Induced-Fit Docking Enables Accurate Free Energy Perturbation Calculations in Homology Models

Tianchuan Xu[†], Kai Zhu[†,§], Alexandre Beautrait[†], Jeremie Vendome[†], Kenneth W. Borrelli[†,‡], Robert Abel[†], Richard A. Friesner [‖], Edward B Miller[†,*]

†Schrödinger, Inc., 1540 Broadway, New York, New York 10036, United States

‖Department of Chemistry, Columbia University, New York, 3000 Broadway, MC 3110, New York 10036, United States

ABSTRACT: Homology models have been used for virtual screening and to understand the binding mode of a known active, however rarely have the models been shown to be of sufficient accuracy, comparable to crystal structures, to support free-energy perturbation (FEP) calculations. We demonstrate here that the use of an advanced induced-fit docking methodology reliably enables predictive FEP calculations on congeneric series across homology models ≥ 30% sequence identity. Further, we show that retrospective FEP calculations on a congeneric series of drug-like ligands is sufficient to discriminate between predicted binding modes. Results are presented for a total of 29 homology models for 14 protein targets, showing FEP results comparable to those obtained using experimentally determined crystal structures for 86% of homology models with template structure sequence identities ranging from 30% to 50%. Implications for the use and validation of homology models in drug discovery projects are discussed, including the use of AlphaFold2 de novo structures.

I. Introduction

In a recent publication[1], we have described an improved approach to small molecule protein-ligand docking, IFD-MD, in which a combination of conformational search and molecular dynamics methods are used to incorporate induced fit effects in the receptor for prediction of the protein-ligand binding mode. For cases where the induced fit effects are dominated by side chain motion of the protein, a success rate (RMSD of one of the top two poses less than 2.5A) of greater than 90% was reported across a wide range of receptors for 415 examples, which constitutes a significant advance in robustness and breadth of applicability over prior induced fit docking methods. Furthermore, we demonstrated that poses generated by IFD-MD are of sufficient quality to enable useful free energy perturbation (FEP) calculations for congeneric series of drug-like ligands.

The data sets investigated in ref. 1 employed high resolution crystal structures as a starting point; the goal was to predict the change in receptor conformation for a new (active) ligand binding to the initial crystal structure. The success of the methodology of ref. 1 raises the question as to whether IFD-MD can be similarly effective when docking a known active ligand into a homology model, and if so, how the accuracy is impacted by the sequence identity between the target receptor and the template used to build the homology model. Docking into a homology model is often considerably more challenging than docking into a crystal structure, as the active site may need to be significantly reorganized to accommodate the ligand. However, as long as side chain motions dominate the reorganization, IFD-MD should in principle be capable of obtaining a reasonably good structure. Whether such structures will support FEP calculations of sufficient accuracy to enable a computationally driven structure-based drug design campaign is an empirical question that can only be answered by the appropriate computational experiments.

In our prior publication by Cappel et al., efforts to explore the impact of a homology model on FEP calculations deferred the problem of binding mode prediction[2]. Docked poses were selected on the basis of their ligand RMSD relative to the known experimental coordinates. In a subsequent publication by Moraca et al., induced-fit docking[3] was used for binding mode prediction with the IFD scoring function used to select the final predicted complex for use in FEP[4]. This paper successfully demonstrated that induced-fit docking enabled the use of FEP in homology models, however the work was limited to members of the phosphodiesterase (PDE) family.

Another pair of publications by Clark et al., demonstrated accurate use of relative binding FEP calculations in antibody-antigen complexes, specifically regarding the gp120 glycoprotein of HIV-1[5,6]. The binding experiments performed in that paper reference a gp120 variant, requiring construction of gp120 homology models. Wild-type gp120-antibody crystal structures had a sequence identity of 50% for the gp120 variant of interest and were used as a template. These papers extended the use of FEP to homology models of protein-protein complexes which avoids the need to cope with small-molecule binding pose determination, although there were other unique challenges like the placement of glycans and mutant side chains.

In the present paper, we investigate a total of 14 protein targets, each of which consists of a congeneric set of active ligands along with a co-crystallized structure for one of those ligands. Seven of the data sets are taken from our 2015 paper in which we introduced our FEP+ methodology[7] plus one homology model of PDE10A which was used as an isolated test case; the remaining six come from internal Schrodinger drug discovery projects. In each case, we evaluate the performance of IFD-MD for several different homology models, based on templates with differing sequence identities (roughly 30%, 40%, and 50%, although templates in all three of these categories are not available for every target), using the ligand for which a co-crystallized structure is available for the IFD-MD calculations (so as to be able to evaluate the RMSD from the experimentally determined structure). We then take the top 5 poses produced by IFD-MD and carry out FEP calculations for the entire congeneric series of ligands for each pose. The final pose is selected using a scoring function which combines several performance metrics from the FEP calculations (correlation coefficient, RMS error) as well as the absolute binding free energy, obtained from our absolute binding free energy perturbation (AB-FEP) module[8]. This protocol is a realistic one to perform for many potential structure-based drug discovery projects, requiring experimental binding affinity data for a congeneric series obtained either from the literature (publication or patent) or in-house experiments.

As is shown below, the interrogation of the top 5 poses, combined with the use of AB-FEP and FEP data from scoring the congeneric series, is sufficient to select a low RMSD pose in nearly all of the cases that we have investigated. Further refinement of the initial IFD-MD generated model is possible, for example by repredicting loop regions near the binding site, and/or running longer molecular dynamics simulations. Such refinement protocols will be the subject of a future publication.

The paper is organized as follows. In Section II, we describe the datasets used to evaluate IFD-MD homology modeling performance. In Section III, we outline the workflow used to carry out the calculations, which integrates the IFD-MD predictions with the relative and absolute FEP calculations that are utilized to rank order the top scoring poses. In Section IV, results and discussion are presented for the various test cases in the data set; the implications of our findings, particularly for structure-based drug discovery efforts, are considered in Section V, the Discussion. Finally, in Section VI, the Conclusion, we summarize our results and outline future directions.

II. Datasets

In the present paper, we focus on two different datasets. For the first dataset, we chose eight retrospective targets from our original FEP+ publication[7], referred to as the public dataset. In that paper, we considered an experimentally assessed congeneric series containing between 11 to 36 ligands for each of these targets, with the spread in binding affinities of the series ranging from 1.7 kcal/mol to 5.1 kcal/mol. As we previously reported, using free energy perturbation starting from crystal structures co-crystallized with at least one member of the congeneric series, we were able to achieve both predictive correlation ($R^2 > 0.35$) with experimental binding affinity and a low root-mean-square error (RMSE < 1.4kcal/mol) for each target. Among the eight targets in ref [7], MCL1 was a unique case. Compared with the natural peptide binder for this target, the small molecule ligands in the data series opened a new sub-pocket via a kink in a nearby helix. Low sequence identity (< 50%) homology modeling templates in the PDB database, either with a small molecule or peptide binder, seldom open this pocket fully. This type of substantial induced conformational change, akin in magnitude to the DFG-in and DFG-out transition in kinases, could not be handled with our existing IFD-MD technology. Consequently, MCL1 was left out of the initial dataset.

A second dataset was created using six targets and congeneric series derived from targets of interest to Schrodinger's internal and collaborative drug discovery programs. For each target, a publicly available structure and congeneric series was available. The term "proprietary dataset" refers to this collection of data. Target names and homology modelling templates associated with the proprietary data set are kept hidden to maintain confidentiality concerning targets of interest to Schrödinger and its collaborators.

Combining both datasets, there are in total 28 homology cases spanning 13 targets. Table 4 and Table 5 list the individual cases (each column represents a case).

For each target, we used BLAST[9] to scan the PDB database and chose a template with a sequence identity close to 50%, 40%, or 30%. With the following exclusions, the template with the highest sequence identity to the cutoff was chosen: (1) Templates with bound ligands (i.e. holo-proteins) were favored over templates without bound ligands (apo-proteins), unless all templates were holo- or apo-proteins; (2) For kinases, only DFG-in structures were used as templates, as all kinase targets in our datasets are DFG-in structures. Table 1 and Table 2 contained the targets and templates for both datasets, as well as information regarding the congeneric series and relative free-energy perturbation statistics using the target crystal structure. When an apo structure was chosen as the homology modeling template, the ligand from the nearest holo template was grafted in to identify the binding site for later work. For example, for the target ptp1b in the public data set, because the template with 30% sequence identity, 3M4U, is an apo structure, the binding site was defined using the ligand from the 40% sequence identity template, 2H02.

Table 1. Public Dataset Composition[a]

System	Native PDB	Native AB-FEP (kcal/mol)	No. of Ligands	No. of Edges	Binding Affinity Range	FEP on crystal structure RMSE[†]/ R^2	50% SeqId Template PDB	40% SeqId Template PDB	30% SeqId Template PDB
Thrombin	2ZFF	-7.88	11	16	1.7	0.90 / 0.50	N/A	3F6U	1FXY
Tyk2	4GIH	-17.58	16	24	4.3	0.63 / 0.89	3ZC6	4ID7	4OT5

Jnk1	2GMX	-15.13	21	31	3.4	0.98 / 0.77	3GC9	4B99	3MTL
CDK2	1H1Q	-10.85	16	25	4.2	1.36 / 0.38	1XO2	4AGU	3JUH
P38	3FLY	-16.61	34	56	3.8	1.02 / 0.61	1PME	4EOK	4BCF
Bace1	4DJW	-9.75	36	58	3.5	1.20 / 0.65	3ZLQ	N/A	1DP5
PTP1B	2QBS	-22.15	23	49	5.1	0.85 / 0.81	N/A	2H02	3M4U

[a] Units for binding affinity and RMSE are kcal/mol. An edge refers to a single FEP calculation between two ligands in the congeneric series.

