* is the hydrophilic score.

$$SiteScore = 0.0733\sqrt{n} + 0.6688e - 0.2p$$

Equation 1. Equation for SiteScore.

We also compare the binding site residue conformations in the five structures by calculating the pairwise all atom RMSD between binding site residues.

2.5 Molecular docking and enrichment metrics

Glide[23,39] is a leading molecular docking program. Here, we use the enrichment results derived from Glide docking as an indication of the usefulness of different structures for virtual screening. Like other docking methods, Glide utilizes a docking grid to increase throughput. In our experiments, the grid box center is defined as the center of mass of the reference ligand in the *holo* structure. The gridbox size is defined as $16\ \text{Å} + 0.8 \times diameter$, where $diameter$ is the maximum distance between any two atoms in the *holo* ligand.

Glide docking was performed in the SP (standard precision) mode. We calculated the area under the receiver-operator characteristic (ROC) curve, Boltzmann-enhanced discrimination of receiver operating characteristic (BEDROC)[40] with α=160.9, and EF 1% using the *enrichment.py* script within the Schrödinger Suite to evaluate the performance of the virtual screening. Virtual screening has been described as an early enrichment problem[40] as true positives recovered early in the ranked list are more valuable given the costs associated with compound acquisition and validation. The BEDROC and EF 1% metrics are more sensitive to early enrichment performance, as compared to popular enrichment measures used in other contexts such as ROC and are better suited for evaluating virtual screening performance.

3. RESULTS and DISCUSSIONS:

AlphaFold2 predicted structures cover most common drug targets. Here, we benchmark the virtual screening performance using AF2 structures in structure-based virtual screening methods, a common approach for hit identification in a drug discovery project. In the following sections, we will describe: (1) binding site quality comparison between *holo*, *apo,* and AF2 structures; (2) molecular docking-based enrichment using AF2 structures; (3) molecular docking enrichment for IFD-MD refined AF2 structures, and (4) challenges and limitations of using AF2 structures.

3.1 Binding site assessment

We analyzed the binding site properties and conformations of *holo, apo*, and AF2 structures of each target. Key binding site properties (volume and SiteScore) were evaluated using SiteMap. The *SiteScore* evaluates whether a site is likely to produce high-affinity ligands, and larger scores correspond to higher probability of having tighter binders.

Fig. 2 shows the comparison of *SiteScores* and volumes of binding sites on AF2 full, AF2 truncated, *apo*, and *holo* structures. Overall, the *holo* structures have slightly higher SiteScores and larger volumes. We have previously shown that Glide enrichment performance can be improved by minimization of the input receptors for docking, generally increasing the binding site volume. An increase in the available space for docking may explain the superior enrichments derived from docking into *holo* structures. This hypothesis would need to be further investigated to demonstrate generality. The *SiteScore* and volumes of binding sites on AF2 structures generally lie between the *apo* and *holo* structures, possibly because both *holo* and *apo* structures were included in the training of the AF2 model. As shown in Fig. 2, the full AF2 structures have slightly better *SiteScore* than the truncated AF2 structures. One plausible explanation is that in the full structure the presence of a terminus structure near the binding site may create a more enclosed or better defined binding site. A more complete prediction of the structure of the binding site may not necessarily produce a more useful structure for virtual screening. After IFD-MD, the AF2 structures show comparable SiteScores and volumes as the *holo* structures.

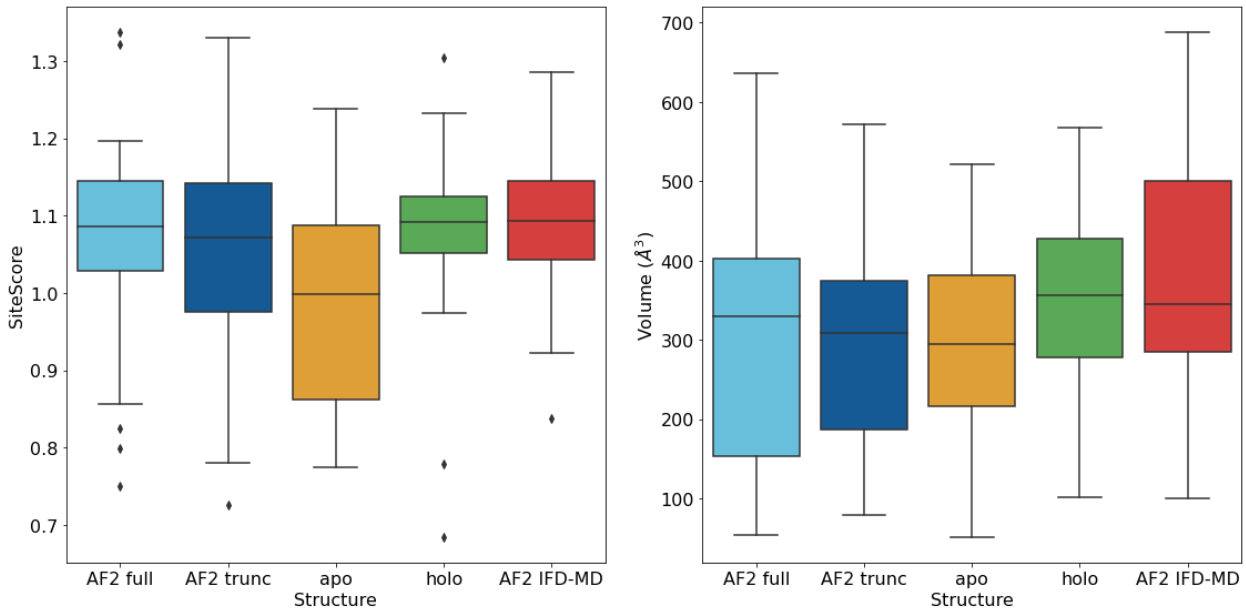


Figure 2. Comparison of *SiteScores* and volumes of binding sites on AF2 full, AF2 truncated, *apo*, *holo*, and the top poses of IFD-MD refined AF2 structures. The AF2 binding sites have SiteScores in between *holo* and *apo* binding sites, while the IFD-MD refinement improves the SiteScore of the AF2 structures.

