FEP Protocol Builder: Optimization of Free Energy Perturbation Protocols using Active Learning

César de Oliveira,[b] Karl Leswing[a], Shulu Feng[a], René Kanters[a], Robert Abel [a] and Sathesh Bhat[a]

[a]Schrodinger, Inc., 1540 Broadway 24th Floor New York, NY 10036, United States.

[b]Schrodinger, Inc., 9868 Scranton Road, Suite 3200, San Diego, CA 92121, United States.

Corresponding Author

**E-mail:* sathesh.bhat@schrodinger.com

KEYWORDS: Free Energy Perturbation, Active Learning, Protocol optimization, FEP+

Abstract

Significant improvements have been made in the past decade to methods that rapidly and accurately predict binding affinity through free energy perturbation (FEP) calculations. This has been driven by recent advances in small molecule force fields and sampling algorithms combined with the availability of low-cost parallel computing. Predictive accuracies of ~1 kcal mol^{-1} have been regularly achieved, which are sufficient to drive potency optimization in modern drug discovery campaigns. Despite the robustness of these FEP approaches across multiple target classes, there are invariably target systems that do not display expected performance with default FEP settings. Traditionally, these systems required labor-intensive manual protocol development to arrive at parameter settings that produce a predictive FEP model. Due to the a) relatively large parameter space to be explored, b) significant compute requirements, and c) limited understanding of how combinations of parameters can affect FEP performance, manual FEP protocol optimization can take weeks to months to complete, and often does not involve rigorous train-test set splits, resulting in potential overfitting. These manual FEP protocol development timelines do not coincide with tight drug discovery project timelines, essentially preventing the use of FEP calculations for these target systems. Here, we describe an automated workflow termed FEP Protocol Builder (FEP-PB) to rapidly generate accurate FEP protocols for systems that do not perform well with default settings. FEP-PB uses active learning to iteratively search the protocol parameter space to develop accurate FEP protocols. To validate this approach, we applied it to pharmaceutically relevant systems where default FEP settings could not produce predictive models. We demonstrate that FEP-PB can rapidly generate accurate FEP protocols for the previously challenging MCL1 system with limited human intervention. We also apply FEP-PB in a real-world drug discovery setting to generate an accurate FEP protocol for the p97 system. FEP-PB is able to generate a more accurate protocol

than the expert user, rapidly validating p97 as amenable to free energy calculations. Additionally, through the active learning process, we are able to gain insight into which parameters are most important for a given system. These results suggest that FEP-PB is a robust tool that can aid in rapidly developing accurate FEP protocols and increasing the number of targets that are amenable to the technology.

Introduction

One of the primary goals in a drug discovery campaign is improving ligand binding affinity towards a given on-target, while simultaneously reducing binding affinity towards off-targets.[1,2] Computational methods that can accurately predict binding affinity can significantly accelerate drug discovery projects by a) reducing the number of compounds that have to be synthesized and assayed, and b) enabling the exploration of vast chemical space *in silico* to rapidly discover key molecules that advance the project.[3,4] Free energy perturbation (FEP) calculations provide a rigorous, accurate approach to predicting binding free energies.[5,6] FEP has recently become more accessible due to simultaneous advances in small molecule force fields, molecular dynamics sampling methods, and GPU cloud computing.[7–11]

A Free Energy Perturbation (FEP) protocol is defined as the input receptor and ligand structures alongside the settings of various simulation parameters involved in a free energy calculation (Table 1). To successfully deploy FEP technology on a drug discovery project, the primary requirement is to have a robust FEP protocol that can predict the experimental binding affinity with sufficient accuracy. In practice, we and others have found that acceptable accuracy for prospective application is a root-mean-square error (RMSE) less than 1.3 kcal/mol along with no evidence of convergence problems.[12] With recent improvements to FEP calculations, a large number of systems routinely produce accurate results with the default FEP protocol.[7,13,14] Additionally, improvements in usability and automation.[7,15,16] have resulted in increases in the adoption of large-scale FEP calculations in drug discovery settings.[13,17,18] For example, we have previously demonstrated how an accurate FEP protocol can rapidly accelerate a drug discovery project through the synthesis of fewer compounds by exploring billions of molecules *in silico.* [19,20] Despite these compelling results, there are still a significant number of structurally enabled