Table 2. Proprietary Dataset Composition[a]

System	Native AB-FEP (kcal/mol)	No. of Ligands	No. of Edges	Binding Affinity Range	FEP on crystal structure RMSE /R^2
System 1	-15.36	10	26	4.5	0.91 / 0.91
System 2	-14.07	18	29	4.9	1.48 / 0.38
System 3	-9.55	11	16	5.9	1.01 / 0.92
System 4	-15.58	11	18	2.8	1.41 / 0.51
System 5	-15.56	12	20	3.5	2.11 / 0.39
System 6	-6.17	14	18	2.5	0.91 / 0.70

[a] Units for binding affinity and RMSE are kcal/mol. An edge refers to a single FEP calculation between two ligands in the congeneric series.

III. Computational Workflow

We expect that obtaining accurate pose prediction and subsequent FEP results for a congeneric series will, in general, be significantly more challenging for a homology model than in the case where one is starting with a high-resolution crystal structure. In a homology model, the environment of the active site may have backbone deviations (as compared to a high-resolution structure) which can increase the difficulty of predicting an accurate binding pose and can introduce noise into the calculation of both relative and absolute binding affinities. The objective of the workflow described below is to use a number of different metrics to evaluate the quality of the pose and thus to overcome noise. We optimize a scoring function which combines all of the information so as to succeed in the highest number of cases possible. Section IV.F reports initial true test set data for the method reported, however future testing of this scoring function on broader set of test cases will be the subject of future work and would be provide additional validation of the methodology.

The first step in the homology modeling workflow is construction of the homology model starting from the template receptor bound to a suitable template ligand. If the template structure is a holo structure, then the ligand that is a component of the high-resolution complex serves as the template ligand. If the template structure is an apo structure (as might be the case for a project on a recently identified drug discovery target), we place a ligand from an alternative source into the template (for the present paper, we do this using an alternative holo template). The homology model itself is then generated from the template via standard alignment and modeling building protocols, as discussed below.

Once the homology model is created, the next step is the use of IFD-MD docking to generate candidate poses for one or more compounds that are known to be active against the target receptor. The ligand in the template structure is removed, and the top 5 poses from IFD-MD docking are selected for further processing. As described in ref. 1, the IFD-MD scoring function itself uses multiple components (most importantly the Prime continuum solvent-based energy[10,11], WScore docking score[12], and stability in metadynamics simulations[13]) to rank order hundreds of candidate poses. In ref. 1, where induced fit docking into a high-resolution crystal structure was the objective, our goal was to reliably place a low RMSD structure (less than 2.5A) among the two top ranked poses. Because of the greater noise expected in docking into a homology model, we revise our goal to be placement of a low RMSD pose among the top 5 ranked structures. Subsequent evaluations of pose performance using both relative FEP (for a congeneric series) and absolute FEP (AB-FEP) enables selection of an optimal low RMSD final pose among the five candidates emerging from IFD-MD.

In Figure 1 is graphical depiction of the workflow outlined above. After all of the descriptors have been produced for each of the five candidate IFD-MD poses, a scoring function is used to combine the information and select the final pose. The formulation of the scoring function is given below in section III.D.

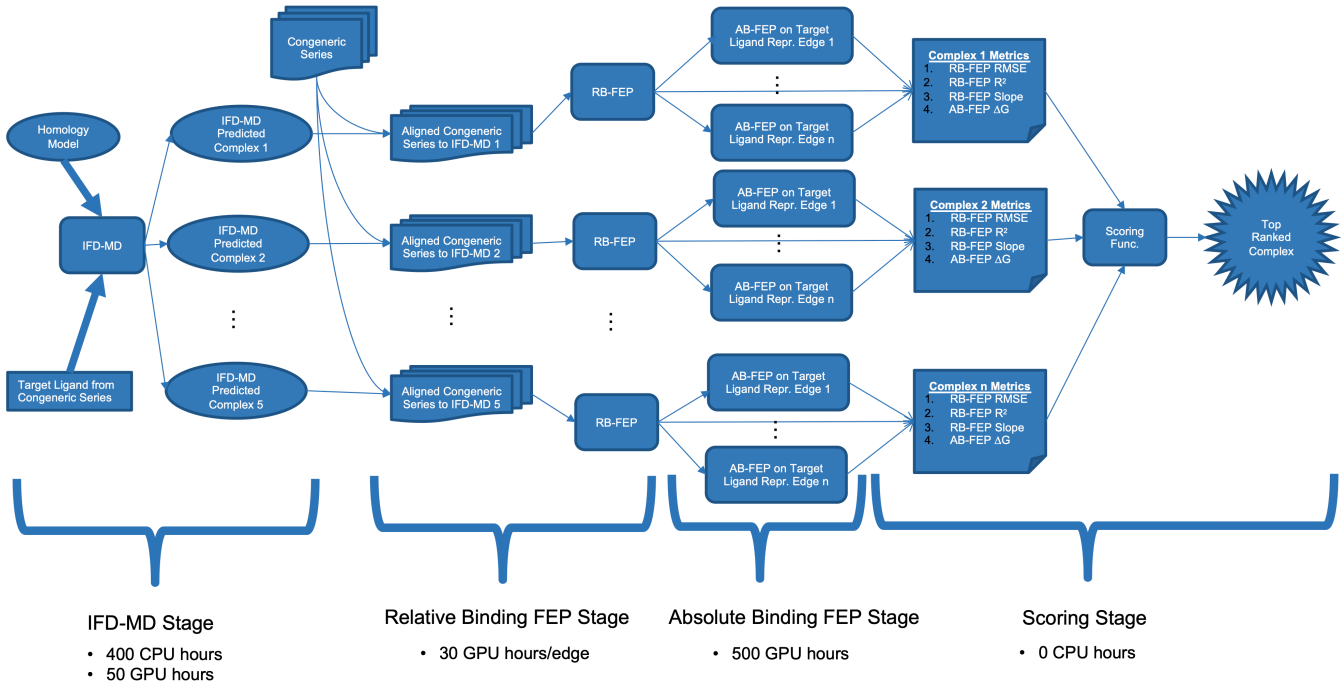

Figure 1. Flow chart of the homology modelling workflow. The workflow is initiated with an existing homology model and a congeneric series. The output is a list of ranked ligand-receptor complexes that are scored using a combination of relative FEP metrics and absolute binding FEP predicted free energy. The homology model can be generated via classical sequence alignment methods or alternatively using template-free methods such as AlphaFold2 predictions. The congeneric series is a list of similar ligands with associated binding affinity or functional activity data, e.g. K_i or EC50. Timings report here are approximate and depend on a number of factors such as the size of the receptor and the number of ligands in the congeneric series. Relative Binding FEP timings are a function of the number of edges, a single FEP calculation performed between two ligands. For each target, the number of edges is listed in Table 1 and Table 2.

III.A. Homology Model Generation

The target and template sequences were aligned using the MUSCLE methodology[14,15]. The MUSCLE alignment was modified to shift gaps found within the secondary structure to instead be adjacent to loops. This correction was applied only when insertion/deletions would lie within 10 Å of the ligand in the homology modeling template. The finalized alignment was supplied to the Prime Homology Modeling module[10] to build a complete atomic model of the receptor. We were interested in performance across a range of homology models and sequence identities for the purposes of this work, therefore we remain agnostic regarding questions about competing sequence alignment methodologies. Other alignment algorithms, such as HHBlits[16], or alternative homology models, such as AlphaFold2[17] structures, could be used without altering the protocol developed in this work. The use of AlphaFold2 structures may increase the success probability of our workflow when the sequence identity of the best template is in the 20-40% range.

The initial homology model was built using Prime[10,18,19]. Residues that were conserved between the target and the template were assigned the template's coordinates, including side chain atoms. The side chain atoms were optimized for residues which were mutated between the template and target. De-novo loop prediction was performed wherever there were insertions or deletions. Throughout model building minimization used the VSGB2 implicit solvent model[11] and OPLS3e[20] force field .All un-templated C or N termini longer than one residue were removed from the homology model.

III.B. Ligand Docking and Binding Site Optimization

Following the construction of the homology model, IFD-MD[1] was used for ligand docking and binding site optimization. IFD-MD requires the presence of some chemical matter in the receptor, referred to as the template ligand, in order to define the binding site and facilitate induced-fit docking. When available, the ligand from the homology modeling template served as the template ligand for IFD-MD docking homology models. We selected for docking with IFD-MD the congeneric series member that was present in a previously published crystal structure, referring to the reference ligand. This enables us to track the ligand's RMSD for retrospective evaluation purposes. In a prospective project, one might dock the series' most potent ligand. An alternative strategy would be to dock multiple ligands from the congeneric series into the template and analyze the results of multiple IFD-MD runs. This could lead to building a superior model in some situations, e.g., when the series has more than one important binding mode. In the present work, docking a single ligand, selected from the congeneric series using the criterion described above, produced satisfactory results in nearly all cases studied.

III.C. Model Validation with Free Energy Perturbation Calculations

In many if not most drug discovery projects, some initial data regarding binding of active ligands to the target receptor is available, either from the literature (publications or patents) or

via an in-house experimental screen. A central finding of the present paper is that the identification of even a relatively small number of active ligands (although more data is always better) can be effectively used to develop and validate a superior homology model - one that is suitable for enabling a structure-based project going forward. In effect, the ligands act as a probe of the receptor active site which can be used to discriminate among various plausible alternatives.