A similar trend was observed in the volume comparison. A target-specific volume comparison is shown in Fig. S1. For most of the targets, the AF2 full and truncated structures have binding sites with similar sizes, as the terminus structures in the AF2 full structure are far from the binding sites. For structures where the terminus structures are inside the binding sites, one would expect that truncating them would make the binding sites larger, which is the case for *THRB* and *BRAF*. Three targets (*CSF1R*, *KIT,* and *RENI*) on our list have additional loops occupying the binding site, leading to significantly lower volumes than the *apo* and *holo* structures (Fig. S2), therefore we removed these from the main analysis and virtual screening study.

3.2 Benchmark of VS performance of AF2 structures

With the active/decoy ligand collections from the DUD-E dataset, we tested the molecular docking enrichment for *holo*, *apo*, full AF2, and truncated AF2 structures. The results are summarized in Table 1. On average, the truncated AF2 structures have higher enrichments than the *apo* structures, while being lower than the *holo* structures where ligands are present. The *apo* structures have the lowest enrichment overall. This is consistent with our expectations as binding sites can reorganize with no ligand present, often folding into different binding site environments. AF2 models are trained on the entire PDB database (as of 30 April 2018) with both *holo* and *apo* structures present, and we expect the AF2 binding

sites to have features resembling those of both the *holo* and *apo* structures. More active ligands can be docked into the truncated AF2 structures than the original AF2 structures as shown in Table 1. From these results the AF2 structures appear to have more *apo* than *holo* characteristics in binding site structural prediction.

Table 1. Average molecular docking enrichment on *holo*, *apo*, full AF2, and truncated AF2 structures.

Averaged over 28 targets	ROC	BEDROC (α=160.9)	EF 1%	Percentage of Actives docked
holo	0.814	0.451	24.81	98.8%
apo	0.722	0.218	11.56	96.4%
AF2 full	0.647	0.245	13.16	84.1%
AF2 truncated	0.715	0.239	13.16	94.5%

Fig. 3 shows the BEDROC (α=160.9) for the 28 targets. There are a few targets where the truncated AF2 structure enrichment is significantly lower than the *holo* structure. A plausible explanation for this observation is that the binding sites in these AF2 structures have different loop conformations and/or side chain conformations compared to the *holo* receptors. In such cases, an induced-fit docking approach should be able to refine the AF2 structures to physically relevant *holo*-like conformations and lead to potential improvements in enrichments as described in the following section. Another possible reason is that the target protein can be found in different states in *holo*/*apo* and AF2 structures, like DFG-in and DFG-out states for kinases. In the example of *MET*, the *holo* structure has a DFG-out conformation and the AF2 structure is in DFG-in conformation. In such cases, a better strategy could be to apply the knowledge of the protein class and perform a partial homology modeling on the binding site as discussed in section 3.4.

3.3 Refining AF2 structures with IFD-MD

The results in the previous section suggest that refining the AF2 structures is desired to reach their full potential for virtual screening. IFD-MD combines pharmacophore modeling, molecular docking, and molecular dynamics to predict the induced-fit binding pose. Here we ran IFD-MD on the 28 targets and investigated the VS performance of using the top-ranked IFD-MD pose as well as an ensemble docking approach with the top 5 IFD-MD poses.

IFD-MD requires a complex structure as the template for the refinement. While there is no obvious way to generate a protein-ligand complex with the AF2 structures, we tested two approaches, superimposing the *holo* ligand in its bound coordinates onto the AF2 structures and using the docked pose of the *holo* ligand into the AF2 structure. The first approach generates a complex template that resembles the *holo* binding pattern, which makes it the best scenario for IFD-MD. The latter approach is less biased and represents a scenario where the binding patterns are completely unknown.

For the first approach, we placed the *holo* ligand into the truncated AF2 structures using the coordinates in the aligned *holo* receptor. Such ligand placements often result in clashes between the ligand and the AF2 structures, which can be resolved later in the IFD-MD refinement. Using either the single top-ranked refined structure, or the five top-scoring refined structures from IFD-MD, we performed molecular docking with the active and decoy ligands from the DUD-E dataset. The results are shown in Table 2. If we use the top-ranked pose from IFD-MD, the average ROC improved from 0.715 to 0.798 and BEDROC (α=160.9) and EF 1% improved as well. To fully utilize the IFD-MD results, we also tested an ensemble docking approach where ligands are docked independently into the top 5 receptor conformations from the IFD-MD. Previously we have reported the benefit of ensemble docking when care is taken to balance docking scores across the ensemble.[24] Then the 5 docking results are merged and ranked based on the best docking score per ligand. These results are shown in Table 2. The average ROC further improved to 0.826, better than the average ROC for the *holo* structures. We show the target specific BEDROC comparison in Fig. 3 and discuss a case study of *EGFR* in the following paragraph.

Table 2. Comparison of VS enrichments for IFD-MD refined AF2 structures with *holo* and the truncated AF2 structures.

Averaged over 28 targets	ROC	BEDROC (α=160.9)	EF 1%
holo	0.814	0.451	24.81
AF2 truncated	0.715	0.239	13.16
IFD-MD refined AF2 (top pose)	0.798	0.310	17.23
IFD-MD refined AF2 (5 poses)	0.826	0.348	19.25