targets for which FEP is currently not amenable, oftentimes due to the lack of an accurate FEP protocol. In cases where the default FEP protocol cannot produce an accurate model,[21] it requires the intervention of expert users to develop a working FEP protocol, which is often highly system-dependent and labor-intensive. There are numerous examples in the literature where significant human expert intervention was required to generate an accurate FEP protocol, and simple changes such as longer simulation timescales were insufficient to improve predictive performance.[13,22] For instance, recent work found that certain water models work well for FEP calculations in polar sites of CYP450, while other water models work better for hydrophobic sites.[23] In terms of adequate sampling. Aldeghi et al. showed that equilibration times longer than 1 ns were necessary to produce accurate FEP predictions for the BRD4 system.[24] Enhanced sampling methods, such as replica exchange with solute tempering (REST),[25] can also be applied to further improve convergence and accuracy of FEP calculations. In multiple studies, Wang et al. showed that the inclusion of select binding site residues in the REST region improves the accuracy of FEP predictions for the thrombin and CDK2 system.[8,26] Cole et al. reported similar findings for the HIV reverse transcriptase (HIVRT).[27] Moreover, Sirimulla et al. found that FEP performance was improved by including the entire ligand in the REST region for several systems.[28] Concerning receptor structures, Schindler et al. reported that the selection of loop conformation for the kinesin Eg5 protein had a significant effect on the FEP accuracy.[13] Similarly, Steinbrecher et al. findings strongly suggested that the selection of protein receptor structures can significantly affect FEP model performance for the JAK2, HSP90, and p38alpha MAP kinase systems.[29] Matricon et al. reported that manual selection of protonation states of multiple histidine residues, alternative conformations of His264 and Glu169, and the TIP3P water model resulted in good retrospective FEP performance for the A2A adenosine receptor binding site.[30] Accurate description of water molecules interactions and dynamics around

receptor binding sites are also of key importance for the reliable prediction of protein-ligand binding affinities. As shown by Bos et al. the use of Grand Canonical Monte Carlo sampling (GCMC)[31,32] to improve water sampling for the solvent occluded binding site of D-Amino-Oxidase was important to produce high-quality FEP prospective predictions.[19] Arriving at these aforementioned optimized FEP protocols usually involves several iterations of manually generating FEP protocols, running the FEP calculations, analyzing the results, and refining subsequent FEP protocols. This process is generally labor and time intensive and usually only explores a small portion of the parameter space, oftentimes guided by intuition versus a more systematic data-driven approach. This relatively long timeframe is often not in line with tight drug-discovery timelines, where design-make-test-analyze (DMTA) cycles are usually on the order of two weeks.[13] In a recent large-scale validation of FEP in real-world drug discovery settings, Schindler et al. acknowledge that they expect prospective accuracy to improve if the FEP protocols were further optimized, but manual optimization would take far longer than project timelines.[13] To address these limitations, we developed FEP Protocol Builder (FEP-PB), which is able to significantly accelerate FEP protocol generation via 1) fully automated generation, testing, and validation of FEP protocols using active learning.[33,34] 2) incorporation of training/test set splits to avoid overfitting and 3) employing a data-driven workflow to rapidly converge on accurate FEP protocols that not only have good RMSE values but also good convergence properties. As a real-world test case, we use the Myeloid cell leukemia sequence 1 (MCL1) system, where default FEP settings were unable to produce a protocol with suitable accuracy. FEP-PB is able to rapidly generate accurate protocols with RMSE < 1.3 kcal/mol with good convergence properties. We also apply FEP-PB to real-world drug discovery settings in the validation of the protein p97 system where default settings were unable to generate an accurate protocol. For this system, we also compare the performance of FEP-PB against the

expert user. We find that FEP-PB is able to generate a more accurate FEP protocol than the expert user and rapidly validate p97 as a target that is amenable to free energy calculations