We utilize two different types of free energy calculations to evaluate the performance of a specified IFD-MD pose for a particular active ligand and its associated congeneric series. Absolute binding free energy perturbation (AB-FEP) calculations are aimed at determining the free energy of binding of the ligand to a fixed receptor conformation (defined by the protein-ligand complex structure that is input into the calculations)[8]. These calculations are more difficult from a technical point of view than the standard relative FEP calculations among multiple ligands, and they are subject to noise arising from noise in the input structure (i.e. deviations from a high resolution experimental crystal structure), a significant concern when using a homology model. Furthermore, AB-FEP calculations do not include the reorganization free energy of the receptor (free energy to transition from the apo structure of the receptor to the holo structure induced by a specific ligand). Despite these issues, the robustness of AB-FEP calculations has reached the point where it is possible to reject most IFD-MD poses which are seriously flawed. In some cases, this capability is extremely helpful because (as we will see below) false positives in relative FEP assessment can arise, particularly when the initial data set of ligand binding affinities is sparse and the activity range of the data set is limited.

Relative binding FEP tests whether the homology model can properly describe changes in binding affinity as various regions of the ligand are modified. With sufficient data, relative FEP provides a sensitive calibration of the quality of the binding pose, and a useful assessment of suitability for prospective FEP calculations. As noted above, AB-FEP calculations are a helpful augmentation to relative FEP when the data set is potentially deficient in some respects.

Schrödinger's free-energy perturbation package, FEP+[7,21] was used to predict the protein-ligand binding affinities across a ligand congeneric series particular to a given target. We preserved an identical FEP map topology across all IFD-MD predicted poses. This topology is identical to what was used for the crystal structure. This removes any bias caused by the topology of the map and ensures comparisons between the crystal structure and alternative predicted binding modes are done fairly. Default FEP settings were used with the exception that 10ns of simulation time were performed instead of 5ns.

The reference ligand, the ligand which was present in the target crystal structure and which was used for reporting the RMSD, is connected to multiple other ligands via perturbation edges in the relative free energy perturbation map. All ligands in all FEP maps contained closed cycles to allow for cycle closure correction[7]. This means each ligand, including the reference ligand, has at least two edges connected to it. For each edge, we extracted a representative structure. The representative structure was extracted for the physical replicas (i.e. lambda=0). This resulted in at least two representative structures for the two or more edges connected to the representative ligand. The enhanced sampling performed in relative FEP in many cases provided a salutary effect, improving the quality of the IFD-MD output structure. Table 4 and Table 5 show the improvement in RMSD from the molecular dynamics run during RB-FEP. Absolute binding free energy was calculated using these representative structures as input resulted in at least two predicted ΔGs for a single ligand. The lowest free energy predicted was used as the absolute binding free energy for that predicted ligand-receptor complex. Figure 1 illustrates this process.

III.D. Determination of the composite scoring function

The composite scoring function integrates the data from relative FEP and AB-FEP calculations in order to select a final pose. The specific descriptors used comprise the AB-FEP free energy and the correlation coefficient (R^2), RMS error (RMSE), and slope for the relative FEP calculations. A linear functional form proved sufficient to select the best pose across the entire data set:

$$score = w_{\mathrm{RMSE}} * RMSE_{\mathrm{RB-FEP}} + w_{\mathrm{R^2}} * \mathrm{R}^2_{\mathrm{RB-FEP}} + w_{\mathrm{slope}} * \mathrm{abs}(1 - \mathrm{slope}_{\mathrm{RB-FEP}}) + w_{\mathrm{AB-FEP}} * \Delta\mathrm{G}_{\mathrm{AB-FEP}}$$

Equation 1. Composite Scoring Function

Where $RMSE_{\mathrm{RB-FEP}}$, $\mathrm{R}^2_{\mathrm{RB-FEP}}$, $\mathrm{slope}_{\mathrm{RB-FEP}}$ are the pairwise RMSE, R^2 and slope of the relative FEP calculations of the representative structure, across the ligand congeneric series. $\Delta\mathrm{G}_{\mathrm{AB-FEP}}$ denotes the representative structure's predicted AB-FEP binding affinity.

We chose the convention that a more negative score was more favorable. Of the descriptors used for the scoring function, the $RMSE_{\mathrm{RB-FEP}}$, and $\Delta\mathrm{G}_{\mathrm{AB-FEP}}$ already follow this convention. For the slope, we used the functional form $\mathrm{abs}(1 - \mathrm{slope}_{\mathrm{RB-FEP}})$ to capture the distance the observed slope was from unity, which would be an ideal slope. The remaining term $\mathrm{R}^2_{\mathrm{RB-FEP}}$, therefore required its weight to be a negative value.

All of the public and proprietary data sets were used to train the model with the exception of the PDE10A homology modeling performed in section IV.F which serves as true test set data for the developed protocol.

The parameters for the scoring function were optimized using a grid search. All predicted models for all grid points in this four-dimensional parameter space were scored within a coarse 20-point per dimension resolution and fixed range of values. The points in this 4D parameter space were then ranked by the number of successful cases and by the separation between satisfactory and unsatisfactory poses. A pose was considered satisfactory if its ligand-heavy-atom RMSD was ≤ 2.6 Å. The separation between satisfactory and unsatisfactory poses is defined as follows:

$$S = \sum_{i=1}^{N} ssr_i$$

Equation 2. Optimization function

Where the summation is over all 28 cases. The score-span ratio (ssr_i) of a single case is defined as:

$$ssr_i = \frac{s_{best} - s_{marginal}}{s_{best} - s_{worst}} I_{gp}$$

Equation 3. Score-span ratio definition

Where s_{best} and s_{worst} are the scores of the pose with the highest rank and the pose with the lowest rank, respectively for a single case. $s_{marginal}$ is defined as the worst score among satisfactory poses with no interspersed bad poses, or as the best score among poses with an RMSD more than 2.6 Å if no pose other than the top-ranked pose has an RMSD of 2.6 Å or less. I_{gp} is an indicator function that returns 1 only when the rank-one pose is a satisfactory pose. The larger the score span ratio, the more the satisfactory poses occupy the span of scores without interspersed unsatisfactory poses and the better job the scoring function performs at discrimination.

Once the points in this 4D parameter space were ranked, a finer search was performed with a resolution of 50 points per parameter. The range of values was set to be the range of observed values for each dimension for the top 20 points in the coarse search of parameter space. Supplemental Table S5 lists the range of parameters and the final values for the coarse and fine resolution grid searches.

Table 3. Results of scoring function optimization excluding certain parameters[a]

Excluded Parameter				$\sum_{i=1}^{N} ssr_i$	Number Successful Cases
w_{RMSE}	w_{R^2}	w_{slope}	$w_{\mathrm{AB-FEP}}$		
				11.73	27/28
X				11.50	27/28
	X			10.66	27/28
		X		11.73	27/28
			X	10.08	25/28
X	X			10.56	27/28
X		X		10.86	27/28
X			X	10.15	25/28
	X	X		9.96	27/28
	X		X	8.60	24/28
		X	X	10.58	23/28
X	X	X		10.72	25/28
X	X		X	8.52	24/28
X		X	X	10.29	24/28
	X	X	X	8.07	21/28

[a] Out of the 28 homology modeling cases, one case failed to produce any poses under 2.6 Å ligand RMSD. This forces the maximum number of successful cases to be 27 out of 28 cases.

All combinations of terms were excluded to explore the contribution of that term to the overall scoring function. Table 3 lists the results of all 15 combinations of excluding each of the four parameters. One case, the 30% sequence identity case for System 6 in the proprietary data set failed to produce any poses under 2.6 Å (see Table 5 and Table S4). This set the maximum number of cases to optimize to be 27 out of 28 total cases. Of the four terms in the scoring function, the $w_{\mathrm{AB-FEP}}$ parameter appeared to be the most significant term. When it is the only excluded parameter, the scoring function is unable to rescue two cases, regardless of the weights of the other three parameters. For the other terms, their exclusion resulted in a poorer score-span ratio, which implies a reduced discrimination between satisfactory and unsatisfactory poses. The slope term was the least consequential of the four terms, showing only a significant impact when excluded along with another term, for example the AB-FEP ΔG weight.

Leave-one-out (LOO) optimization was performed to investigate the stability of results when individual cases were excluded from parameter optimization. Leave-one-out optimization excludes a single case from parameter optimization. Once a new set of a parameters is found, the excluded case is then scored and classified as a success or failure. Out of the 27 cases that have at least one 2.6 Å RMSD pose or better, two cases failed during leave one out optimization. Those two cases are the BACE 26% sequence identity case and System 2 32% sequence identity.

For the BACE 26% sequence identity case, LOO optimization led to the selection of the 4.98 Å pose with superior RB-FEP RMSE and R^2 compared to the preferred 1.96 Å pose (Table S3). The 1.96 Å pose has a highly discriminating AB-FEP ΔG that is nearly 12 kcal/mol superior to this competing 4.98 Å pose. However, with LOO, the weight of the AB-FEP ΔG term was insufficient to overcome the inferior RB-FEP statistics of the 1.96 Å pose.

System 2, 32% sequence failed LOO optimization with the 3.54 Å pose selected rather than the 2.35 Å pose (Table S4). Here, proper weighting of the RB-FEP slope with the AB-FEP ΔG was not achieved when this case was excluded from optimization.