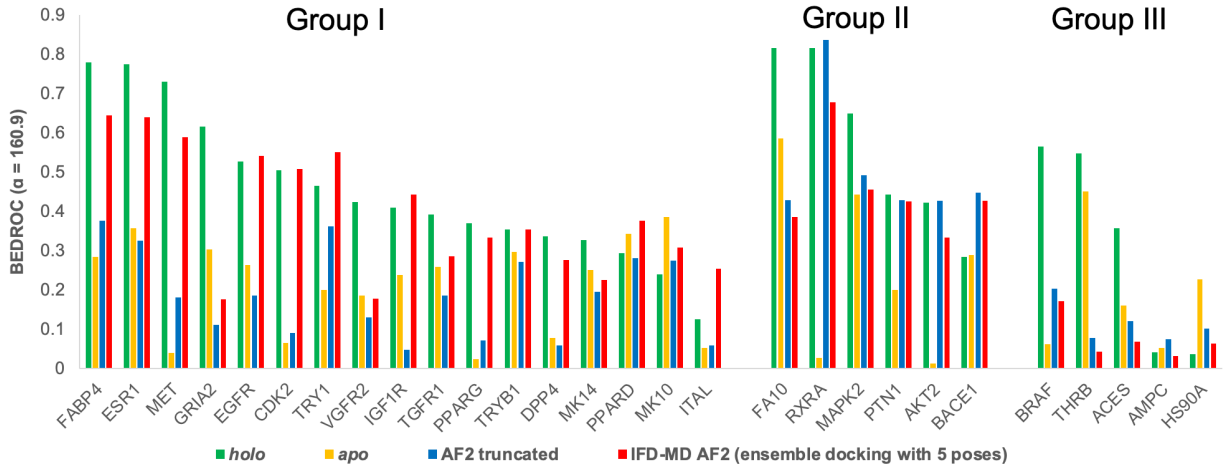


Figure 3. Target specific BEDROC comparison. IFD-MD refined structures have significantly improved enrichments over the original AF2 structures. Group I on the left side shows targets where the IFD-MD refinement (red) improves the VS enrichment over the truncated AF2 structures (blue). For targets in Group II in the middle, the truncated AF2 structures have comparable enrichments with the *holo* structures (green) and the IFD-MD refined structures perform similarly. Targets in Group III on the right are targets where the truncated AF2 structures have low enrichments and the IFD-MD refinement is not able to improve the VS performance. *Apo* structures are also included for reference (yellow).

EGFR was one of the targets in our data set where binding site refinement by IFD-MD had a significant effect on virtual screening performance of the AF2 structure. The *apo, holo*, and AF2 structures exhibited several structural differences affecting the performance (see Fig. 4). In the *apo* structure the C-helix is in a different orientation relative to the *holo* structure, with Glu762 and Met766 clashing with the ligand position seen in the *holo* structure. The activation loop is also in a different conformation in these structures, and Phe856 of the DFG motif in the *apo* structure is also clashing with the ligand position, hence the observed lower enrichment (BEDROC = 0.262). The C-helix and activation loop in the AF2 structure occupy an intermediate position between those in the *apo* and *holo* structures, with the same residues and additionally Cys775 and Thr790 forming suboptimal contacts with the ligand pose from the *holo* structure, hence a similar enrichment to the *apo* structure (BEDROC = 0.185). In the full AF2 structure, the low confidence C-terminal also overlaps with the piperidine moiety of the ligand in the binding site, rendering the enrichment very low (BEDROC = 0.023).

During IFD-MD refinement Met766 in the C-helix moves to a different rotamer state to create space for the ligand and Glu762 is pushed further away from the binding site in a similar rotamer. Phe856 in the DFG motif moves away as a rigid body, while Asp855 assumes a different rotamer position. The Thr790 gatekeeper residue and Cys775 are both moved into the rotamer state observed in the *holo*

structure. IFD-MD refinement brings about both an expansion of the binding site and rotamer state changes of multiple residues. The top pose from IFD-MD shows a flipped indazole ring relative to the experimental binding mode and a moderately increased early enrichment (BEDROC = 0.379), while the second IFD-MD pose corresponds to the experimental binding mode. Thus ensemble docking into multiple IFD-MD structures brings the virtual screening enrichment of the AF2 structure (BEDROC = 0.541) to a similar level to that of the *holo* structure (BEDROC = 0.526).

