A Reliable and Accurate Solution to the Induced Fit Docking Problem for Protein-Ligand Binding

Edward B. Miller[†], Robert B. Murphy[‡], Daniel Sindhikara[†], Kenneth W. Borrelli[†], Matthew J. Grisewood[†], Fabio Ranalli[†], Steven L. Dixon[†], Steven Jerome[‡], Nicholas A. Boyles[⊥], Tyler Day[†], Phani Ghanakota[†], Sayan Mondal[†], Salma B. Rafi[⊥], Dawn M. Troast[§], Robert Abel[†], Richard A. Friesner[*,∥]

†Schrödinger, Inc., 120 West 45th Street, New York, New York 10036, United States

‡ Schrödinger, Inc., 10201 Wateridge Circle, San Diego, California 92121, United States

⊥Schrödinger, Inc., 101 SW Main Street, Portland, Oregon 97204, United States

§Morphic Therapeutic, 35 Gatehouse Drive, Waltham, Massachusetts 02451, United States

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

ABSTRACT: We present a reliable and accurate solution to the induced fit docking problem for protein-ligand binding by combining ligand-based pharmacophore docking, rigid receptor docking, and protein structure prediction with explicit solvent molecular dynamics simulations. This novel methodology in detailed retrospective and prospective testing succeeded to determine protein-ligand binding modes with a root-mean-square-deviation within 2.5 Å in over 90% of cross-docking cases. We further demonstrate these predicted ligand-receptor structures were sufficiently accurate to prospectively enable predictive structure-based drug discovery for challenging targets, substantially expanding the domain of applicability for such methods.

I Introduction

The prediction of ligand binding modes and protein-ligand complex structures is fundamental to modern approaches to small molecule structure-based drug discovery. Experimental methods such as x-ray crystallography, or cryo-electron microscopy (cryo-EM) can provide starting points for such predictions at or near an atomic level of resolution. However, the cost of experimentally obtaining structures for new ligands ranges from non-trivial to extremely large (e.g. for many GPCR projects). The ability to rapidly and reliably generate accurate ligand binding modes (given an experimental structure as a starting point) across a wide range of systems would qualitatively transform the impact of computational methodologies, opening up many new targets to a structure-based approach, and enabling structures to be obtained for a large number of active compounds (e.g. from a virtual screening or HTS campaign). And an extension of this approach to enable the generation of accurate ligand binding modes for homology models would greatly increase the number of interesting drug targets amenable to structure-based discovery.

If the receptor remains relatively rigid when a new ligand is introduced, the structure prediction problem becomes tractable in most cases with a minimal level of computational effort, as the great majority of small molecule drug-like ligands have a small number of rotatable bonds. A number of sophisticated docking programs (e.g. Gold[1], Glide[2], Dock[3], and FlexX[4]) have been developed to address rigid receptor docking. There is some variation in performance (much of which depends upon the scoring function) and advances in this area continue to be made (e.g. the introduction of explicit waters into the docking model as in the WScore methodology[5]). As long as induced fit effects do not preclude docking due to steric clashes, the prediction capability of all of these programs is respectable (and for the best performers, is quite robust and accurate).

The situation is very different when induced fit effects are important. For individual cases, where one is willing to invest a large amount of human and/or computation time, anecdotal successes continue to be reported. For example, the DESRES group has used molecular dynamics simulations to dock ligands into a receptor conformation requiring induced fit movements to achieve the correct pose[6]. However, on the order of 30-300 microseconds of total simulation per ligand were required to reliably observe binding events in their simulations. The robustness of the protocol across many different protein-ligand complexes is not known, since the computational requirements precluded studying more than a small number of examples. Various approaches have been used to predict poses in docking competitions such as GPCRdock[7,8], with the best performing groups achieving low RMSDs. However, no evidence has been presented that these methods can be applied to a large number of ligands in an automated fashion, across a wide range of receptors.

The Schrodinger Induced Fit Docking (IFD) methodology was developed a number of years ago to address the induced fit docking challenge. IFD has been utilized successfully by many pharmaceutical and biotechnology companies in projects for more than a decade. However, it lacks the robustness and ac-

curacy that would render it a true solution, as opposed to a useful tool, as will be demonstrated below by performance assessments using large data sets. Problems can arise in both sampling (can a correct pose be generated starting from the initial structure) and scoring (can this pose be selected as top ranking, or, failing that, ranking among the top two poses). After extensive experimentation within the historical IFD framework, we concluded that major innovations are required to address the problematic cases.

To address this clear challenge, we have combined several technologies, including pharmacophore analysis[9–11], docking[5], metadynamics[12], molecular dynamics[13,14], and energy-guided protein structural modeling[15–17] into a self-consistent and highly computationally efficient approach to predict protein-ligand binding modes within a potentially significantly flexible and dynamic protein biding site. We refer to this new methodology as IFD-MD. The method is computationally much more efficient than brute-force MD simulations, and can easily be completed overnight using modest cloud computing resources.

The paper is organized as follows. In Section II, we briefly summarize the computational workflow. Section III describes the performance of the algorithm on publicly available data sets, application to free energy perturbation calculations, and usage in the context of drug-discovery projects, both retrospectively and prospectively. More specifically, Section III.B discusses the performance on training and test sets taken from the PDB. Section III.C describes the computational cost of the algorithm, including a recently developed modification which provides a ~5x reduction in computational effort with no loss of accuracy. Section III.D describes the use of the IFD-MD workflow to generate predicted ligand-receptor complexes, and compares their utility in carrying out predictive free energy perturbation (FEP) calculations for congeneric series (typically employed to support hit-to-lead and lead optimization efforts in structure based drug discovery projects) to the same FEP calculations starting from a cocrystallized structure of one of the ligands. Finally, Section III.E describes the performance of the workflow in retrospective evaluation, where prediction of proprietary crystal structures is performed, and in prospective evaluation where the workflow predictions preceded the availability of any solved structure.