Both cases which failed LOO optimization are built from low-sequence identity templates and show the limitations of this small data set.

IV. Results and Discussion

IV.A. Overview.

In Table 4 and Table 5, we present results for the IFD-MD homology modeling workflow for all of the cases included in the public and proprietary data sets. As a comparison point, relative and absolute FEP results for the native structures of the targets are given in Table 1 and Table 2. We also present the results of rigid receptor Glide docking (in Table 4 and Table 5) of the target ligand for the same set of test cases. The Glide results display a high fraction of large errors in pose prediction (RMSD > 3Å), indicating that the homology models neglecting induced-fit effects are unsuitable for subsequent investigation of ligand binding without refinement of the binding site, along the lines that we carry out here. These Glide poses resulted almost entirely in poor FEP performance (Supplemental Table S1 and Table S2)

From the standpoint of structure-based drug discovery, a minimal basis for proceeding to use the homology model in a

project would be reasonable results (with regard to RMSE, R^2, and slope) for relative FEP calculations. We defined reasonable relative FEP performance to be an RMSE ≤ 1.5 kcal/mol and an $R^2 \geq 0.3$.

Table 4. Public Data Set FEP Results and Ligand RMSDs[a]

		Systems										
		BACE		TYK2			P38			JNK1		
Template		3ZLQ	1DP5	3ZC6	4ID7	4OT5	1PME	4EOK	4BCF	3GC9	4B99	3MTL
Sequence Identity		54%	26%	51%	41%	33%	49%	38%	31%	47%	41%	29%
$RMSE_{RB\text{-}FEP}$ (kcal/mol)		1.01	1.48	0.64	1.36	0.87	1.36	1.52	1.13	1.08	1.42	0.95
R^2_{RB-FEP}		0.48	0.02	0.89	0.67	0.78	0.31	0.30	0.43	0.63	0.59	0.60
$slope_{RB-FEP}$		0.88	0.14	0.97	0.29	0.72	0.56	0.66	0.58	1.14	1.34	0.94
$\Delta G_{AB\text{-}FEP}$ (kcal/mol)		-16.97	-18.75	-12.62	-12.18	-10.55	-13.08	-15.34	-10.81	-14.43	-16.2	-16.41
RMSD	GlideSP (Å)	3.16	2.37	3.82	4.81	6.34	2.70	6.76	8.74	5.54	5.84	6.16
	IFD-MD (Å)	2.17	2.60	1.78	3.41	2.18	1.39	2.42	2.01	3.12	1.29	1.44
	Final (Å)	1.40	1.96	0.74	0.77	0.84	1.06	1.49	1.31	1.22	0.93	0.84
		Systems										
		CDK2			Thrombin		PTP1B					
Template		1XO2	4AGU	3JUH	3F6U	1FXY	2H02	3M4U				
Sequence Identity		50%	40%	34%	42%	39%	38%	27%				
$RMSE_{RB\text{-}FEP}$ (kcal/mol)		1.30	1.17	1.27	0.79	0.76	0.67	1.05				
R^2_{RB-FEP}		0.45	0.54	0.45	0.35	0.48	0.89	0.71				
$slope_{RB-FEP}$		0.35	0.49	0.46	0.73	0.95	1.04	0.84				
$\Delta G_{AB\text{-}FEP}$ (kcal/mol)		-13.56	-11.19	-13.41	-8.17	-7.36	-18.28	-14.02				
RMSD	GlideSP (Å)	5.42	1.50	5.58	7.11	2.90	7.43	6.36				
	IFD-MD (Å)	2.14	1.86	1.45	1.53	2.43	1.09	2.18				
	Final (Å)	1.56	1.87	1.61	1.16	2.08	1.41	1.24				

[a] Free energy perturbation and ligand RMSDs for the cases in the public data set. The relative-binding FEP statistics (RB-FEP) R^2 and slope refer to the Pearson correlation coefficient of the best fit linear regression that maps experimental ΔG to relative binding FEP predicted ΔG. The slope corresponds to the slope of this best fit line. RMSDs are reported as ligand heavy atom RMSD compared to the crystal structure for the reference ligand in the congeneric series. The GlideSP RMSD is the rigid-receptor docking RMSD. Rigid-receptor docking is not part of the workflow but is shown here for comparison to illustrate the need for induced-fit effects to be considered. The final RMSD is the ligand heavy-atom RMSD of the lowest free energy MD frame scored using absolute-binding FEP (AB-FEP).

Table 5. Proprietary Data Set FEP Results and Ligand RMSDs[a]

		System 1		System 2			System 3		System 4	System 5	System 6
Sequence Identity		54%	33%	52%	41%	32%	52%	40%	27%	28%	30%
Ligand RMSD (Å)		2.45	1.53	1.43	2.17	2.35	1.30	1.50	1.15	2.04	3.72
$RMSE_{RB\text{-}FEP}$ (kcal/mol)		0.80	1.05	1.44	1.53	1.35	1.49	1.23	0.50	3.08	0.79
R^2_{RB-FEP}		0.88	0.85	0.46	0.57	0.48	0.96	0.88	0.88	0.35	0.54
$slope_{RB-FEP}$		0.95	0.64	0.70	1.01	0.66	1.48	0.59	0.68	1.27	0.47
$\Delta G_{AB\text{-}FEP}$ (kcal/mol)		-13.81	-10.25	-11.65	-12.39	-8.54	-9.53	-8.22	-11.85	-13.91	-7.6
RMSD	GlideSP (Å)	No pose	No pose	1.37	9.53	1.74	0.54	5.28	6.08	7.93	5.37

	IFD-MD (Å)	1.23	2.72	2.36	2.30	2.40	3.08	1.01	1.55	2.52	2.88
	Final (Å)	2.45	1.53	1.43	2.17	2.35	1.30	1.50	1.15	2.04	3.72

[a] Free energy perturbation and ligand RMSDs for the cases in the public data set. The relative-binding FEP statistics (RB-FEP) R^2 and slope refer to the Pearson correlation coefficient of the best fit linear regression that maps experimental ΔG to relative binding FEP predicted ΔG. The slope corresponds to the slope of this best fit line. RMSDs are reported as ligand heavy atom RMSD compared to the crystal structure for the reference ligand in the congeneric series. The GlideSP RMSD is the rigid-receptor docking RMSD. Rigid-receptor docking is not part of the workflow but is shown here for comparison to illustrate the need for induced-fit effects to be considered. The final RMSD is the ligand heavy-atom RMSD of the lowest free energy MD frame scored using absolute-binding FEP (AB-FEP).

A significant fraction (86%) of the test cases below satisfy the criteria that RMSE ≤ 1.5 kcal/mol and $R^2 \geq 0.3$ (Table 6 and Table 7); two of the cases that do not (bace1, 26% sequence identity and System 5, 28% sequence identity of the proprietary data set) involve building homology models when the sequence identity of the template and target is below 30%. It is not surprising to see a degradation in model quality in this regime. Nevertheless, the results for all four cases with unsatisfactory FEP performance have some favorable features – all four cases have a good ligand RMSD under 2.5 Å. These partially validated results suggest that further refinement of the structure (e.g. prediction of surrounding loops, more extensive MD simulation, etc.) could improve the structures sufficiently to support acceptable predictive relative FEP performance. Alternatively, a better starting point for the homology model (e.g. an AlphaFold2[17] structure) might similarly yield superior results to those shown here. We intend to explore both of these directions in future work. For the specific protocol described in this paper, a sequence identity of 30% or higher appears to be sufficient to yield a useful model over 80% of the time. Statistically, one would not necessarily expect every 30% sequence identity model to work as well as those shown here but given the availability of experimental binding data for the congeneric series (necessary to execute the workflow in any case), suitable performance on the relative FEP metrics can readily be identified. Further tests of the model in prospective use (involving modifications of the ligand at different locations than in the congeneric series) should then enable a more rigorous probe of the overall quality of the ligand pose.

Table 6. Homology modeling success rate by FEP Performance[a]

	Total Cases	50% Seq. ID	40% Seq. ID	30% Seq. ID
Public Dataset	18	5/5	5/6	6/7
Proprietary Dataset	10	3/3	1/2	4/5

[a] Each homology case is rounded to the nearest sequence identity bin listed here. Success is defined to have a RMSE ≤ 1.5 kcal/mol and an $R^2 \geq 0.3$.

Table 7. Homology modeling success rate by ligand RMSD[a]

	Total Cases	50% Seq. ID	40% Seq. ID	30% Seq. ID
Public Dataset	18	5/5	6/6	7/7
Proprietary Dataset	10	3/3	2/2	4/5

[a] Each homology case is rounded to the nearest sequence identity bin listed here. Success is defined to have a ligand heavy-atom RMSD ≤ 2.5 Å.

While some degradation of the relative FEP results as compared to the crystal structure data can be observed, it is remarkable how frequently results of comparable quality are obtained, in some cases even for lower sequence identities. This result is not achieved via IFD-MD alone; the molecular dynamics simulation of the FEP step is important in a number of cases to generate an accurate ligand pose. However, results of this type require at least a plausible initial guess, unless one is willing to expend extraordinarily amounts of computational resources; there is no possibility that a relatively short MD trajectory (nanoseconds as opposed to many microseconds) could refine a 5-7A pose of the type that is generated by rigid receptor docking in more than half of the test cases.