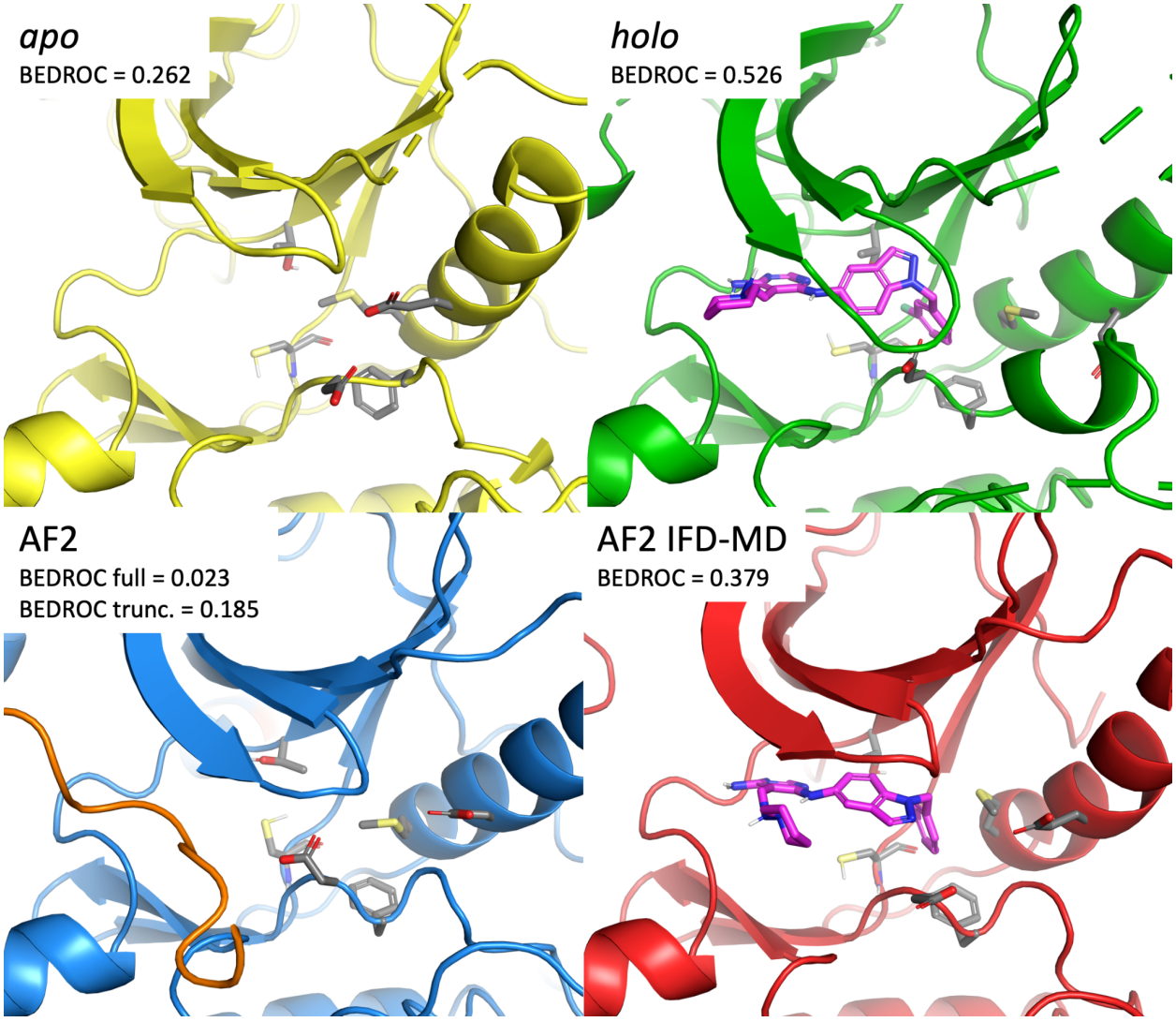


Figure 4. Binding site comparison and early enrichments of the *apo* (yellow), *holo* (green), AF2 (blue, truncated region shown as orange ribbon), and IFD-MD refined AF2 structures (red) for *EGFR*. Ligand shown as magenta sticks in the experimental binding mode in the *holo* structure, and the top predicted binding mode in the IFD-MD refined AF2 structure. Residues with significantly different orientations affecting the ligand binding mode are highlighted as gray sticks.

In the second approach using IFD-MD to refine AF2 structures, we docked the *holo* ligand into the AF2 structures with Glide and used the resulting pose as the template for IFD-MD. AF2 structures of *ESR1* failed to dock with the respective *holo* ligand with Glide due to dramatic binding site differences. Enrichment improvements in IFD-MD refined AF2 structures are similar to the first approach described earlier as shown in Table S2. When comparing the docked poses and the original *holo* ligand binding mode, the docked ligands do not always recover the same binding mode as in the *holo* structures. The docked ligands have RMSD lower than 3.0 Å with the reference *holo* ligand in only 5 targets. The pose accuracy is lower than a typical redocking experiment of a native ligand back into its native receptor. This suggests that for the purpose of refining the protein structure to enable virtual screenings in hit discovery, the enrichment performance is not sensitive to the choice of template complex used in IFD-MD.

To eliminate the selection bias introduced by the *holo* ligand, we tested the IFD-MD refinement workflow with an active ligand that is dissimilar to the *holo* ligand. A subset of targets was selected for this additional analysis, to limit computational costs. We developed a simple protocol for identifying a subset where *holo* structures were available in the PDB with a co-crystallized ligand dissimilar to active ligands in the DUD-E dataset. Using PLDB, a component of the Schrodinger Suite, we queried the entire PDB with the following query settings: (1) *holo* structures only, (2) ligand binding affinity better than 1μM, and (3) electron density map(s) available. The results of this query were processed in the following way: First, the PLDB-identified ligand was extracted. Next, RDKit fingerprints[41] were generated using default settings. The ligand choices would likely have been impacted by the choice of fingerprint featurization. Nevertheless, no attempt was made to optimize the fingerprint type. Default RDKit fingerprints are used for convenience, given the popularity of the open source cheminformatics package. RDKit fingerprints are related to Daylight-style fingerprints, but incorporate several significant changes. Then the maximum Tanimoto similarity between the query and the set of all DUD-E active ligands available for a given target was calculated. Targets where the maximum similarity to any ligand of the DUD-E actives set was less than 0.5 were considered. Finally, a selection of these targets were made to ensure target diversity. We selected *MK10*, *PPARD,* and *TYRB*. Together, they comprise the "diversity subset" for subsequent analysis. For the targets in the diversity subset, we searched the PDB for complex structures where the ligands are dissimilar to the assigned *holo* ligand from the DUD-E dataset. We repeated the IFD-MD refinement workflow using the alternative PDB structure. The results are shown in Table 3. This analysis suggests that the improved enrichments using IFD-MD refined structures are not merely a result of fitting to specific ligands in the DUD-E dataset.

Table 3. VS enrichment for AF2 structures refined by IFD-MD using reference ligands topologically dissimilar from the DUD-E ligands.

Averaged over *MK10*, *PPARD*, and *TYRB*	ROC	BEDROC (α=160.9)	EF 1%
holo	0.831	0.296	16.33
apo	0.809	0.341	17.33
AF2 truncated	0.788	0.253	12.67
IFD-MD refined AF2 (top pose)	0.864	0.421	22.00
IFD-MD refined AF2 (5 poses)	0.888	0.517	27.67
IFD-MD refined AF2 (top pose) with DUD-E ligand	0.814	0.307	14.33
IFD-MD refined AF2 (5 poses) with DUD-E ligand	0.847	0.406	14.67

AF2 structures showed great potential for hit discovery after IFD-MD refinement. We wanted to estimate whether IFD-MD refinement can also enhance the performance of virtual screening of the *apo* structures. Deposited PDB structures often have missing segments and sometimes those missing segments prevent the application of all-atom molecular dynamics methods, such as IFD-MD. Among the 28 targets in this study, we found 9 targets without such missing segments: *AMPC*, *DPP4*, *IGF1R*, *ITAL*, *PPARG*, *PTN1*, *TRY1,* and *TRYB1*. We superimposed the *holo* ligand in its bound coordinates onto the *apo* structures as the template for IFD-MD. The virtual screening results from this experiment are shown in Table 4. Relative to the *apo* structures for these 9 targets, IFD-MD refinement improves the average EF 1% from 7.31 to 10.34 if only the top-ranked IFD-MD pose is used, and to 13.44 if the best-performing of top 5 IFD-MD poses are considered. Meanwhile, the IFD-MD refined AF2 structures have an average EF 1% of 15.59 with the top pose and 19.00 with top 5 poses, surpassing the average EF 1% of the *holo* structures for these 9 targets.

Table 4. Comparison of the VS enrichment for IFD-MD refined *apo* structures with *holo*, *apo* and AF2 structures.