II Description of the IFD-MD Workflow

The current IFD-MD workflow addresses the following challenge: how can one predict the binding mode of a ligand into a protein, where the sidechains of the protein binding site may reorganize upon binding the ligand (small modifications of the backbone conformation are also often handled adequately)? Because of the large number of degrees of freedom available to the ligand and the protein, a brute force computational strategy is likely intractable for many cases, and certainly will not meet the throughput needs of modern pharmaceutical drug discovery. However, by combining a wide variety of computational techniques, including, pharmacophore modeling, docking, metadynamics, molecular dynamics, and energy-guided protein structural modeling, we have been able to develop methodology to approach this problem which retains high accuracy and good computational efficiency.

The workflow begins by requiring an input holo-structure. This contains the starting receptor and what is referred to as a template ligand. The ligand whose binding mode is being predicted is referred to as the target ligand.

A summary of the new approach can be outlined as follows:

1. Ligand pharmacophore analysis, using the template ligand as a reference conformation, is used to generate a large and diverse number of possible target ligand binding modes within the ligand binding site of the protein. Initially, many of these binding modes will contain clashes with the template receptor structure; these clashes provide useful information as to where modifications of the receptor conformation may be needed to construct the correct binding mode of the target ligand.
2. Along these lines, protein sidechains which clash with the pharmacophore-docked ligand positions are refined to produce alternative protein conformations.
3. Glide[2] rigid receptor docking is then used to dock the ligand into theses catalogued alternative protein conformations, followed by energy-guided relaxation of the protein side chains and redocking of the ligand into each of these computationally determined trial protein-ligand binding modes, each with a self-consistently determined docking score and protein structural modeling energy value, calculated using the standard Prime energy model (based on the OPLS-3e[18] force field, and the VSGB 2.0 continuum solvation model[19]).
4. These trial protein-ligand binding modes are then clustered by mutual RMSD, and twenty cluster representatives with favorable docking scores and protein structural modeling energy values are advanced into a detailed rescoring procedure (steps 5-7).
5. The detailed rescoring of these twenty trial binding modes begins with each binding mode being subjected to 500 ps molecular dynamics simulation and grand canonical water equilibration. Ten different trial MD runs are generated, the last MD frame produced by each run is evaluated using the WScore scoring function which we have discussed in previous work[5]. Explicit waters from the MD simulation can, however, be used to modify WScore assessment of the desolvation of polar or charged groups.
6. These equilibrated binding modes are then advanced into metadynamics simulations where a gradually increasing force assesses the local stability of the trial binding mode. A quantitative analysis of the quality of the protein-ligand contacts maintained is again performed for binding mode geometries observed during simulations;
7. Finally, the results of above analysis steps 4-6- specifically, the best WScore obtained in the 500 psec equilibration runs, and two metadynamics metrics evaluating pose displacement and hydrogen bond persistence, are combined with the Prime energy and Glide docking scores, plus a number of other empirical terms, to yield a composite scoring function that provides the final ranking of the candidate poses.

A detailed discussion of the complete scoring workflow, the utilized functional forms and their parametrizations is available in the supporting information.

III Results and Discussion

III.A Construction of the Training and Test Sets

Our IFD-MD training and test sets are constructed from 280 PDB complexes, spanning 41 targets, which were previously utilized in the development of the WScore docking methodology[5]. For these complexes, extensive effort has been spent on protein preparation, including the key task of assigning protonation states. The preparation of each complex has been validated via self-docking. A complete list of targets and PDB codes are given in the supporting information in Tables S3 through S5. Additionally, all inputs and outputs for both the training and test set, as well as an expanded test set (Section III.C) are available in the supporting information as PDB-formatted files.

Cross-docking cases are assembled such that the dominant receptor motion involved in the binding site involve side chain atoms. This includes cases where there is no induced-fit effect observed at all. These cases serve as negative controls, ensuring that the algorithm could successfully dock a ligand without performing unnecessary receptor conformational changes. Additional information regarding the construction of the training sets and tests sets can be found in the supporting information in Section XI.

The training set utilized here is composed of 258 cross-docks across 41 targets. The test set consists of 157 cross-docks across 24 targets, a subset of the 41 targets used for the training set. Methods development and parametrization was performed exclusively on the training set leaving the test set results as an independent evaluation. The supporting information lists the targets and the number of cross-docks for each target within the training and test set.

A proxy for the difficulty of the cross-dock is the Tanimoto similarity between the template ligand and the target ligand. Across all cross-docking cases, the median Tanimoto similarity is 0.04 with 75% of cross-docks having a similarity less than 0.07 and 90% of cross-docks with a similarity less than 0.23. The Tanimoto similarity is computed using linear fingerprints with Daylight invariant atom types[20,21] and bonds distinguished by their formal bond order. Based on these low similarities, the template and target ligands are almost entirely non-congeneric.

III.B IFD-MD Performance

For the development of IFD-MD, cross-docks are divided into either a training or test set with the test set excluded for the optimization of any parameters. The separation between training and test-set is roughly a 60%/40% split.

The results shown here are those obtained after the complete IFD-MD workflow has been performed. All RMSDs are ligand-heavy atom RMSD relative to the native crystal structure, the crystal structure of the bound target ligand. To account for solvent exposed tails, each predicted structure is also compared against a 100 ns MD trajectory run on the native crystal structure. This is used to quantitatively determine whether the proposed solvent-exposed tail lies within the same energy basin as the crystal structure within explicit solvent MD. When this occurs, the RMSD of the ligand is adjusted in comparison to the native MD, rather than the native crystal structure. There are six cases where this rescues a failure; these cases are listed in the supporting information, Table S1.

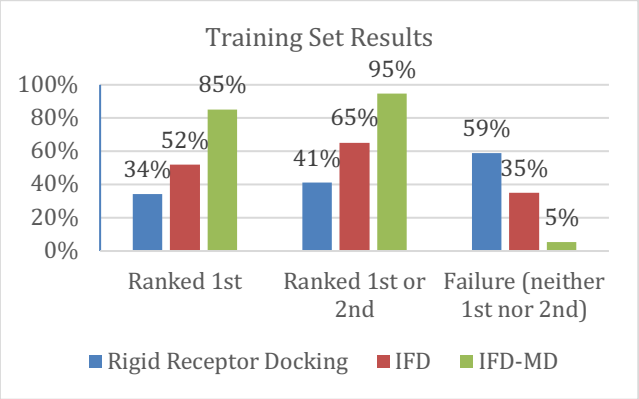

Figure 1. Training set results consisting of 258 cross-docks run using IFD-MD, the previous IFD release, and rigid receptor docking. A case is considered successful if its ligand heavy-atom RMSD is within 2.5 Å of the crystal structure. Note that these cross-docks are a training set only for IFD-MD. They were not used for any parameter optimization with IFD.