The AB-FEP results are quite variable, with some of the low RMSD poses showing a better signal than others (see for example, the P38 results in Table 4**)**. The AB-FEP results for crystal structures are obtained in a simulation with all of the key interactions in place, as well as an optimally relaxed environment. Because AB-FEP is strongly restrained around the initial pose (necessary in order to obtain converged results in reasonable computation times), sampling away problems with the initial structure is difficult, and the quality of the AB-FEP result is going to be strongly dependent upon the quality of the initial pose. The initial results shown below are encouraging in that a strong signal is frequently obtained, and in a few cases can profitably be used to choose between poses with noisy relative FEP results (only one of which has a reasonable RMSD). It is also quite possible that further refinement of the structure, around a localized region, could be used to generate improved AB-FEP results; indeed, AB-FEP can be used as a signal to select the best alternative among such refined poses. In fact, success along these lines is already apparent from the current protocol in which AB-FEP is used to select among a number of representative MD frames from a given IFD-MD pose. Experiments along these lines will be carried out in subsequent work.

Most of the remaining subsections below in this section address the various points discussed above in more detail, including illustrative examples. The final subsection, IV.F, considers a potential failure mode; a specific case, a congeneric series of ligands binding to the PDE10A receptor. Here, use of the data as provided leads to ambiguity in selecting the final pose; in particular, a pose with a very poor RMSD is

competitive in both absolute and relative FEP performance with a quite different pose with a sub 1 Å RMSD. We analyze this case in detail and show that the problem lies in the fact that the ligand is very hydrophobic (leading to plausible AB-FEP binding free energies in multiple poses) and that the binding affinity is highly correlated with molecular weight (in essence, more binding affinity can be obtained by adding hydrophobic groups to the core, even in very different orientations of the core). This problem can be overcome by selecting a subset of ligands with similar molecular weights but differentiated binding affinities, for which the low RMSD pose qualitatively outperforms in the incorrect pose in relative FEP correlation with experiment.

IV.B. IFD-MD Allows for Accurate Binding Mode Prediction in Imperfect Homology Models

Below we report the RMSD of the ligand for each stage of the homology modeling workflow. Among all the stages, the induced-fit docking of a ligand from the congeneric series was critical. The induced-fit docking produced pose becomes the scaffold upon which all members of the congeneric series are aligned. While the additional dynamics performed during relative binding FEP and absolute binding FEP can correct some flaws in the structure, in any reasonable simulation time, significant kinetic barriers are likely present preventing a total reorganization of the binding mode. In 24 of 28 homology modeling cases from both datasets, IFD-MD generated poses with ligand heavy-atom RMSD less than 2.5Å as one of the top five poses (Table 4 and Table 5). For comparison, we report the RMSD of rigid-receptor docking, performed using GlideSP. Out of the 28 homology modeling cases, only 5 cases can produce a sub-2.5 Å ligand RMSD. This illustrates the need for an induced fit technology to produce an accurate initial pose (Table S1 and Table S2) .

The example of cyclin-dependent kinase 2 (CDK2) modeling using the 34% sequence identity 3JUH template demonstrated how IFD-MD poses preserved critical binding site interactions during FEP simulations. Figure 2A shows the crystal structure of CDK2 relative to the initial homology model, before any induced-fit docking has occurred. The linkage/hinge area was of poor quality in the original homology model and can be described as having two flaws:

1. The backbone amine and carbonyl groups were pointed upward, away from the binding site

2. The auxiliary hydrogen bond (topmost in the figure) was impossible to form because the second carbonyl group retreated from the binding site.

As a result, rigid receptor docking with a flexible ligand and a rigid protein receptor failed to predict any plausible pose in even the top 20 predictions. Figure 2B shows the top rigid receptor docked pose. This pose, unable to form hydrogen bonds with the hinge, rotates the polar groups on the ligand to instead face solvent. By comparison, the top two poses produced by IFD-MD both had reasonable hinge hydrogen bond interactions. The top IFD-MD pose has a ligand RMSD of 5.21 Å and is shown in Figure 2C while the second-ranked IFD-MD pose is a 1.45 Å RMSD pose and is shown in Figure 2D. Both poses formed similar hydrogen bonds with the hinge. Notable here was that the ligand perturbations occurred on the benzene ring which are in a distinctly different environment. As expected, these two poses were markedly different in terms of their performance in FEP. The 1.45 Å pose demonstrated strong recapitulation of experimental data with R^2 of 0.45 and RMSE of 1.27 kcal/mol, whereas the negative control pose performed poorly with an R^2 of 0.16 and an of RMSE of 4.09 kcal/mol. The FEP correlation plots for these two poses are shown in Figure 2E and F.

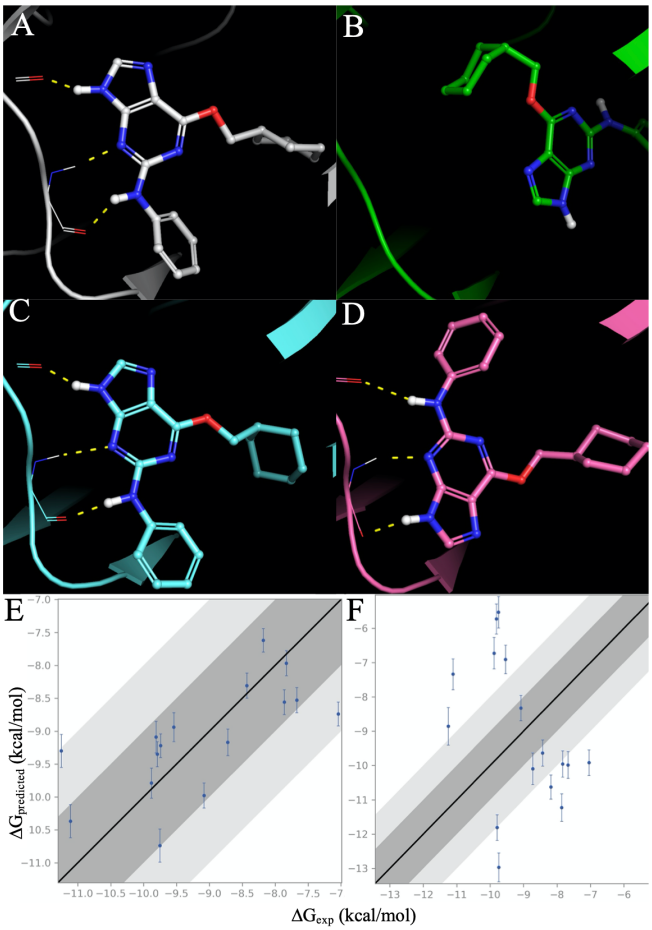

Figure 2. Docking and FEP performance for the CDK2 case using PDB 3JUH, 34% sequence identity. (A) Interactions between the native CDK2 crystal structure (PDB:1H1Q) and its native ligand (both in white), as well as the initial homology model (green). (B): The top-ranked pose for rigid receptor docking (GlideSP) into the initial homology model. Unable to form a hinge interaction, the pose has retreated from the hinge and moved the polar atoms to be solvent exposed. (C) Hinge interactions of the 1.45 Å RMSD IFD-MD pose. (D) Hinge interactions of the 5.21 Å RMSD IFD-MD pose. (C) The FEP results for the 1.45 Å IFD-MD pose. The RMSE is 1.27 kcal/mol and the R^2 is 0.45. (F) The FEP results for the 5.21 Å IFD-MD pose. The RMSE is 4.09 kcal/mol and the R^2 is 0.16.

IV.C. Ligands Requiring Simple Motions Could be Fixed by the Sampling Performed in FEP

When IFD-MD predicts no pose with an RMSD of less than 2.5 Å among the top 5, it is possible that a relatively short MD simulation can transform an apparently poor IFD-MD pose into a much-improved pose with a low RMSD, for example by

a predominantly translational motion of the ligand relative to the receptor, typically coupled to some side chain motions and ligand peripheral group rotation. For example, we observed this starting from a 3.12 Å RMSD IFD-MD ligand pose docked into a homology model of c-Jun N-terminal kinase-1 (JNK1) modeled using PDB ID 3GC9 template (47% sequence identity). While the IFD-MD pose has an RMSD of 3.12 Å, its variation from the crystal structure was due to a simple folding of the chloro-dimethyl benzene ring, which was addressed later in the initial FEP equilibrium step. As a result, the FEP predictions following the equilibrium stage were comparable to the crystal structure FEP results (R^2 of 0.63 vs. 0.77 in the native, RMSE 1.08 kcal/mol vs. 0.77 kcal/mol, respectively). The ligand RMSD improved from 3.12 Å to 1.22 Å (Figure 3).

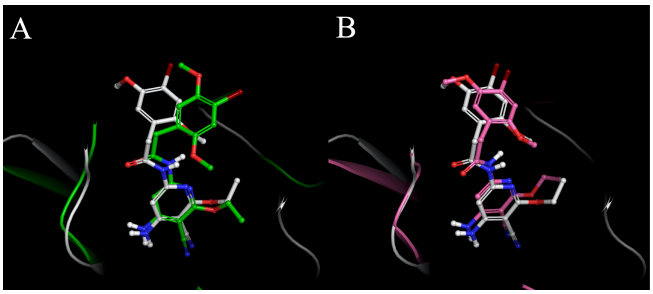

Figure 3. Comparison of the binding mode for the homology model of JNK1 using the 47% sequence identity template 3GC9. (A) Comparison of the IFD-MD pose (green) and the native structure (white). The RMSD of this pose is 3.12 Å, which is dominated by the incorrect position of the terminal chloro-dimethyl benzene ring. (B) The representative frame of the IFD-MD pose after running FEP. The RMSD has improved to 1.22 Å.