METHODS

System Preparation

The starting receptor structures for FEP-PB were the PDB structure with ID 4HW3[35] for MCL1 and a proprietary Cryo-EM structure for p97, similar to the publicly available Cryo-EM PDB structure with ID 7RLI. The p97 structure contained an allosteric ligand bound at a site distal from where orthosteric ligands bind. It was not clear if the presence of this allosteric ligand would improve FEP performance so two receptor structures were prepared, one with the allosteric ligand present and one with the allosteric ligand absent. The OPLS3e force field was used for the proteins and the ligands.[36] The proteins were prepared using the Protein Preparation Wizard during which the initial protonation states were assigned assuming a pH of 7.0. The systems were solvated in a water box with a buffer width of 5 Å for the complex simulations and 10 Å for the solvent simulations. The systems were relaxed and equilibrated using the default Desmond relaxation protocol. The entire system with the solute molecules restrained to their initial positions was first minimized using the Brownie integrator and then simulated at 10 K using an NVT ensemble followed by an NPT ensemble. The system was then simulated at room temperature using the NPT ensemble with the restraints retained. The Bennett acceptance ratio method (BAR) was used to calculate the free energy.[37] Errors were estimated for each free energy calculation using both bootstrapping and the BAR analytical error prediction.

Ligand Datasets

For the MCL1 system, 42 ligands were selected from Friberg *et al.*[35] (Figure S1). For the p97 system, 15 ligands were selected from Zhou *et al.*[38] (Figure S3). Experimental binding affinity data are expressed as Gibbs Free Energy (ΔG) of binding and were obtained from the published IC_{50} or K_i (inhibition constant) values. Ligand structures were processed using LigPrep to enumerate all stereoisomers, and protonation states were assigned using Epik. OPLS3e force field torsion parameters for each ligand were generated using the Force-Field Builder module (FFBuilder).[36,39]

FEP-PB Active Learning Workflow.

Figure 1 provides an overview of the automated FEP Protocol builder workflow. For the first step, the ligand dataset is randomly split into equally sized training and test sets. The train-test split procedure is appropriate to increase the probability of producing robust models and also to avoid potential overfitting and data dredging.[40,41] The training set is used to run the active learning optimization stage of the workflow, while the test set is used to evaluate the performance of the top optimized protocols. Using the input structures and information about the parameter space to explore, the workflow generates 50 (or a user-defined number) random FEP protocols. The FEP protocols are generated and stored using the javascript object notation (JSON) file format.[42] The protocols are randomly generated to help ensure the parameter combinations that are sampled are representative of the entire distribution of possible protocols. Each random protocol is defined by assigning a unique value for each one of the parameters displayed in Table 1. Once the random protocols are generated, an FEP simulation calculation is run using each random protocol. The FEP simulation timescale for the active-learning portion

of the workflow is set to 1 ns. This is done primarily to 1) keep compute costs low, and 2) have the workflow complete in a reasonable timeframe in line with real-world drug discovery projects. After each FEP run is completed, edgewise RMSE values comparing predicted free energies of binding with their respective experimental values are collected. The RMSE values along with their respective protocols are then used to train a machine learning (ML) model. The ML model is built with TPOT,[43] a Python automated machine learning tool (AutoML) tool that uses genetic programming to optimize machine learning pipelines. The workflow then executes a random search through the FEP protocol parameter space followed by the scoring of generated FEP protocols using the ML model. The top 50 protocols with the lowest ML predicted RMSE values are selected. Next, the workflow automatically sets up and runs FEP calculations based on the top 50 ML-selected FEP protocols. In this work, we repeated these steps two times. After the final iteration, the ten top FEP protocols are selected based on the lowest RMSE computed from the 1 ns FEP calculations. To ensure the lowest RMSE scoring protocols also have good convergence properties, we extended the simulation time of the top 10 protocols to 20 ns. Finally, the top extended FEP protocols with the lowest RMSE and good convergence properties are run against the test set. The FEP protocol with the best performance on the training set and acceptable performance against the test set (RMSE < 1.3 kcal/mol) is selected as the overall top-performing protocol.