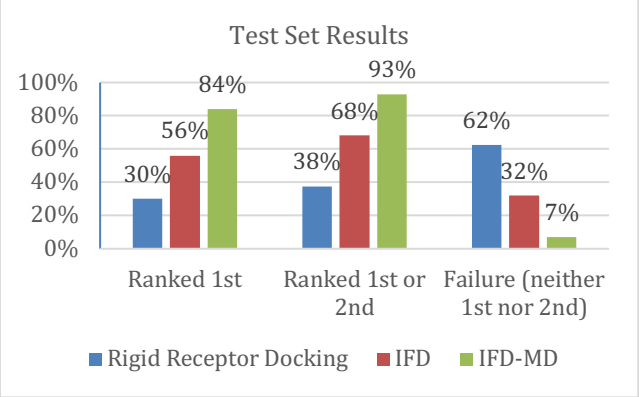

Figure 2. Test set results consisting of 157 cross-docks run using IFD-MD, the previous IFD release, and rigid receptor docking. A case is considered successful if its ligand heavy-atom RMSD is within 2.5 Å of the crystal structure.

Figure 1 shows the performance of IFD-MD compared to IFD and rigid receptor docking for the training set of 258 cross-docks. It should be noted that these cases are a training set for only IFD-MD and that neither the parameters for IFD nor rigid receptor docking are altered as a result of these cases. Figure 2 shows a similar plot for the 157 cross-docks in the test set.

Comparing the training set and test set results, IFD-MD results degrade by 2% suggesting parameters are not materially overfit to the training set. For either the training or the test-set, IFD-MD significantly outperforms both rigid receptor docking and IFD. In particular, IFD performs slightly better than 50% in producing a top ranked pose under 2.5 Å while IFD-MD is above 80% for a rank 1 pose, and above 90% if one explores the top two poses.

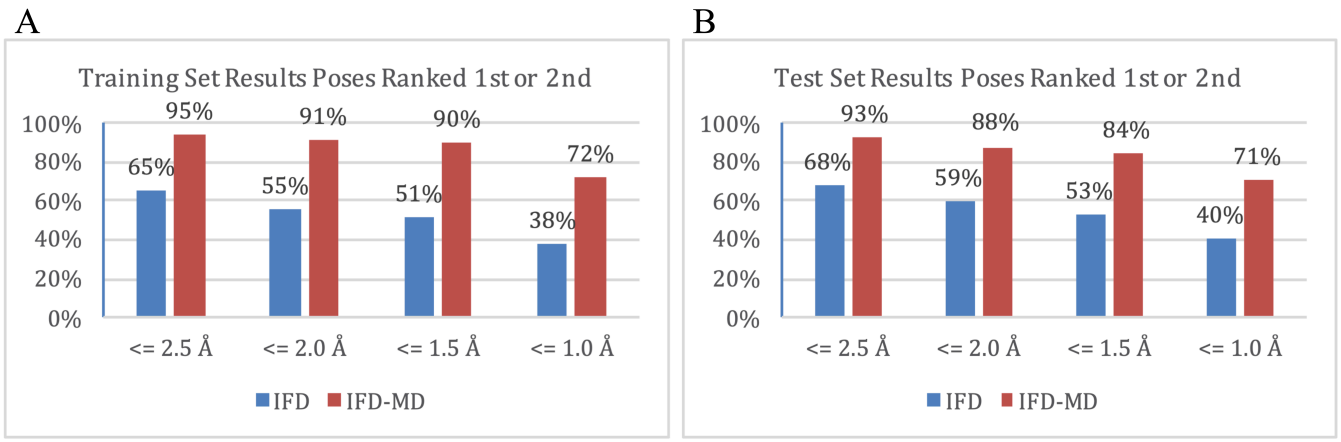

Figure 3. Percent of cases by ligand heavy atom RMSD for the top two ranked predictions. (A) Results for the 258 training set cross-docks. (B) Results for the 157 test set cross-docks.

Another interesting question is how well IFD-MD and the older IFD methodology perform on cases where rigid receptor docking succeeds. There are 165 cases in total, across the training and test sets, in this category. IFD-MD produces a sub-2.5 Å pose on 96% of these cases, as opposed to IFD which succeeds in 82% of cases. In prospective applications, where induced-fit effects may not always be required, the ability to maintain the accuracy of rigid receptor docking is necessary for robustness.

We next evaluate the comparative performance of IFD-MD across a wide range of RMSD cutoffs. Figure 3 shows training and test set results plotted by RMSD. For both the training and test set, IFD-MD has approximately 90% of the cases 2 Å RMSD or better compared to less than 60% of the cases with IFD for both training and test set.

Finally, we discuss an example illustrating our reasons for choosing a 2.5 Å RMSD cutoff as our primary success criterion. Figure 4 displays an HIV-RT cross-dock where the ligand from PDB ID 1FKP is docked into the receptor for PDB ID 2B5J. Comparing 1FKP with 2B5J, there is backbone motion of a hairpin which in the native 1FKP structure packs closer into the binding site compared to the receptor in PDB ID 2B5J.

Figure 4 shows the result from this cross-dock. IFD-MD, unable to execute the hairpin motion, shifts the ligand slightly to compensate, resulting in a 2.4 Å ligand heavy-atom RMSD. With the native ligand forming no hydrogen bonds to the receptor, the IFD-MD ligand reproduces the same hydrophobic and aromatic CH interactions as the native-ligand.

A prediction of this type provides useful information from the point of view of visual inspection, and possibly is a starting point for virtual screening or free energy perturbation calculations. We therefore made the somewhat arbitrary decision to classify such cases as successes. A more complete picture of performance is provided in Figure 3.