IV.D. Limited Binding Affinity Range Could Lead to Unreliable FEP Predictions

Although thrombin modeled on the 3F6U template (42% sequence identity) generated a low-RMSD IFD-MD pose with (RMSD 1.53Å), its RB-FEP performance was slightly inferior to that of a 7.74Å IFD-MD pose (R^2 0.35 vs. 0.41, RMSE 0.96 kcal/mol vs. 0.59 kcal/mol (Table S3). This 7.74 Å pose is a 180º flip (Figure 4). The perturbations in the congeneric series were located adjacent to the solvent-exposed region of the benzene ring. Between the near symmetry of the ligand and narrow, 1.8 kcal/mol activity range of the congeneric series, the relative binding FEP performance of the two poses are nearly equivalent. This is significantly smaller than the typical range of 3-4 kcal/mol activity range for the congeneric series used in our FEP validation publication.

To remedy this, other ligands with perturbations at distinct places should be added to break the symmetry (e.g. perturbations on the amine group in this case). This will increase the binding affinity range of the congeneric series concurrently. This is also a situation in which the AB-FEP score of the different poses can be helpful in differentiating the alternative poses, as is in fact the case here (Table S3).

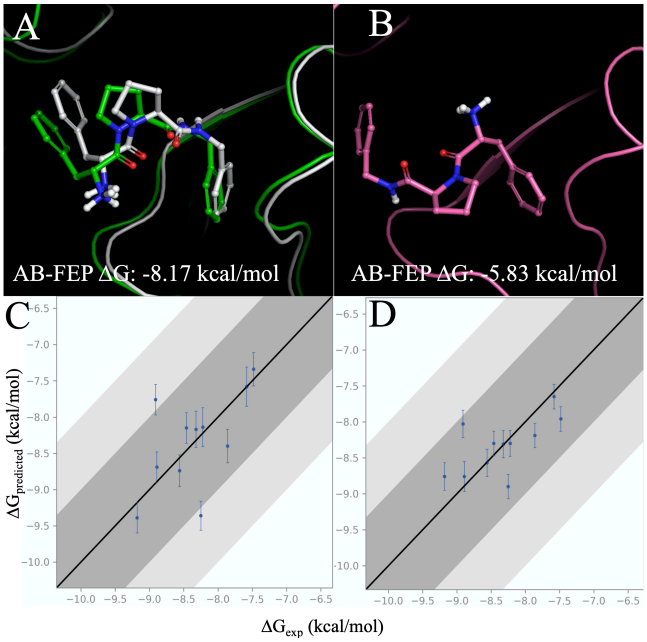

Figure 4. Comparison of a 1.53 Å RMSD pose and a 7.74 Å RMSD pose for thrombin modeled from the 42% sequence identity template, 3F6U. (A). Visualization of the 1.53 Å RMSD pose (green) compared to the native (white). (B) Visualization of the 7.74 Å RMSD pose (pink). (C) RB-FEP plot of predicted ΔG versus experimental ΔG for the 1.53 Å RMSD pose. (D) RB-FEP plot of predicted ΔG versus experimental ΔG for the 7.74 Å RMSD pose

IV.E. Predicted Ligand Poses with Large RMSDs Can Perform Well in RB-FEP if the Perturbations are Localized

System 6 from the proprietary data (30% sequence identity), performed well with RB-FEP although the IFD-MD RMSD and ligand representative RMSD of both poses were greater than 2.5 Å (Table 5). Even though the representative MD frame selected by AB-FEP is 3.72 Å, the large RMSD is localized to the end of the ligand opposite where the perturbations are resulting in very native-like R^2 of 0.47, and RMSE of 0.59 kcal/mol FEP performance. Figure 5 illustrates this situation. Figure 5A shows the crystal structure while Figure 5B shows the pose generated at the end of the complete homology modeling protocol. The congeneric series here is perturbing the terminal chlorinated benzene ring pointing out towards the reader. That ring is in a native-like position and free to execute small rotations being relatively solvent exposed. However, the opposite end of the molecule is incorrectly placed and in a completely buried environment. It is unlikely that any reasonable amount of simulation time can resolve this. While this case is technically a success in terms of satisfaction of our set FEP performance metrics, prospective use of this predicted model would presumably perform poorly once perturbations were made across multiple vectors on the molecule.

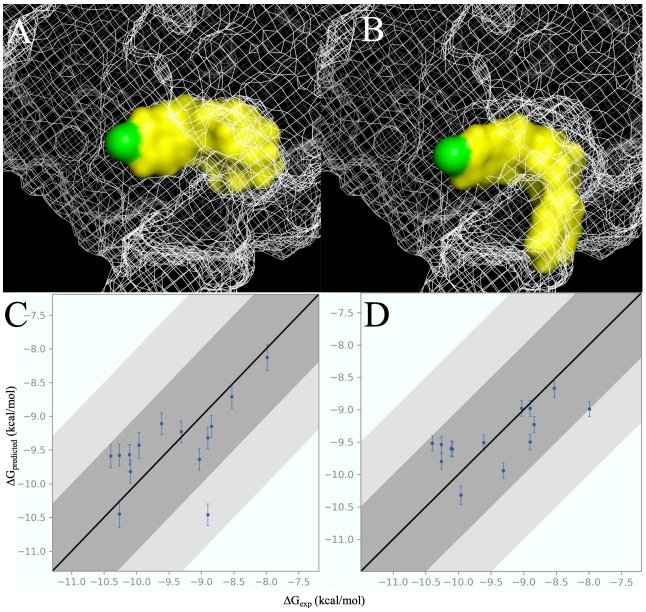

Figure 5. Predicted versus experimental binding mode for system 6 (30% sequence identity) and RB-FEP performance. (A) Mesh and surface representation of the native structure for System 6. (B) Mesh and surface representation of the IFD-MD predicted. This pose has a 3.72 Å ligand RMSD dominated by the incorrect placement of the end of the molecule opposite the chlorinated benzene ring. However, the perturbations in the congeneric series are localized entirely to this chlorinated benzene ring. (C) RB-FEP performance using the native crystal structure. (D) RB-FEP performance using the IFD-MD generated predicted structure.

IV.F. Non-Specific Hydrophobic Contacts Can Degrade the Ability to Discriminate Between Competing Binding Mode Predictions

As mentioned in section III.B, a broad activity range is needed to robustly challenge homology models. This is because a simple null model, for example picking the average ΔG for all ligands, would provide a competitive R^2 and RMSE for series with a narrow activity range.

In general, it is preferable to avoid congeneric series that can be simply explained by a basic descriptor such as molecular weight, polarity, or similarity. Regardless of the series' length, if, a strong correlation exists between physical properties and binding affinity, a simple null model may mask the poor performance of an incorrect homology model. This was noticed when building phosphodiesterases (PDE) homology models. The goal was to construct high-accuracy homology models of PDE10A (PDB ID: 5C2H) using crystal structures from other PDE families, similar to what was done in our previous publication[4]. Here, the PDE10A homology modeling was kept outside of scoring function optimization and therefore served as a test case.

The binding site contacts of the crystal structure 5C2H can be classified into three components:

1. interactions between the pyrimidine core and both Gln716 and Phe719,

2. stacking of the thiazole substituent in the R1 pocket between the hydrophobic side chains of Tyr514, His515 and Phe686,

3. hydrophobic interactions in the R2 pocket (also known as selectivity pocket).

The initial ligand congeneric series was constructed using a 28-ligand subset (Table S6) from a 95-ligand set[22]. The homology modeling template was taken from PDE2A crystal structure, 5U7K, a 37% sequence identity homologue to PDE10A. As shown in Table 8, FEP calculations reported similar performance for both poor- and high-quality models. Figure 6 shows the correlation plots for the two poses. The plot shows that even on a per-ligand basis, the predicted ΔGs are quite similar across the favorably large activity range of 8 kcal/mol.

Table 8. Relative and absolute binding FEP performance for two PDE10A predicted binding modes

Ligand RMSD (Å)	RB-FEP		AB-FEP ΔG (kcal/mol)
	RMSE (kcal/mol)	R^2	
0.97	1.29	0.89	-12.89
4.41	1.32	0.89	-12.67

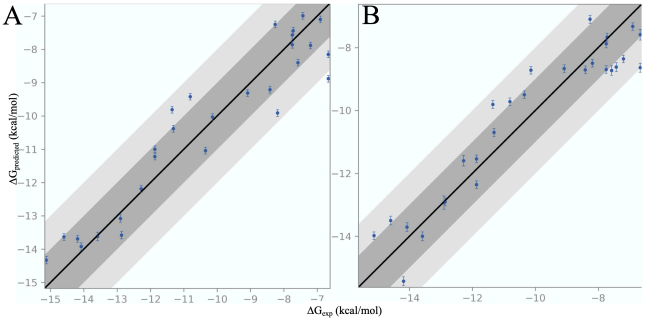

Figure 6. RB-FEP performance (predicted versus experimental ΔG) for PDE10A models. (A). RB-FEP performance for the 0.97 Å ligand RMSD pose (B) RB-FEP performance for the 4.41 Å ligand RMSD pose.

As mentioned, two of the key interactions of PDE10A inhibitors (i.e. ligands) were the contacts in two hydrophobic pockets on both ends of the ligands, which were exactly where the perturbations occurred. Larger hydrophobic ligand substituents resulted in higher ligand efficiency and affinity, which explained the strong correlation between ligand molecular weight and experimental binding affinity.