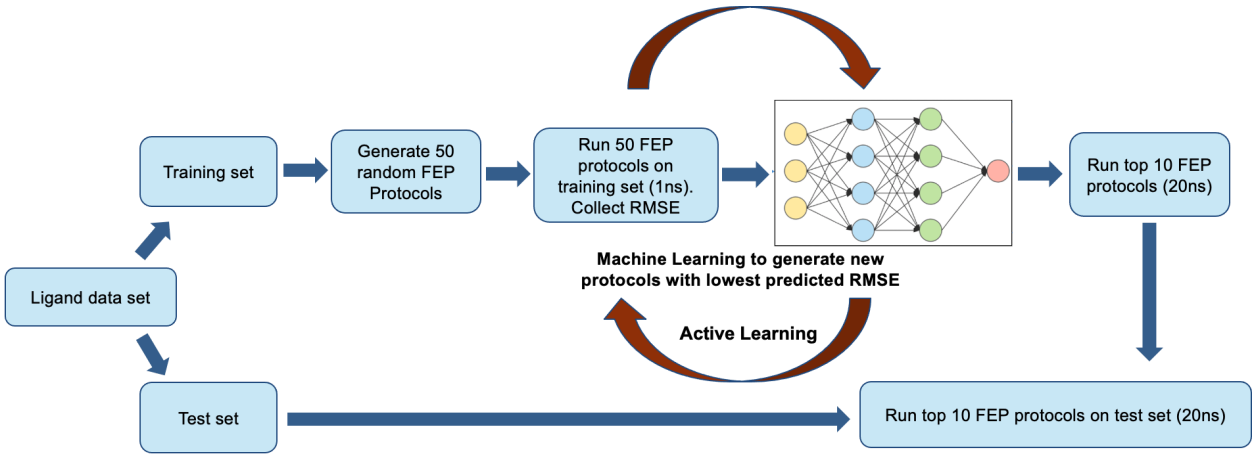

Figure 1. FEP-PB active learning workflow.

FEP+ Parameters for Active Learning Optimization

During protocol optimization, FEP-PB is designed to take into account both FEP simulation parameters and receptor structural modifications, such as multiple receptor structures, residue orientations, and protonation states. Table 1 outlines the FEP parameters explored during the active learning optimization process. Below is a short description of each parameter and how it relates to the overall FEP protocol accuracy.

Receptor structure. The choice of the initial ligand-protein structure can be of key importance to the FEP protocol accuracy, especially for more dynamic systems.[44] Given an ensemble of structures (which can be derived experimentally or from *in-silico* approaches), it is not often straightforward to know which structure will give the optimal FEP performance, especially for systems that display a high degree of conformational flexibility. Therefore, FEP-PB accommodates one or more protein-ligand complex structure files containing a specific

conformation of the protein target and a set of congeneric ligands bound to the receptor binding site. The ligand binding poses can be varied for each specific receptor structure. This allows for active learning to select the optimal receptor structure and ligand binding modes that will produce the most accurate FEP protocol.

Custom core. In most cases, FEP calculations involve the perturbation of a peripheral functional group or the central part (core) of the ligand. In the FEP+ implementation, perturbation pathways are automatically generated using a mapping algorithm as described by Wang et al.[7] Compounds with high FEP similarity scores are connected by edges, which represent alchemical transformations between connected ligands. FEP similarity score for a pair of ligands uses a combination of 2D and 3D features such as their maximum common substructure (MCS) and the quality of their three-dimensional alignment in the receptor binding pocket.[7] The combination of these features determines which atoms of the ligands are mutated during the alchemical transformation. The *custom core* feature in FEP+ enables control of which atoms are included in the mutation region of the molecule by specifying the core of the ligand. The *core,* in this context, corresponds to the ligand atoms that are not perturbed along the FEP calculation. Multiple custom cores can be supplied in the SMARTS format,[45] alongside using the default assignment.

Equilibration timescales. The equilibration phase is generally used to bring the system to the desired temperature and pressure conditions. It is difficult to ascertain *a priori* what the appropriate equilibration time is for a given system, often resulting in either equilibration times that are unnecessarily long, wasting compute resources, or too short, which results in simulations that do not accurately represent the system of interest.[46] Therefore, we allow active

learning to select from a variety of timescales for the equilibration phase of the free energy calculations.