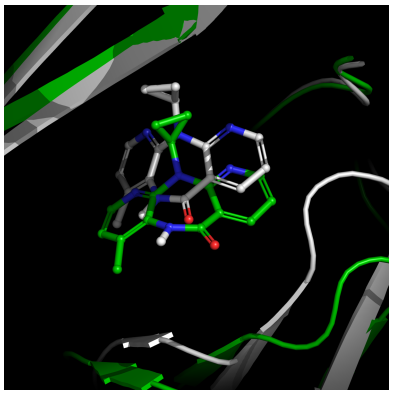

Figure 4. Top ranked IFD-MD prediction from an HIV-RT cross-dock consisting of the ligand from PDB ID 1FKP docked into the receptor from PDB ID 2B5J. The 1FKP crystal structure is shown in white and the IFD-MD prediction in green. The IFD-MD prediction is a 2.4 Å ligand heavy-atom RMSD pose. In the bottom right is the motion of a hairpin moving further into the binding site which is not present in 2B5J and is not reproduced by IFD-MD. Rather, IFD-MD shifts the ligand further down to compensate, forming the same overall interactions as the native.

III.C Computational Cost

The computational cost of the complete algorithm is a function of the size of the receptor, and for the MD calculations, the performance of the available GPU hardware. Averaged across the 415 cases within both the training and test set, the computational cost for an IFD-MD calculation is 400 CPU hours, 250 GPU hours, using an NVIDIA P100 GPU card. A significant portion of the algorithm is parallelizable such that the entire algorithm can be completed in one to two days depending on available compute resources.

Very recently, we constructed an intermediate scoring function to improve the algorithm's efficiency. This intermediate scoring function, the efficacy of which is significantly enhanced by the use of WScore results (which require only the 500 psec MD simulations) triages the top twenty poses down to five using a function that ranks the poses without the expensive metadynamics calculations. The remaining top five poses are subsequently scored with the complete scoring function whose accuracy is established in Section III.B. Combined with other

minor improvements, the computational cost of the MD component has been reduced by a factor of five such that the current computational cost is 400 CPU hours, 50 GPU hours. A typical configuration of 100 CPUs and 8 GPUs would see a wall-clock time of 12 hours. With additional CPU resources and the use of state-of-the-art NVIDIA V100 cards, the wall-clock time can be reduced further to 6 hours.

The intermediate scoring function was parameterized using the complete training and test set. The accuracy of the complete IFD-MD algorithm was preserved with 96% of cases having a pose within 2.5 Å ranked 1st or 2nd across the combined training and test set.

An expanded test set was therefore also introduced containing 19 novel cross-docks which were excluded from all parameter optimization, including the parameterization of the intermediate scoring function. These cross-docks are drawn from 11 new targets not present in the training or test set. The performance across this set is shown in Figure 5. The identity of these cross-docks is listed in supporting information Table S7. Given the limited size of this expanded set and the modest increase in difficulty evident in the rigid receptor docking results, the performance on this expanded test set appears consistent with our previous data sets. We note that the introduction of entirely new targets also tests IFD-MD in a new dimension as compared to the prior test set, one that is highly relevant to use in drug discovery projects where novel targets are often being addressed.

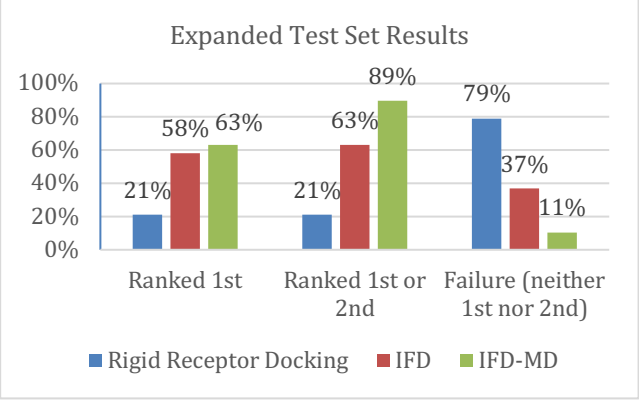

Figure 5. Performance of IFD-MD with an expanded test set introduced after the creation of an intermediate scoring function to reduce computational cost. The expanded test set is composed of 19 cross-docks across 11 targets. These targets were not present in the previous training or test sets.

III.D Use of IFD-MD Structures in Free Energy Perturbation Calculations

While RMSD represents one metric assessing the performance of IFD-MD in flexible ligand docking, a second relevant measure of the utility of this approach for structure based drug discovery efforts is the performance of IFD-MD structures in free energy perturbation (FEP) calculations. Consider for example a typical situation in which binding affinities for a congeneric ligand series are available in a publication or a patent, but a co-crystallized structure of one of the congeneric ligands with the target receptor is not available. In such a case, validation of FEP performance for the series requires generation of accurate and reliable binding modes as a starting point for the FEP simulations. In this section, we carry out initial tests of IFD-MD for four systems that we have studied previously[22], using a co-crystal structure for each system containing a ligand bearing little resemblence to the congeneric series of interest.

The four receptors that we investigate here have been pharmaceutical targets for many years: BACE, CDK2, PTP1B, and Thrombin. We choose a template receptor for each system by locating a PDB complex for which the co-crystallized ligand has a low similarity to the congeneric series under study while at the same time does not have major backbone motions (vs. those of the congeneric ligand crystal structure) that would make it difficult for the current version of IFD-MD to successfully dock the series. We define low similarity by evaluating the tanimoto similairity across all pairs within the congeneric series. We require that the template ligand have a lower tanimoto similarity to all ligands in the series than the least similar pair within the series. We choose the largest ligand in the series to dock with IFD-MD; this is not necessarily the optimal strategy, but it works well for the cases that we investigate here. Development of a robust protocol will require investigating many additional systems, including more challenging cases such as those using homology models as a template. We will report on an initial study along these line in an upcoming publication.

A key idea in using FEP in conjunction with IFD-MD is that the correlation coefficient obtained from the FEP calculations for different IFD-MD poses can be used to discriminate between the top few poses, thereby selecting the best structure to use in subsequent prospective FEP simulations. In the present paper, we interrogate only the top two poses; as has been shown above, IFD-MD obtains a low RMSD pose within the top two ranks in ~95% of cases. However, one could examine lower ranked poses as well, which would certainly be suggested if the FEP results for both of the top two poses have poor correlation with the experimental binding data.