Averaged over 9 targets	ROC	BEDROC (α=160.9)	EF 1%
holo	0.786	0.318	17.31
apo	0.724	0.142	7.31
IFD-MD refined *apo* (top pose)	0.736	0.190	10.34
IFD-MD refined *apo* (5 poses)	0.802	0.247	13.44
AF2 truncated	0.692	0.171	9.10
IFD-MD refined AF2 (top pose)	0.824	0.287	15.59
IFD-MD refined AF2 (5 poses)	0.832	0.333	19.00

So far, we have focused on the 28 targets discussed above where refinement can be automated. We also tested the IFD-MD workflow on the six targets containing organic cofactor(s) or metal(s) in the binding site. Manual preparation of these structures was needed to place the cofactor or metal properly back into the AF2 structures as described in section 2.2. Virtual screening results shown in Table 5 indicate that the addition of the cofactors to the AF2 structures generally slightly improved virtual screening performance both overall and in terms of early enrichment. This is not surprising, since the cofactors are often involved in essential ligand interactions. In this limited data set the presence of metals was more important for good virtual screening performance than the presence of organic cofactors. It should also be noted that for *COMT* the transposition of the magnesium ion and the SAM cofactor decreased performance. Possibly in this case the simple transposition without refinement was not a good enough placement protocol for the cofactors. The IFD-MD refined structures stripped of cofactors exhibited already improved results relative to the original AF2 structures, but could not recapitulate the performance of the *holo* structures. However, the IFD-MD refined structures with the cofactors had similar or even improved performance relative to the *holo* structures especially in terms of early enrichment metrics. The seemingly higher performance of the IFD-MD refined structures over the *holo* structures originates from the *DYR* target, where the *holo* structure selected by DUD-E has a small co-crystallised ligand and a more enclosed binding site than some other *holo* structures of the target. For the *COMT* target the presence of cofactors improved virtual screening performance for the IFD-MD

structure as opposed to the AF2 structure, underlining the importance of refinement after cofactor transposition in this case.

Table 5. Comparison of the VS enrichment for metal- and cofactor- containing targets.

Averaged over 6 targets	ROC	BEDROC (α=160.9)	EF 1%
apo	0.643	0.087	5.27
holo	0.846	0.312	17.47
AF2 truncated	0.641	0.100	5.48
AF2 truncated + cofactor(s)	0.675	0.167	8.50
IFD-MD refined AF2 (top pose)	0.752	0.203	10.13
IFD-MD refined AF2 (top pose) + cofactor(s)	0.840	0.367	21.50

3.4 Limitations

Encouraging results were obtained for these well-established targets in the DUD-E data set, however, some challenges in the application of AlphaFold2 to virtual screening against novel targets merit consideration: 1) uncertainty in relative domain orientation prediction, 2) predictions limited to a single protein conformation for each input sequence, 3) binding site modifications by other entities such as homo- and hetero-multimers, cofactors, mutations, post-translational modifications, and additionally 4) unknown structure-function relationships for truly novel protein targets.

When coevolutionary information between two domains of a protein is low, the relative orientation of these domains in the AF2 predicted structures can be uncertain, as shown by the predicted aligned error (PAE) maps provided by AF2 predictions[42]. In our data set *AKT2, BRAF, EGFR*, and *VGFR2* were all multi-domain kinases, where the additional domains folded onto the surface of the kinase domain, however, judging from the predicted aligned error map these were uncertain predictions. For

BRAF, this completely abolished the virtual screening performance. Incorrect relative domain orientations appear to be a fairly common issue for AF2 structures spanning the cell membrane. To alleviate this issue, one might select only the domain of interest from the AF2 structures and perform truncation using both the per residue confidence values and the predicted aligned error maps. However, if the binding site is only transiently formed at flexible domain interfaces, the modeling of such an interface may still be a challenging task.

Besides flexible domain movements, proteins also often exhibit multiple conformational states. Sometimes this is inherently related to the function of the protein, such as the activation of GPCRs and the gating mechanism of ion channels, or may serve as a regulatory mechanism such as DFG-in and out conformations of kinases. In our data set the *MET* target was in the DFG-in conformation in the AF2 structure, while in the DFG-out conformation was present in the *holo* structure, resulting in markedly different virtual screening performance. IFD-MD was not designed to generate vastly different conformations of the protein backbone, but to refine the binding pockets and predict binding poses of non-native ligands. To model such larger rearrangements, different methods are needed. The open source AF2 implementation may be used to generate multiple model predictions. For example, it has been used to predict active and inactive state GPCR structures using annotated structure templates[43]. This, however, relies on ample characterization of conformational states of the protein family, which is still lacking for many drug target classes. *Ab initio* loop prediction and molecular dynamics simulations may also be used to re-predict or to generate multiple conformational states of proteins, however, depending on the time scale of these rearrangements, some states may be practically inaccessible in unbiased simulations. It is also expected that coupled prediction of folded protein and ligand structure in a complex by machine learning methods will be an intensive research area in the future. In this case ligand induced conformational changes may be inherently predicted by the machine learning algorithm.

The binding site of proteins may not only consist of standard amino acids of a single chain. They may be located on interfaces of homo- or hetero-multimers, interfaces of proteins with DNA or RNA, they may contain metals or organic cofactors, mutations, and post-translational modifications at PPIs and ligand binding sites, all of which may affect the properties of such pockets. Some advances have already been made to alleviate these shortcomings with the initially published AF2 structures, such as AlphaFold-Multimer for the prediction of multimeric structures[44]. As cofactor-containing targets were present in our data set, we were able to study the effect of these on hit identification. It is estimated that 30-40% of enzymes in the proteome require metal ions to perform their biological function in cells, and more than 10% of the human proteome requires organic cofactors for its function. Although cofactor binding sites are not automatically predicted in AF2 structures, the model is trained to predict the structures of proteins as they might appear in PDB, therefore backbone and side chain coordinates are

frequently consistent with the presence of cofactors[19]. Based on sequence and structural similarity the AlphaFill database provides modified AF2 structures with transposed cofactors needed for the structural integrity of proteins[45]. For completely novel protein classes, however, the identification of cofactor sites from sequence or predicted 3D structure and their placement is a challenging task, but several de novo algorithms and web servers do exist for this purpose[46,47]. A proof-of-concept study for grafting on post-translational modifications has also been published[47]. Effects of mutations on the structure seem not to be well predicted by AF2, although it is important to note that it was not trained to do so[48,49]. To uncover changes to binding site structures due to mutations, MD-based methods, such as protein FEP[50] and protein stability prediction methods may be used.

The current study focused on well-established targets with multiple crystal structures, both *apo* and *holo*, from well characterized target classes. However, the true transformative potential of AF2 lies in the prediction of protein structures that were previously understudied or elusive to structure determination methods. AF2 structures show great potential for enabling crystallography and cryo-EM structure determination, therefore the structural coverage of the human proteome is expected to increase. If no experimental structure of these novel targets exists, however, the previously discussed challenges are exacerbated. Domain organization, existence of multiple conformational states, complexing partners, the presence, stoichiometry, and location of cofactors and even ligand binding sites may be completely unknown. Here again we expect bioinformatics, molecular dynamics, and new machine learning-based methods to provide valuable insights for structure refinement, conformational sampling, binding site detection, placement of cofactors and ligands, and uncovering induced fit effects.