Water models. Due to the impact of water thermodynamics on small molecule binding,[47] the application of different water models in FEP calculations can significantly impact the accuracy of the results. The workflow allows for the inclusion of several popular water models.[48–51]

Water Sampling. Accurate modeling of water molecules plays a key role in the accuracy of free energy calculations. Often the most important water molecules are those that occupy the binding site, particularly those which are kinetically trapped and cannot exchange with bulk solvent within the relatively short molecular simulation timescales.[52–54] Grand Canonical Monte Carlo (GCMC) sampling allows for more accurate water sampling in buried pockets.[31,32,55] We allow active learning to turn this approach on or off.

Lambda schemes. During FEP calculations, the relative free energy between a pair of ligands is estimated by performing a series of molecular dynamics simulations at different coupling parameter (lambda) values.[56] This coupling parameter is a crucial part of the FEP approach as it allows the gradual transformation of one molecule into another. Moreover, the lambda parameter plays a critical role in FEP convergence and accuracy. In this work, a lambda scheme refers to a set of values that are applied to specific types of alchemical transformation. For instance, the “12, 16, 24” lambda scheme refers to r-group, core-hoping, and charge perturbations respectively.

Ligand and Protein REST. The replica exchange with solute tempering (REST) method uses a temperature-dependent scaling of the Hamiltonian, which allows one to effectively heat the molecule, fragment, or protein residue of interest while the remainder of the system remains "cold". In this way, the number of replicas required depends only on a small subset of the total system degrees of freedom. It has been shown that the selection of protein residues to place in the REST region can significantly improve or degrade performance.[25–27,57] Since this is highly system dependent, we allow the active learning to select up to three binding site residues to place in the REST region. The binding site is defined as any residues within 5 Å of a representative ligand. For the ligand REST, this approach facilitates sampling of ligand binding modes that are separated by high free energy barriers and ensures that computed free energy changes are considerably less dependent on the starting conditions and the chosen mutation pathway. We allow the active learning workflow to select from three ligand REST settings 1) Default: For a perturbation between a pair of ligands, the region of the ligand directly involved in the mutation is included in the REST region. If the number of heavy atoms in the perturbed functional group is less than the cutoff value (25 heavy atoms), then the algorithm tries to include one more rotatable bond until the total number of heavy atoms exceeds the cutoff value;[7] 2) Complex: default, Solvent: all. The entire ligand is placed in the REST region during the solvent leg of the simulation while the default setting is applied in the complex leg, and 3) Complex: all, Solvent: all. The entire ligand is placed in the REST region during the solvent and complex leg of the simulation.

Ligand Enhanced Sampling. In typical molecular dynamics simulations, the ligand conformational space is sampled by rotating internal degrees of freedom. Unfortunately, the high energy barriers of certain torsion angles can cause a limited sampling of certain conformations,

which can result in subsequent inaccuracies in the free energy calculations. Ligand torsion scaling addresses this problem by selectively scaling energy barriers on the aforementioned torsion angles, allowing better sampling of conformational space. The use of torsion scaling is highly dependent on factors such as the conformational flexibility of the ligand as well as the flexibility of the protein binding site. We allow active learning to turn this option on or off.

Reference Ligands. Reference ligands in free energy calculations are ligands that are used as reference points to compute the free energy difference between two ligands or a ligand-protein complex. Reference ligands are useful in accounting for systematic errors as well as normalizing the predicted results. From large sets of congeneric ligands, it is often unclear which molecules to select as reference ligands, and this can have significant effects on the accuracy of the predictions.[16] Therefore, we allow active learning to select the optimal pair of reference ligands from a larger set of ligands provided by the user.