Table 1 summarizes the IFD-MD and FEP data for the top two poses for the four receptors enumerated above. This data includes the RMSD of the ligand, and the RMSE and correlation coefficient characterizing the FEP simulation, for the pose in question. It can be seen that the quality of the FEP results effectively discriminate between the top two poses in cases where one of those poses is clearly superior in RMSD. This is shown in greater detail for the CDK2 example in Figure 6. Table 2 summarizes the average performance of IFD-MD across all four systems (using the best poses as selected from the FEP results) and compares this performance with the same averages obtained using co-crystallized structures of one of the congeneric ligands as a starting point (the same complex that is used in ref. [22]). The results are very comparable in terms of both root mean square error (RMSE) and correlation coefficient. Thus, at least for these four systems, deployment of FEP using an IFD-MD starting pose (as would make sense if there were no co-crystal structures of a congeneric ligand) is shown to be a viable option.

Table 1. Performance of Free Energy Calculations Using IFD-MD Determined Binding Modes

	Systems			
	BACE	CDK2	PTP1B	Thrombin
No. of Compounds	36	16	23	11
Binding Affinity Range (kcal/mol)	3.5	4.2	5.1	1.7
Min Tanimoto Similarity Within Series	0.3	0.53	0.22	0.6
Max Tanimoto Similarity of Template Ligand to Series	0.07	0.08	0.05	0.03
Template Complex PDBID	2ZDZ	4BCO	1C87	1C4U
Crystal Structure Series Member	4DJW	1H1Q	2QBS	2ZFF
IFD-MD Pose 1 RMSD (Å)	1.52	0.92	1.63	1.27
IFD-MD Pose 2 RMSD (Å)	0.943	2.81	2.82	2.62
Crystal Structure FEP R^2/$RMSE_{pairwise}$ (kcal/mol)	0.60 / 1.45	0.35 / 1.39	0.87 / 0.69	0.53 / 0.94
IFD-MD Pose 1 FEP R^2/$RMSE_{pairwise}$ (kcal/mol)	0.47 / 0.99	0.60 / 1.09	0.61/ 1.16	0.36 / 1.34
IFD-MD Pose 2 FEP R^2/$RMSE_{pairwise}$ (kcal/mol)	0.17 / 1.11	0.04 / 1.71	0.31 / 1.57	0.04 / 1.17

R^2 is the coefficient of determination between experimental ΔG and predicted ΔG. $RMSE_{pairwise}$ is the root-mean-squared-error between experimental ΔG and predicted ΔG for all ligand pairs. RMSD refers to the ligand heavy-atom RMSD between the IFD-MD aligned member of the series for which there is a published crystal structure, listed under Crystal Structure Series Member.

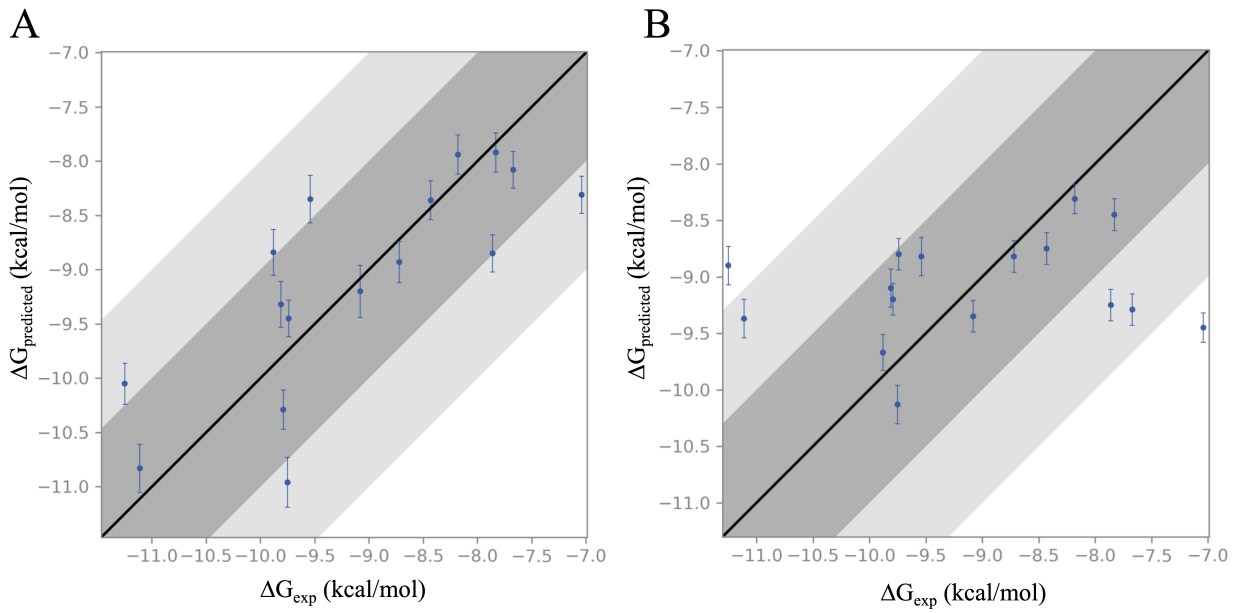

Figure 6. Plot of predicted ΔG versus experimental ΔG for congeneric ligands using the IFD-MD predictions for CDK2. In prospective application, where the ligand RMSD is unknown, correlation plots such as these allow for the discrimination between the correct and incorrect IFD-MD prediction using only retrospective affinity data around for a congeneric series. (A) Correlation plot using IFD-MD Pose 1. This pose is a 0.92 Å ligand-heavy atom RMSD for the member of the series for which a crystal structure is available, PDB ID 1H1Q. The R^2 is 0.60 and the RMSE 1.09 kcal/mol. (B) Correlation plot using IFD-MD Pose 2. This pose is a 2.81 Å ligand-heavy atom RMSD. The R^2 is an inferior 0.04 and the RMSE is 1.71 kcal/mol.