4. Data and Software Availability

The modeling in this work was carried out against two publicly available datasets: The DUD-E (http://dude.docking.org) and the AF2 database (https://alphafold.ebi.ac.uk). AF2 structures were modified from their raw state using the simple truncation protocol described in Section 2. Calculations were performed using the 2021-4 release of the Schrödinger Core Modeling Suite. Any non-default settings related to modeling calculations are described in Section 2. Supporting Information comprising (1) The prepared *apo*, *holo*, AF2 and truncated AF2 structures (2) set of IFD-MD refined AF2 models, and (3) complete virtual screening results is available, free of charge, here: https://schrodinger.com/other_downloads

5. CONCLUSION

The AlphaFold2 technology and publicly available database of predicted structures by many available metrics represent the current benchmark for protein structure prediction. However, it remains to be seen whether improvements in CASP-style metrics will translate to consistent success in drug discovery applications where experimentally derived structures are unavailable. As a step toward answering this critical question, we investigated virtual screening performance using AF2 structures in place of experimental structures. Out of the box, many AF2 structures produce low enrichment in our testing. They may contain low confidence loops occluding part of the binding site, missing co-factors, and uncertainty in relative domain orientation. We investigated some of these limitations here. Where a comparison could be made, we find that unrefined AF2 structures deliver similar enrichments to that of an *apo* experimentally derived structure, significantly below the enrichments using an experimentally derived *holo* structure. Meanwhile, the application of IFD-MD can induce a binding site conformation that delivers enrichments much closer to the *holo* structure. This is also supported by our finding that the average binding site volume of the IFD-MD refined AF2 structure is closer to a *holo* structure than the raw AF2 structure. This represents a repurposing of the IFD-MD protocol which was originally developed to model induced fit effects to improve cross docking accuracy. To achieve optimal enrichment, we use the top 5 models generated by IFD-MD, as opposed to considering the top 2 models as recommended in the usual application. Otherwise, no parameter adjustments are needed for IFD-MD. In prospective applications pilot screening with known actives could be used to identify the specific IFD-MD model to take forward into a production screen. To achieve optimal enrichment, all 5 IFD-MD models could be carried forward in an ensemble docking workflow.

Further work will be needed to determine whether the protocol we describe here could be used to coax additional accessible protein conformations given a single AF2 structure. It also remains to be seen whether this approach could enable successful hit discovery campaigns against a target without a representative in the AF2 training set. We plan for subsequent efforts to further explore the potentials and limitations of AF2 structures for hit identification, including the application for targets not represented in the PDB at the time of AF2 model training.

REFERENCES:

(1) Schuller, M.; Correy, G. J.; Gahbauer, S.; Fearon, D.; Wu, T.; Díaz, R. E.; Young, I. D.; Carvalho Martins, L.; Smith, D. H.; Schulze-Gahmen, U.; Owens, T. W.; Deshpande, I.; Merz, G. E.; Thwin, A. C.; Biel, J. T.; Peters, J. K.; Moritz, M.; Herrera, N.; Kratochvil, H. T.; QCRG Structural Biology Consortium; Aimon, A.; Bennett, J. M.; Brandao Neto, J.; Cohen, A. E.; Dias, A.; Douangamath, A.; Dunnett, L.; Fedorov, O.; Ferla, M. P.; Fuchs, M. R.; Gorrie-Stone, T. J.; Holton, J. M.; Johnson, M. G.; Krojer, T.; Meigs, G.; Powell, A. J.; Rack, J. G. M.; Rangel, V. L.; Russi, S.; Skyner, R. E.; Smith, C. A.; Soares, A. S.; Wierman, J. L.; Zhu, K.; O'Brien, P.; Jura, N.; Ashworth, A.; Irwin, J. J.; Thompson, M. C.; Gestwicki, J. E.; von Delft, F.; Shoichet, B. K.; Fraser, J. S.; Ahel, I. Fragment Binding to the Nsp3 Macrodomain of SARS-CoV-2 Identified through Crystallographic Screening and Computational Docking. *Sci Adv* **2021**, *7* (16). https://doi.org/10.1126/sciadv.abf8711.

(2) C, S.; S, D. K.; Ragunathan, V.; Tiwari, P.; A, S.; P, B. D. Molecular Docking, Validation, Dynamics Simulations, and Pharmacokinetic Prediction of Natural Compounds against the SARS-CoV-2 Main-Protease. *J. Biomol. Struct. Dyn.* **2022**, *40* (2), 585–611.

(3) Gupta, S.; Singh, A. K.; Kushwaha, P. P.; Prajapati, K. S.; Shuaib, M.; Senapati, S.; Kumar, S. Identification of Potential Natural Inhibitors of SARS-CoV2 Main Protease by Molecular Docking and Simulation Studies. *J. Biomol. Struct. Dyn.* **2021**, *39* (12), 4334–4345.

(4) Keretsu, S.; Bhujbal, S. P.; Cho, S. J. Rational Approach toward COVID-19 Main Protease Inhibitors via Molecular Docking, Molecular Dynamics Simulation and Free Energy Calculation. *Sci. Rep.* **2020**, *10* (1), 17716.

(5) Kumar, D.; Kumari, K.; Vishvakarma, V. K.; Jayaraj, A.; Kumar, D.; Ramappa, V. K.; Patel, R.; Kumar, V.; Dass, S. K.; Chandra, R.; Singh, P. Promising Inhibitors of Main Protease of Novel Corona Virus to Prevent the Spread of COVID-19 Using Docking and Molecular Dynamics Simulation. *J. Biomol. Struct. Dyn.* **2021**, *39* (13), 4671–4685.

(6) Lyu, J.; Wang, S.; Balius, T. E.; Singh, I.; Levit, A.; Moroz, Y. S.; O'Meara, M. J.; Che, T.; Algaa, E.; Tolmachova, K.; Tolmachev, A. A.; Shoichet, B. K.; Roth, B. L.; Irwin, J. J. Ultra-Large Library Docking for Discovering New Chemotypes. *Nature* **2019**, *566* (7743), 224–229.