To validate this hypothesis, we then chose a different subset (Table S7) from the 95-ligand congeneric series, with the intent of picking compounds within a narrow molecular weight range but a broad activity range (Figure 7). For this subset of ligands, the two models showed markedly different performance in relative-binding FEP. Here, the 0.97 Å RMSD maintained an excellent RMSE and R^2 while the 4.41 Å pose showed no correlation between predicted ΔG and experimental ΔG (Table 9).

Table 9. Relative binding FEP performance for two PDE10A predicting binding modes using a subset of ligands for which there is weak correlation between affinity and molecular weight.

Ligand RMSD (Å)	RB-FEP	
	RMSE (kcal/mol)	R^2
0.97	1.36	0.51
4.41	2.05	0.02

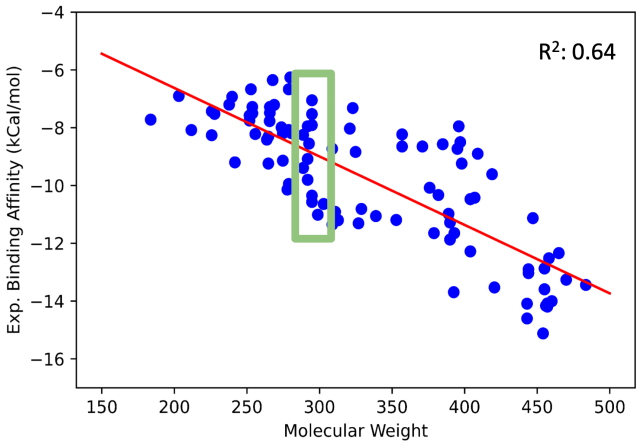

Figure 7. Correlation plot between molecular weight and binding affinity for the 95-ligand PDE10A congeneric series. Boxed in green in a subset of ligands for which there is a broad activity range but no correlation to molecular weight.

V. Discussion

Homology models have seen increasing utilization in drug discovery projects over the past decade. The most common applications of a homology model are to provide a reasonable guess for the binding modes of known active compounds (typically obtained via docking calculations), and more generally to enable optimization of hits and leads via a "structure-guided" proces[23–25]. As the goal of these efforts is to provide physical insight as opposed to quantitative information about binding affinity, it is difficult to assess the robustness of the homology models or the ligand binding modes obtained from them.

Homology models have also been employed to carry out virtual screening of the target, ordinarily via docking calculations for in house or purchasable chemical libraries[26,27]. A number of these virtual screening efforts have reported successfully finding hits, typically in the low micromolar regime. However, using a homology model as generated, without any structural refinement of the active site, will often lead to steric clashes of even tight binding ligands with the protein, making it impossible to recover important classes of active compounds. An example of such a situation is discussed in ref [28] in which a homology model was built for the receptor CDC7 (at a time when no crystal structure of this target was available). The authors concluded that pharmacophore-based docking was superior to docking into their homology model for just this reason.

The results in the present paper unambiguously demonstrate that the application of the IFD-MD methodology to a homology model is capable of transforming the active site to a configuration suitable for a given active ligand with a high degree of robustness. Furthermore, the resulting configuration is often improved further by subsequent molecular dynamics simulation and is a useful starting point for relative binding free energy perturbation calculations for a congeneric series that is compatible with the induced fit structure that IFD-MD has produced.

The opportunities to apply IFD-MD to high quality computationally derived protein structures have recently been qualitatively expanded by the publication of the AlphaFold2 protein structure prediction methodology by Jumper et al.[17]. Results from the CASP14 competition suggested that the transformative combination of machine learning and physics based computation pioneered in AlphaFold2 is capable of producing remarkably accurate (although still imperfect) backbone structures in a blind prediction even when there is no close homologue to the target in the Protein Data Bank- for example, when the closest sequence identities are in the 20-30% range[29]. Initial preliminary investigations that we have carried out support the proposition that, at the very least, AlphaFold2 structures are in many cases a significant improvement over those derived from alternative prediction approaches for a wide range of interesting pharmaceutical targets.

However, the AlphaFold2 methodology does not take ligands, or the associated induced fit effects of ligand binding on the receptor structure, into account. In order to render AlphaFold2 structures suitable for driving structure-based drug design projects (e.g. via virtual screening and/or FEP computations for lead optimization) it is necessary to reorganize the binding site to accommodate a given ligand series, precisely as has been carried out in the present paper. Furthermore, there are cases where refinement of loop regions is also required, and this potentially can be accomplished by a physics-based approach such as is available in the Prime program[10,18,30]. Given that AlphaFold2 models have been produced for every known protein in the human genome, there is in principle an opportunity to create effective structure-based drug design projects for hundreds, if not thousands, of important pharmaceutical targets that previously were accessible only via ligand-based approaches.

A key aspect of the refinement protocol that we have described above is the use of binding data from a ligand series to choose between multiple options for the protein-ligand complex structure. The ambiguity and noise that is present in a typical homology model, even with the most recent advances, can be addressed by differentiating proposed options with ligand-based information. As the structural ambiguity grows (e.g. multiple options for loop conformations), the need for resolution of options with an even greater quantity (and diversity) of ligand binding data can only increase.

One way to interpret these results is that the IFD-MD and FEP calculations provide a way to combine protein structure prediction and ligand binding information. In a pharmacophore methodology, one is in effect using ligand binding data to try to map out the interactions of the ligand in the binding pocket; indeed, the COMFA approach[31,32], an early 3D ligand-based algorithm, a "molecular field" representing the inferred attractiveness of various regions of the receptor to different types of ligand atoms was assembled from the ligand binding

data. In the IFD-MD/FEP based protocol proposed here, rather than assembling a guess for the "molecular field" from scratch (an extremely difficult task given a limited amount of ligand binding data), we instead use the ligand binding data to choose between a highly restricted number of alternatives which have been constructed by using all of the information in the PDB and in sequence databases (as processed and deployed for example by the AlphaFold2 algorithm). Provided that suitable candidate structures can in fact be generated by the protein-based component of the algorithm (more than one may be needed, for example to accommodate DFG-in and DFG-out binders to kinases), this approach should provide increased reliability and accuracy as compared to a pharmacophore-based approach, while avoiding the problem of sterically precluding the binding of key active compounds via the reorganizational capability of IFD-MD.

VI. Conclusion

We have shown that the IFD-MD induced fit docking methodology is capable of obtaining a good initial pose within the top 5 scoring poses, typically within 2.5 Å of the native structure, upon docking of a known active ligand into a homology model, in cases where the sequence identify of the target to the template is in the range of 30-50%. This domain of applicability may well be extended via the use of AlphaFold2 structures.

The correct pose can then be identified via a combination of relative binding FEP calculations for a congeneric series of active ligands, and AB-FEP calculations on the candidate poses. The molecular dynamics simulations used to equilibrate the relative binding FEP often lead to improved RMSDs as compared to the original poses generated by IFD-MD, and representative structures from the MD simulations are therefore used in the AB-FEP calculations.

Our consistent results across a wide range of targets and ligand series engender optimism that the combined IFD-MD/FEP approach described herein can be utilized in a diversity of applications in drug discovery projects based on homology modeling, including virtual screening, hit to lead discovery, and lead optimization. A key question is whether the protocols can succeed for particularly challenging but important target classes such as GPCRs. Further work will address this issue in the near future.

ASSOCIATED CONTENT

Supporting Information

Detailed results for each homology modeling case and the individual poses generated by IFD-MD and GlideSP, parameter values for the composite scoring function, congeneric subseries selected for PDE10A validation (PDF)

Coordinates of public retrospective predictions (ZIP)

AUTHOR INFORMATION

Corresponding Author

* Edward B. Miller – Email: Ed.Miller@schrodinger.com

Present Addresses

‡ Odyssey Therapeutics, 301 Binney Street, Cambridge, MA 02142

§ Amgen Inc., One Amgen Center Drive, Thousand Oaks, CA 91320

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare the following competing financial interest(s): R.A.F. has a significant financial stake in Schrödinger, Inc., is a consultant to Schrödinger, Inc., and is on the Scientific Advisory Board of Schrödinger, Inc.

ABBREVIATIONS

IFD-MD, induced-fit docking molecular dynamics; FEP, free-energy perturbation; AB-FEP, absolute binding free-energy perturbation; RB-FEP, relative binding free-energy perturbation; LOO, leave-one-out optimization.

REFERENCES

(1) 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. *J. Chem. Theory Comput.* **2021**, *17* (4), 2630–2639. https://doi.org/10.1021/acs.jctc.1c00136.

(2) Cappel, D.; Hall, M. L.; Lenselink, E. B.; Beuming, T.; Qi, J.; Bradner, J.; Sherman, W. Relative Binding Free Energy Calculations Applied to Protein Homology Models. *J. Chem. Inf. Model.* **2016**, *56* (12), 2388–2400. https://doi.org/10.1021/acs.jcim.6b00362.

(3) Sherman, W.; Day, T.; Jacobson, M. P.; Friesner, R. A.; Farid, R. Novel Procedure for Modeling Ligand/Receptor Induced Fit Effects. *J. Med. Chem.* **2006**, *49* (2), 534–553. https://doi.org/10.1021/jm050540c.

(4) Moraca, F.; Negri, A.; de Oliveira, C.; Abel, R. Application of Free Energy Perturbation (FEP+) to Understanding Ligand Selectivity: A Case Study to Assess Selectivity Between Pairs of Phosphodiesterases (PDE's). *J. Chem. Inf. Model.* **2019**, *59* (6), 2729–2740. https://doi.org/10.1021/acs.jcim.9b00106.