Table 2. Performance of Free Energy Calculations using the Best IFD-MD Determined Binding Mode

System	Crystal Structure R^2 / RMSE (kcal/mol)	Best IFD-MD Pose R^2 / RMSE (kcal/mol)
BACE	0.6 / 1.45	0.49 / 0.99
CDK2	0.35 / 1.39	0.60 / 1.09
PTP1B	0.87 / 0.69	0.61 / 1.16
Thrombin	0.53 / 0.94	0.36 / 1.34
Average	0.59 / 1.12	0.52 / 1.15

The best IFD-MD pose is determined using the R^2 among the poses shown in Table 1.

III.E Application of IFD-MD to Proprietary Drug Discovery Projects

In a recent publication authored by scientists at Merck KGaA, the utility of free energy calculations in prospective drug discovery projects was assessed[23]. In that publication, the authors state that the main reason limiting the use of free energy calculations in their discovery projects is a lack of structural data. In

particular, protein conformational changes, unresolved atoms in an existing crystal structure, or uncertainty in the binding exist as possible challenges. The induced fit docking algorithm we present in this publication is intended to reduce the severity of these problems in active drug discovery projects.

To explore the utility of IFD-MD in active projects we engaged in retrospective and prospective evaluation. For retrospective evaluation, we examine the performance of IFD-MD in predicting the structure of a proprietary ligand using only the publicly available crystal structures that would have been present at the start of the project. Here, we have available for comparison a crystal structure of the proprietary ligand bound to its target receptor.

Table 3. Retrospective Performance of IFD-MD on Five Proprietary Systems

System	Rigid Receptor Docking	IFD	IFD-MD
System 1	1/5	3/5	5/5
System 2	2/14	10/14	14/14
System 3	13/15	13/15	15/15
System 4	5/20	6/20	7/20
System 5	11/20	15/20	20/20
Total	32/74	47/74	61/74

For each of the five systems, the numbers shown indicate the number of cross-docks successful versus the number of cross-docks attempted. A successful cross-dock is one in which the predicted ligand heavy-atom RMSD was 2.5 Å or better. Multiple cross-docks indicate the use of alternative public structures as the template structure for the cross-dock calculation. For example, for System 1, five separate public co-crystal structures were used as a starting structure for the three docking methods listed.

Table 3 lists the performance of IFD-MD for five proprietary systems investigated in the course of Schrödinger's in-house drug discovery efforts. Unlike the construction of prior cross-docks in this paper, no consideration was given here to the presence or absence of backbone motion or to whether the template ligand fully occupied the same binding site as the target ligand.

Compared to both rigid receptor docking and IFD, IFD-MD outperforms in all five systems. In only one system, titled System 4, does IFD-MD fail to successfully model the ligand-receptor complex 100% of the time. Notably, System 4 included cases involving significant backbone motion outside the scope of the present work. We anticipate that further validation of these predictions with free-energy perturbation calculations would mitigate the uncertainty of knowing which predictions were suitable for further modeling.

In prospective evaluation, IFD-MD was used to generate a predicted ligand-receptor complex before the availability of any crystal structure, public or proprietary. In the first case presented, System 6, a potent ligand was discovered whose binding mode was unknown. Congeneric ligand series around this novel ligand was available to validate the IFD-MD model using free energy perturbation calculations.

Figure 7 plots the predicted ΔG against the experimental ΔG using IFD-MD to generate the complex structure for System 6. The plot shown here has an R^2 of 0.52 and a pairwise RMSE of 1.42 kcal/mol. While the IFD-MD and free energy perturbation calculations took a matter of days to complete, weeks later a crystal structure was eventually obtained. The IFD-MD prediction was observed to have an RMSD of 1.24 Å relative to the crystal structure, consistent with its performance when used to support free energy calculation-based optimization of project compounds.

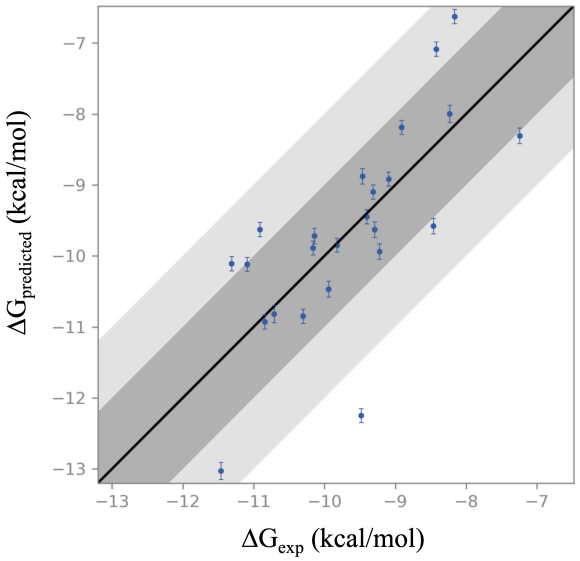

Figure 7. Plot of predicted ΔG versus experimental ΔG for congeneric ligands using the IFD-MD prediction for System 6. The plot shown here has an R^2 of 0.52 and a pairwise RMSE of 1.42 kcal/mol.

In a final prospective evaluation, IFD-MD was used to predict the ligand-receptor complex of a target ligand bound to an identified off-target receptor with known adverse effects. This off-target will be referred to as System 7. Here, IFD-MD and free energy calculations were both used before either a crystal structure or experimental affinity data of the shown compounds was available for the off-target. The experimental binding affinity data was subsequently measured and found to be highly correlated with the predicted values.