(7) Stein, R. M.; Kang, H. J.; McCorvy, J. D.; Glatfelter, G. C.; Jones, A. J.; Che, T.; Slocum, S.; Huang, X.-P.; Savych, O.; Moroz, Y. S.; Stauch, B.; Johansson, L. C.; Cherezov, V.; Kenakin, T.; Irwin, J. J.; Shoichet, B. K.; Roth, B. L.; Dubocovich, M. L. Virtual Discovery of Melatonin Receptor Ligands to Modulate Circadian Rhythms. *Nature* **2020**, *579* (7800), 609–614.

(8) Sadybekov, A. A.; Sadybekov, A. V.; Liu, Y.; Iliopoulos-Tsoutsouvas, C.; Huang, X.-P.; Pickett, J.; Houser, B.; Patel, N.; Tran, N. K.; Tong, F.; Zvonok, N.; Jain, M. K.; Savych, O.; Radchenko, D. S.; Nikas, S. P.; Petasis, N. A.; Moroz, Y. S.; Roth, B. L.; Makriyannis, A.; Katritch, V. Synthon-Based Ligand Discovery in Virtual Libraries of over 11 Billion Compounds. *Nature* **2022**, *601* (7893), 452–459.

(9) Gentile, F.; Yaacoub, J. C.; Gleave, J.; Fernandez, M.; Ton, A.-T.; Ban, F.; Stern, A.; Cherkasov, A. Artificial Intelligence–enabled Virtual Screening of Ultra-Large Chemical Libraries with Deep Docking. *Nature Protocols*. 2022, pp 672–697. https://doi.org/10.1038/s41596-021-00659-2.

(10) Yang, Y.; Yao, K.; Repasky, M. P.; Leswing, K.; Abel, R.; Shoichet, B. K.; Jerome, S. V. Efficient Exploration of Chemical Space with Docking and Deep Learning. *J. Chem. Theory Comput.* **2021**, *17* (11), 7106–7119.

(11) Wang, R. Y.-R.; Song, Y.; Barad, B. A.; Cheng, Y.; Fraser, J. S.; DiMaio, F. Automated Structure Refinement of Macromolecular Assemblies from Cryo-EM Maps Using Rosetta. *eLife*. 2016. https://doi.org/10.7554/elife.17219.

(12) Callaway, E. Revolutionary Cryo-EM Is Taking over Structural Biology. *Nature* **2020**, *578* (7794), 201.

(13) Kim, H.-U.; Jung, H. S. Cryo-EM as a Powerful Tool for Drug Discovery: Recent Structural Based Studies of SARS-CoV-2. *Appl Microsc* **2021**, *51* (1), 13.

(14) Zhao, Z.; Tajkhorshid, E. GOLEM: Automated and Robust Cryo-EM-Guided Ligand Docking with Explicit Water Molecules. *Biophysical Journal*. 2021, p 290a. https://doi.org/10.1016/j.bpj.2020.11.1861.
(15) van Zundert, G. C. P.; Moriarty, N. W.; Sobolev, O. V.; Adams, P. D.; Borrelli, K. W. Macromolecular Refinement of X-Ray and Cryoelectron Microscopy Structures with Phenix/OPLS3e for Improved Structure and Ligand Quality. *Structure* **2021**, *29* (8), 913–921.e4.
(16) Porta-Pardo, E.; Ruiz-Serra, V.; Valentini, S.; Valencia, A. The Structural Coverage of the Human Proteome before and after AlphaFold. *PLoS Comput. Biol.* **2022**, *18* (1), e1009818.
(17) Muhammed, M. T.; Aki-Yalcin, E. Homology Modeling in Drug Discovery: Overview, Current Applications, and Future Perspectives. *Chem. Biol. Drug Des.* **2019**, *93* (1), 12–20.
(18) Zhang, Y.; Skolnick, J. The Protein Structure Prediction Problem Could Be Solved Using the Current PDB Library. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (4), 1029–1034.
(19) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583–589.
(20) Baek, M.; DiMaio, F.; Anishchenko, I.; Dauparas, J.; Ovchinnikov, S.; Lee, G. R.; Wang, J.; Cong, Q.; Kinch, L. N.; Schaeffer, R. D.; Millán, C.; Park, H.; Adams, C.; Glassman, C. R.; DeGiovanni, A.; Pereira, J. H.; Rodrigues, A. V.; van Dijk, A. A.; Ebrecht, A. C.; Opperman, D. J.; Sagmeister, T.; Buhlheller, C.; Pavkov-Keller, T.; Rathinaswamy, M. K.; Dalwadi, U.; Yip, C. K.; Burke, J. E.; Garcia, K. C.; Grishin, N. V.; Adams, P. D.; Read, R. J.; Baker, D. Accurate Prediction of Protein Structures and Interactions Using a Three-Track Neural Network. *Science* **2021**, *373* (6557), 871–876.
(21) Trott, O.; Olson, A. J. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J. Comput. Chem.* **2010**, *31* (2), 455–461.
(22) Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. Improved Protein-Ligand Docking Using GOLD. *Proteins: Structure, Function, and Bioinformatics*. 2003, pp 609–623. https://doi.org/10.1002/prot.10465.
(23) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J. Med. Chem.* **2004**, *47* (7), 1750–1759.
(24) Murphy, R. B.; Repasky, M. P.; Greenwood, J. R.; Tubert-Brohman, I.; Jerome, S.; Annabhimoju, R.; Boyles, N. A.; Schmitz, C. D.; Abel, R.; Farid, R.; Friesner, R. A. WScore: A Flexible and Accurate Treatment of Explicit Water Molecules in Ligand-Receptor Docking. *J. Med. Chem.* **2016**, *59* (9), 4364–4384.
(25) Guterres, H.; Park, S.-J.; Jiang, W.; Im, W. Ligand-Binding-Site Refinement to Generate Reliable Holo Protein Structure Conformations from Apo Structures. *J. Chem. Inf. Model.* **2021**, *61* (1), 535–546.
(26) Mysinger, M. M.; Carchia, M.; Irwin, J. J.; Shoichet, B. K. Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. *J. Med. Chem.* **2012**, *55* (14), 6582–6594.
(27) Miller, E. B.; Murphy, R. B.; Sindhikara, D.; Borrelli, K. W.; Grisewood, M. J.; Ranalli, F.; Dixon, S. L.; Jerome, S.; Boyles, N. A.; Day, T.; Ghanakota, P.; Mondal, S.; Rafi, S. B.; Troast, D. M.; Abel, R.; Friesner, R. A. Reliable and Accurate Solution to the Induced Fit Docking Problem for Protein–Ligand Binding. *Journal of Chemical Theory and Computation*. 2021, pp 2630–2639. https://doi.org/10.1021/acs.jctc.1c00136.
(28) Xu, T.; Zhu, K.; Beautrait, A.; Vendome, J.; Borrelli, K.; Abel, R.; Friesner, R.; Miller, E. Induced-Fit Docking Enables Accurate Free Energy Perturbation Calculations in Homology Models.