(5) Clark, A. J.; Gindin, T.; Zhang, B.; Wang, L.; Abel, R.; Murret, C. S.; Xu, F.; Bao, A.; Lu, N. J.; Zhou, T.; Kwong, P. D.; Shapiro, L.; Honig, B.; Friesner, R. A. Free Energy Perturbation Calculation of Relative Binding Free Energy between Broadly Neutralizing Antibodies and the Gp120 Glycoprotein of HIV-1. *J. Mol. Biol.* **2017**, *429* (7), 930–947. https://doi.org/10.1016/j.jmb.2016.11.021.

(6) Clark, A. J.; Negron, C.; Hauser, K.; Sun, M.; Wang, L.; Abel, R.; Friesner, R. A. Relative Binding Affinity Prediction of Charge-Changing Sequence Mutations with FEP in Protein–Protein Interfaces. *J. Mol. Biol.* **2019**, *431* (7), 1481–1493. https://doi.org/10.1016/j.jmb.2019.02.003.

(7) Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J.; Romero, D. L.; Masse, C.; Knight, J. L.; Steinbrecher, T.; Beuming, T.; Damm, W.; Harder, E.; Sherman, W.; Brewer, M.; Wester, R.; Murcko, M.; Frye, L.; Farid, R.; Lin, T.; Mobley, D. L.; Jorgensen, W. L.; Berne, B. J.; Friesner, R. A.; Abel, R. Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field. *J. Am. Chem. Soc.* **2015**, *137* (7), 2695–2703. https://doi.org/10.1021/ja512751q.

(8) Chen, W.; Cui, D.; Abel, R.; Friesner, R. A.; Wang, L. Accurate Calculation of Absolute Protein-Ligand Binding Free Energies. *ChemRxiv* **2022**. https://doi.org/10.26434/chemrxiv-2022-2t0dq.

(9) Coordinators, N. R. Database Resources of the National Center

for Biotechnology Information. *Nucleic Acids Res.* **2016**, *44* (D1), D7–D19. https://doi.org/10.1093/nar/gkv1290.

(10) Schrödinger Release 2021-4: Prime. Schrödinger, Inc: New York, NY 2020.

(11) Li, J.; Abel, R.; Zhu, K.; Cao, Y.; Zhao, S.; Friesner, R. A. The VSGB 2.0 Model: A next Generation Energy Model for High Resolution Protein Structure Modeling. *Proteins Struct. Funct. Bioinforma.* **2011**, *79* (10), 2794–2812. https://doi.org/10.1002/prot.23106.

(12) 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. https://doi.org/10.1021/acs.jmedchem.6b00131.

(13) Clark, A. J.; Tiwary, P.; Borrelli, K.; Feng, S.; Miller, E. B.; Abel, R.; Friesner, R. A.; Berne, B. J. Prediction of Protein–Ligand Binding Poses via a Combination of Induced Fit Docking and Metadynamics Simulations. *J. Chem. Theory Comput.* **2016**, *12* (6), 2990–2998. https://doi.org/10.1021/acs.jctc.6b00201.

(14) Edgar, R. C. MUSCLE: A Multiple Sequence Alignment Method with Reduced Time and Space Complexity. *BMC Bioinformatics* **2004**, *5* (1), 113. https://doi.org/10.1186/1471-2105-5-113.

(15) Edgar, R. C. MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput. *Nucleic Acids Res.* **2004**, *32* (5), 1792–1797. https://doi.org/10.1093/nar/gkh340.

(16) Remmert, M.; Biegert, A.; Hauser, A.; Söding, J. HHblits: Lightning-Fast Iterative Protein Sequence Searching by HMM-HMM Alignment. *Nat. Methods* **2012**, *9* (2), 173–175. https://doi.org/10.1038/nmeth.1818.

(17) 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. https://doi.org/10.1038/s41586-021-03819-2.

(18) 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 Struct. Funct. Bioinforma.* **2004**, *55* (2), 351–367. https://doi.org/10.1002/prot.10613.

(19) Jacobson, M. P.; Friesner, R. A.; Xiang, Z.; Honig, B. On the Role of the Crystal Environment in Determining Protein Side-Chain Conformations. *J. Mol. Biol.* **2002**, *320* (3), 597–608. https://doi.org/10.1016/S0022-2836(02)00470-9.

(20) Roos, K.; Wu, C.; Damm, W.; Reboul, M.; Stevenson, J. M.; Lu, C.; Dahlgren, M. K.; Mondal, S.; Chen, W.; Wang, L.; Abel, R.; Friesner, R. A.; Harder, E. D. OPLS3e: Extending Force Field Coverage for Drug-Like Small Molecules. *J. Chem. Theory Comput.* **2019**, *15* (3), 1863–1874. https://doi.org/10.1021/acs.jctc.8b01026.

(21) Schrödinger Release 2021-4: FEP+. Schrodinger, Inc: New York, NY 2020.

(22) Shipe, W. D.; Sharik, S. S.; Barrow, J. C.; McGaughey, G. B.; Theberge, C. R.; Uslaner, J. M.; Yan, Y.; Renger, J. J.; Smith, S. M.; Coleman, P. J.; Cox, C. D. Discovery and Optimization of a Series of Pyrimidine-Based Phosphodiesterase 10A (PDE10A) Inhibitors through Fragment Screening, Structure-Based Design, and Parallel Synthesis. *J. Med. Chem.* **2015**, *58* (19), 7888–7894. https://doi.org/10.1021/acs.jmedchem.5b00983.

(23) Dong, L.; Feng, R.; Bi, J.; Shen, S.; Lu, H.; Zhang, J. Insight into the Interaction Mechanism of Human SGLT2 with Its Inhibitors: 3D-QSAR Studies, Homology Modeling, and Molecular Docking and Molecular Dynamics Simulations. *J. Mol. Model.* **2018**, *24* (4), 86. https://doi.org/10.1007/s00894-018-3582-2.

(24) de la Fuente, T.; Martín-Fontecha, M.; Sallander, J.; Benhamú, B.; Campillo, M.; Medina, R. A.; Pellissier, L. P.; Claeysen, S.; Dumuis, A.; Pardo, L.; López-Rodríguez, M. L. Benzimidazole Derivatives as New Serotonin 5-HT6 Receptor Antagonists. Molecular Mechanisms of Receptor Inactivation. *J. Med. Chem.* **2010**, *53* (3), 1357–1369. https://doi.org/10.1021/jm901672k.

(25) Batran, R. Z.; Dawood, D. H.; El-Seginy, S. A.; Maher, T. J.; Gugnani, K. S.; Rondon-Ortiz, A. N. Coumarinyl Pyranopyrimidines as New Neuropeptide S Receptor Antagonists; Design, Synthesis, Homology and Molecular Docking. *Bioorg. Chem.* **2017**, *75*, 274–290. https://doi.org/https://doi.org/10.1016/j.bioorg.2017.09.017.

(26) Li, S.; Zhou, Y.; Lu, W.; Zhong, Y.; Song, W.; Liu, K.; Huang, J.; Zhao, Z.; Xu, Y.; Liu, X.; Li, H. Identification of Inhibitors against P90 Ribosomal S6 Kinase 2 (RSK2) through Structure-Based Virtual Screening with the Inhibitor-Constrained Refined Homology Model. *J. Chem. Inf. Model.* **2011**, *51* (11), 2939–2947. https://doi.org/10.1021/ci2002445.

(27) Nguyen, T. L.; Gussio, R.; Smith, J. A.; Lannigan, D. A.; Hecht, S. M.; Scudiero, D. A.; Shoemaker, R. H.; Zaharevitz, D. W. Homology Model of RSK2 N-Terminal Kinase Domain, Structure-Based Identification of Novel RSK2 Inhibitors, and Preliminary Common Pharmacophore. *Bioorg. Med. Chem.* **2006**, *14* (17), 6097–6105. https://doi.org/https://doi.org/10.1016/j.bmc.2006.05.001.

(28) Lindvall, M.; McBride, C.; McKenna, M.; Gesner, T. G.; Yabannavar, A.; Wong, K.; Lin, S.; Walter, A.; Shafer, C. M. 3D Pharmacophore Model-Assisted Discovery of Novel CDC7 Inhibitors. *ACS Med. Chem. Lett.* **2011**, *2* (10), 720–723. https://doi.org/10.1021/ml200029w.

(29) 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.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Applying and Improving AlphaFold at CASP14. *Proteins* **2021**, *89* (12), 1711–1721. https://doi.org/10.1002/prot.26257.

(30) Miller, E. B.; Murrett, C. S.; Zhu, K.; Zhao, S.; Goldfeld, D. A.; Bylund, J. H.; Friesner, R. A. Prediction of Long Loops with Embedded Secondary Structure Using the Protein Local Optimization Program. *J. Chem. Theory Comput.* **2013**, *9* (3), 1846–1864. https://doi.org/10.1021/ct301083q.

(31) Cramer, R. D.; Patterson, D. E.; Bunce, J. D. Comparative Molecular Field Analysis (CoMFA). 1. Effect of Shape on Binding of Steroids to Carrier Proteins. *J. Am. Chem. Soc.* **1988**, *110* (18), 5959–5967. https://doi.org/10.1021/ja00226a005.

(32) Clark, M.; Cramer, R. D.; Jones, D. M.; Patterson, D. E.; Simeroth, P. E. Comparative Molecular Field Analysis (CoMFA). 2. Toward Its Use with 3D-Structural Databases. *Tetrahedron Comput. Methodol.* **1990**, *3* (1), 47–59. https://doi.org/10.1016/0898-5529(90)90120-W.