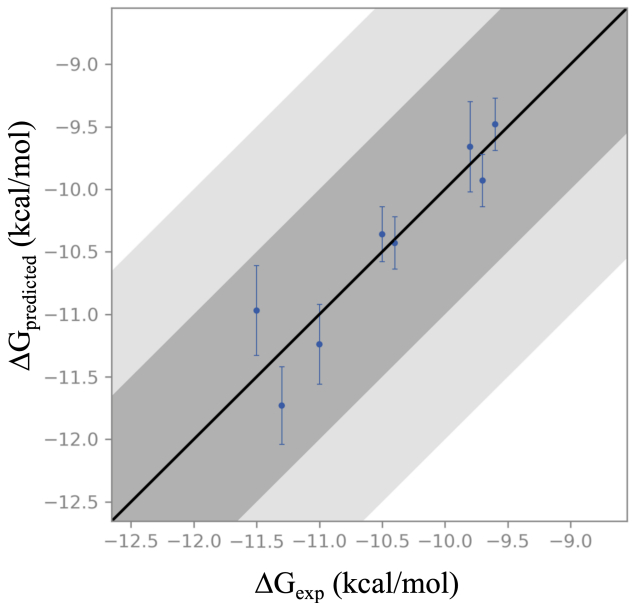

Figure 8. Plot of predicted ΔG versus experimental ΔG for congenic ligands using the IFD-MD prediction for System 7. The plot shown here has an R^2 of 0.87 and a pairwise RMSE of 0.45 kcal/mol.

Figure 8 plots the predicted ΔG against the experimental ΔG using the top-ranked ligand-receptor complex from IFD-MD for System 7. The plot shown here has an R^2 of 0.87 and a pairwise RMSE of 0.45 kcal/mol. Based on the quality of these results, the project team has chosen to not to pursue crystallization of the target ligand bound to this receptor, and instead intends to use the IFD-MD structural model to pursue further optimization of the off-target selectivity of the matter as needed.

IV Conclusion

By combining state of the art computational chemistry tools for sampling and scoring, utilizing empirical, continuum solvent based, and explicit molecular dynamics models, we have developed a robust and accurate methodology, IFD-MD, for induced fit docking of ligands into protein receptors. The methodology has been validated using large training and test sets taken from the PDB, displaying failure rates on the order of 5-7%. It should be noted that many of the systems classified as "failures" could in fact be viewed as marginal successes, for example cases where top scoring poses have RMSDs in the range of 2.5-2.8 Å, or where low RMSD poses are ranked in the 3rd-5th position as opposed to first or second. Furthermore, IFD-MD determined binding modes can yield results that are comparable in quality to an experimentally determined crystal structure, and can support successful application of advance structure-based drug discovery techniques, including free energy calculations. These results hold up not only in an independent test set composed of PDB structures, but also in retrospective and prospective applications arising from Schrödinger's internal and collaborative drug discovery projects.

There are a number of paths forward to systematically improve IFD-MD with regard to accuracy, domain of applicability, and computational performance. Firstly, as larger training and test sets are employed, we expect that the individual components will become better at discriminating correct and incorrect poses, thus increasing the reliability of the overall methodology. Secondly, we intend to introduce loop refinement into the algorithm, enabling cases with backbone clashes to be treated successfully, and further improving scoring effectiveness.

A key question is what sort of impact the improved IFD-MD induced fit docking capabilities can have on structure-based drug discovery projects. At Schrödinger, we are ramping up usage of IFD-MD across our project portfolio. Likewise, we anticipate exciting applications throughout the pharmaceutical industry, including for example rapid structure determination of hits arising from experimental high throughput or DNA-encoded library screening, and using known active molecules to refine the ligand binding sites of homology models.

ASSOCIATED CONTENT

Supporting Information

Detailed methodology as well as performance of individual components of the algorithm and scoring function, additional tables listing the complete composition of the data sets, parameter values for the composite scoring function, and individual free energy perturbation results using an IFD-MD receptor rather than a crystal structure. (DOCX)

Coordinates of Public Retrospective Predictions, PDB Format (ZIP)

AUTHOR INFORMATION

Corresponding Author

*Phone: 212-854-7606. Fax: 212-854-7454. E-mail: raf8@columbia.edu.

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; IFD, Schrödinger's previous IFD software; FEP, free energy perturbation;

REFERENCES

(1) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and Validation of a Genetic Algorithm for Flexible Docking. *J. Mol. Biol.* **1997**, *267* (3), 727–748. https://doi.org/10.1006/JMBI.1996.0897.

(2) 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. https://doi.org/10.1021/jm0306430.

(3) Allen, W. J.; Balius, T. E.; Mukherjee, S.; Brozell, S. R.; Moustakas, D. T.; Lang, P. T.; Case, D. A.; Kuntz, I. D.; Rizzo, R. C. DOCK 6: Impact of New Features and Current Docking Performance. *J. Comput. Chem.* **2015**, *36* (15), 1132–1156. https://doi.org/10.1002/jcc.23905.

(4) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A Fast Flexible Docking Method Using an Incremental Construction Algorithm.

J. Mol. Biol. **1996**, *261* (3), 470–489. https://doi.org/10.1006/JMBI.1996.0477.

(5) 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.

(6) Shan, Y.; Kim, E. T.; Eastwood, M. P.; Dror, R. O.; Seeliger, M. A.; Shaw, D. E. How Does a Drug Molecule Find Its Target Binding Site? *J. Am. Chem. Soc.* **2011**, *133* (24), 9181–9183. https://doi.org/10.1021/ja202726y.