https://doi.org/10.26434/chemrxiv-2022-mq9n3.
(29) *New Features*. https://www.schrodinger.com/newfeatures (accessed 2022-05-05).
(30) Crystallography: Protein Data Bank. *Nature New Biology*. 1971, pp 223–223. https://doi.org/10.1038/newbio233223b0.
(31) Database, A. P. S. *AlphaFold Protein Structure Database*. https://alphafold.ebi.ac.uk/ (accessed 2022-05-04).
(32) Sastry, G. M.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. Protein and Ligand Preparation: Parameters, Protocols, and Influence on Virtual Screening Enrichments. *J. Comput. Aided Mol. Des.* **2013**, *27* (3), 221–234.
(33) Søndergaard, C. R.; Olsson, M. H. M.; Rostkowski, M.; Jensen, J. H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. *J. Chem. Theory Comput.* **2011**, *7* (7), 2284–2295.
(34) Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theory Comput.* **2011**, *7* (2), 525–537.
(35) Cappel, D.; Dixon, S. L.; Sherman, W.; Duan, J. Exploring Conformational Search Protocols for Ligand-Based Virtual Screening and 3-D QSAR Modeling. *J. Comput. Aided Mol. Des.* **2015**, *29* (2), 165–182.
(36) Jacobson, M. P.; Pincus, D. L.; Rapp, C. S.; Day, T. J. F.; Honig, B.; Shaw, D. E.; Friesner, R. A. A Hierarchical Approach to All-Atom Protein Loop Prediction. *Proteins* **2004**, *55* (2), 351–367.
(37) Halgren, T. New Method for Fast and Accurate Binding-Site Identification and Analysis. *Chem. Biol. Drug Des.* **2007**, *69* (2), 146–148.
(38) Halgren, T. A. Identifying and Characterizing Binding Sites and Assessing Druggability. *Journal of Chemical Information and Modeling*. 2009, pp 377–389. https://doi.org/10.1021/ci800324m.
(39) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004**, *47* (7), 1739–1749.
(40) Truchon, J.-F.; Bayly, C. I. Evaluating Virtual Screening Methods: Good and Bad Metrics for the "Early Recognition" Problem. *J. Chem. Inf. Model.* **2007**, *47* (2), 488–508.
(41) *The RDKit Book — The RDKit 2022.03.1 documentation*. https://www.rdkit.org/docs/RDKit_Book.html (accessed 2022-06-06).
(42) Tunyasuvunakool, K.; Adler, J.; Wu, Z.; Green, T.; Zielinski, M.; Žídek, A.; Bridgland, A.; Cowie, A.; Meyer, C.; Laydon, A.; Velankar, S.; Kleywegt, G. J.; Bateman, A.; Evans, R.; Pritzel, A.; Figurnov, M.; Ronneberger, O.; Bates, R.; Kohl, S. A. A.; Potapenko, A.; Ballard, A. J.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Clancy, E.; Reiman, D.; Petersen, S.; Senior, A. W.; Kavukcuoglu, K.; Birney, E.; Kohli, P.; Jumper, J.; Hassabis, D. Highly Accurate Protein Structure Prediction for the Human Proteome. *Nature* **2021**, *596* (7873), 590–596.
(43) Heo, L.; Feig, M. Multi-State Modeling of G-Protein Coupled Receptors at Experimental Accuracy. https://doi.org/10.1101/2021.11.26.470086.
(44) Evans, R.; O'Neill, M.; Pritzel, A.; Antropova, N.; Senior, A.; Green, T.; Žídek, A.; Bates, R.; Blackwell, S.; Yim, J.; Ronneberger, O.; Bodenstein, S.; Zielinski, M.; Bridgland, A.; Potapenko, A.; Cowie, A.; Tunyasuvunakool, K.; Jain, R.; Clancy, E.; Kohli, P.; Jumper, J.; Hassabis, D. Protein Complex Prediction with AlphaFold-Multimer. https://doi.org/10.1101/2021.10.04.463034.
(45) Hekkelman, M. L.; de Vries, I.; Joosten, R. P.; Perrakis, A. AlphaFill: Enriching the AlphaFold Models with Ligands and Co-Factors. https://doi.org/10.1101/2021.11.26.470110.
(46) Zhang, Y.; Zheng, J. Bioinformatics of Metalloproteins and Metalloproteomes. *Molecules* **2020**, *25* (15). https://doi.org/10.3390/molecules25153366.
(47) Rauer, C.; Sen, N.; Waman, V. P.; Abbasian, M.; Orengo, C. A. Computational Approaches to Predict Protein Functional Families and Functional Sites. *Curr. Opin. Struct. Biol.* **2021**, *70*, 108–122.
(48) Pak, M. A.; Markhieva, K. A.; Novikova, M. S.; Petrov, D. S.; Vorobyev, I. S.; Maksimova, E. S.;

Kondrashov, F. A.; Ivankov, D. N. Using AlphaFold to Predict the Impact of Single Mutations on Protein Stability and Function. https://doi.org/10.1101/2021.09.19.460937.
(49) Buel, G. R.; Walters, K. J. Can AlphaFold2 Predict the Impact of Missense Mutations on Structure? *Nat. Struct. Mol. Biol.* **2022**, *29* (1), 1–2.
(50) Scarabelli, G.; Oloo, E. O.; Maier, J. K. X.; Rodriguez-Granillo, A. Accurate Prediction of Protein Thermodynamic Stability Changes upon Residue Mutation Using Free Energy Perturbation. *J. Mol. Biol.* **2022**, *434* (2), 167375.