(7) Kufareva, I.; Rueda, M.; Katritch, V.; Stevens, R. C.; Abagyan, R.; Yoshikawa, Y.; Furuya, T.; Lee, H.; Roy, A.; Grime, J.; Rebehmed, J.; Zhang, Y.; Roumen, L.; de Esch, I. J. P.; Leurs, R.; de Graaf, C.; Li, Y.; Hou, T.; Mysinger, M. M.; Weiss, D. R.; Irwin, J. J.; Shoichet, B. K.; McRobb, F. M.; Capuano, B.; Crosby, I. T.; Chalmers, D. K.; Yuriev, E.; Wang, Q.; Mach, R. H.; Reichert, D. E.; Chuang, G. Y.; Rognan, D.; Simms, J.; Sexton, P.; Wootten, D.; Latek, D.; Ghoshdastider, U.; Filipek, S.; LenServer; Kirkpatrick, A.; Trzaskowski, B.; Griffith, A.; Kim, S. K.; Abrol, R.; Goddard, W. A.; Vaidehi, N.; Lam, A.; Bhattacharya, S.; Li, H.; Balaraman, G.; Niesen, M.; Pal, S.; Solovyev, V.; Vorobjev, Y.; Bakulina, N.; Beuming, T.; Costanzi, S.; Shi, L.; Higgs, C.; Salam, N.; Lupyan, D.; Sherman, W.; Ding, F.; Kota, P.; Ramachandran, S.; Dokholyan, N. V.; Carlsson, J.; Coleman, R. G.; Fan, H.; Schlessinger, A.; Irwin, J. J.; Sali, A.; Shoichet, B. K.; Tikhonova, I.; Pogozheva, I.; Lomize, A.; Hall, N. E.; Muddassa, M.; Zhang, Y.; Nim Pae, A.; Lee, J.; Lopez, L.; Obiol-Pardo, C.; Selent, J.; Mahboob, S.; Werner, T.; Bret Church, W.; Brylinski, M.; Ando, T.; Guerler, A.; Zhou, H.; Skolnick, J.; Xhaard, H.; Jurkowski, W.; Elofsson, A.; Murad, A. K.; Drwal, M.; Dupree, T. B.; Griffith, R.; Ostopovici-Halip, L.; Bologa, C.; Chen, K. M.; Sun, J.; Barth, P.; Yarov-Yarovoy, V.; Baker, D.; Vroling, B.; Sanders, M. P. A.; Nabuurs, S. B.; Nikiforovich, G. V. Status of GPCR Modeling and Docking as Reflected by Community-Wide GPCR Dock 2010 Assessment. *Structure* **2011**, *19* (8), 1108–1126. https://doi.org/10.1016/j.str.2011.05.012.

(8) Michino, M.; Abola, E.; Brooks, C. L.; Dixon, J. S.; Moult, J.; Stevens, R. C.; participants, G. D. 2008. Community-Wide Assessment of GPCR Structure Modelling and Ligand Docking: GPCR Dock 2008. *Nat. Rev. Drug Discov.* **2009**, *8* (6), 455–463. https://doi.org/10.1038/nrd2877.

(9) Dixon, S. L.; Smondyrev, A. M.; Knoll, E. H.; Rao, S. N.; Shaw, D. E.; Friesner, R. A. PHASE: A New Engine for Pharmacophore Perception, 3D QSAR Model Development, and 3D Database Screening: 1. Methodology and Preliminary Results. *J. Comput. Aided. Mol. Des.* **2006**, *20* (10), 647–671. https://doi.org/10.1007/s10822-006-9087-6.

(10) Sastry, G. M.; Dixon, S. L.; Sherman, W. Rapid Shape-Based Ligand Alignment and Virtual Screening Method Based on Atom/Feature-Pair Similarities and Volume Overlap Scoring. *J. Chem. Inf. Model.* **2011**, *51* (10), 2455–2466. https://doi.org/10.1021/ci2002704.

(11) Dixon, S. L.; Smondyrev, A. M.; Rao, S. N. PHASE: A Novel Approach to Pharmacophore Modeling and 3D Database Searching. *Chem. Biol. Drug Des.* **2006**, *67* (5), 370–372. https://doi.org/10.1111/j.1747-0285.2006.00384.x.

(12) 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.

(13) Bowers, K. J.; Chow, E.; Xu, H.; Dror, R. O.; Eastwood, M. P.; Gregerson, B. A.; Klepeis, J. L.; Kolossvary, I.; Moraes, M. A.; Sacerdoti, F. D.; Salmon, J. K.; Shan, Y.; Shaw, D. E. Proceedings of the 2006 ACM/IEEE Conference on Supercomputing, SC'06. In *Proceedings of the 2006 ACM/IEEE Conference on Supercomputing, SC'06*; Association for Computing Machinery: New York, NY, USA, 2006.

(14) Schrödinger Release 2020-3: Desmond Molecular Dynamics System. D. E. Shaw Research: New York, NY 2020.

(15) Schrödinger Release 2020-3: Prime. Schrödinger, LLC: New York, NY 2020.

(16) 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.

(17) 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.

(18) 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.

(19) 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.

(20) Weininger, D.; Weininger, A.; Weininger, J. L. SMILES. 2. Algorithm for Generation of Unique SMILES Notation. *J. Chem. Inf. Comput. Sci.* **1989**, *29* (2), 97–101. https://doi.org/10.1021/ci00062a008.

(21) Daylight Chemical Information Systems Inc. *Language for Describing Molecular Patterns*; Aliso Viejo, CA, 2008.

(22) 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.

(23) Schindler, C.; Baumann, H.; Blum, A.; Böse, D.; Buchstaller, H.-P.; Burgdorf, L.; Cappel, D.; Chekler, E.; Czodrowski, P.; Dorsch, D.; Eguida, M.; Follows, B.; Fuchß, T.; Grädler, U.; Gunera, J.; Johnson, T.; Lebrun, C. J.; Karra, S.; Klein, M.; Kötzner, L.; Knehans, T.; Krier, M.; Leiendecker, M.; Leuthner, B.; Li, L.; Mochalkin, I.; Musil, D.; Neagu, C.; Rippmann, F.; Schiemann, K.; Schulz, R.; Steinbrecher, T.; Tanzer, E.-M.; Lopez, A. U.; Follis, A. V.; Wegener, A.; Kuhn, D. Large-Scale Assessment of Binding Free Energy Calculations in Active Drug Discovery Projects. January 7, 2020. https://doi.org/10.26434/chemrxiv.11364884.v1.

Insert Table of Contents artwork here

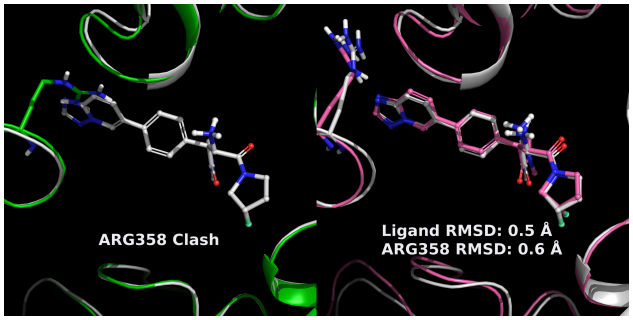

Figure 9. For Table of Contents Only