Ligand/Protein Force Field. Accurate free energy predictions are highly dependent on the accuracy of the potential energy function. The potential energy of the system of interest is calculated using a molecular mechanics force field, which is an approximation of the quantum mechanical forces. The level and implementation of these approximations can vary between different force field implementations, and it is not always clear which implementation is the most accurate *a priori.*[36,58] We allow the active-learning workflow to select between two implementations of the protein force field, OPLS3e, and OPLS-AA-M. The OPLS-AA-M incorporates new dihedral parameters into the OPLS-AA force field.[58] We also allow the FEP protocol builder workflow to evaluate different versions of the small molecule force fields, the default and advanced versions of the small-molecule forcefield. The default small-molecule

force field consists of the parameters included in the official release of OPLS3e.[36] Since the coverage of OPLS3e is not complete, we observed that for some ligands used in this work, there was at least one missing torsion parameter. For the default mode, we deploy FFbuilder[39] to provide force field parameters only for torsions that were not represented explicitly or by atom-type equivalency in the OPLS3e force field. Alternatively, for the advanced mode, we deploy FFbuilder to recalculate force-field parameters for all torsions of the molecule, regardless if the torsion was represented or not in the OPLS3e force field.

HIS tautomer and protonation states. The imidazole side chain of histidine can exist in different tautomeric forms which have different chemical and physical properties. The correct assignment of these forms can significantly impact the accuracy of free energy calculations. In general, x-ray crystallography cannot distinguish between different tautomeric forms of histidine within a protein structure. Therefore, we allow active learning to sample structures with the histidine proton at the epsilon position (HIE), the delta position (HID), and the double protonated tautomer (HIP) where both nitrogen atoms of the imidazole ring are protonated.

GLN and ASN orientation. The side-chain amide oxygen and nitrogen atoms of asparagine and glutamine residues are indistinguishable at standard crystallographic resolution, and can routinely be assigned incorrectly.[16] Incorrect orientations of asparagine or glutamine residues have little effect on the measured quality of a crystal structure, but can significantly destabilize a protein during a molecular dynamics simulation. Additionally, the energetic barrier to correctly rotate these groups might be too high to achieve during the simulation. Therefore, we sample different starting orientations of these groups in the FEP protocol generation.

Assay Top and Bottom. Oftentimes, significant amounts of experimental binding data are reported at the highest (top) or lowest (bottom) concentration of the ligand in the experimental assay. This data is traditionally excluded from the validation of free energy calculations since the exact experimental value is outside the detection limit of the experimental assay. Here we use a correction function such that if the FEP predicted value is greater than the bottom of the assay or lower than the top of the assay, it is considered a perfect prediction (Equation 1). This allows for active learning to build an accurate model based on both quantitative and qualitative data. This is particularly useful for generating models that accurately predict compounds that are very weak binders or inactive since one of the main purposes of FEP calculations is to enrich for more potent compounds. This parameter was not applied to the systems in this study since all the available experimental data for both systems was quantitative.

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{QUALIFIEDERROR(x_i, \hat{x_i})}{n}}$$

$$QUALIFIEDERROR = \begin{cases} 0 & \text{if } x \leq \hat{x} \text{ and } \hat{x} = assay_{low} \\ 0 & \text{if } x \geq \hat{x} \text{ and } \hat{x} = assay_{high} \\ (x_i - \hat{x_i})^2 & \text{otherwise} \end{cases}$$

Equation 1. Correction function for experimental data at the top or bottom of the assay detection limit.

Parameters used in the Active Learning Optimization	
FEP+	**Input values**
Receptor structure(s)	Single or multiple protein receptors
Ligand training/test set	Label ligand membership in training or test set
Custom Core	SMARTS Pattern(s)
Production timescale (ns)	1
Equilibration timescale (ns)	0.5, 1.0, 5.0
Water models	SPC, TIP4P, TIP4PEW, TIP5P
Water Sampling	GCMC ON or OFF
Lambda schemes	(12, 16, 24) or (16, 20, 28)
Ligand REST	Default, Solvent, Complex schemes
Protein REST	Randomly select 0, 1, 2 or 3 active residues
Ligand Enhanced Sampling	Torsion scaling ON or OFF
Reference Ligands	Any two ligands in dataset
Protein Force Field	OPLS3e or OPLS-AA-MM
Ligand Force Field	FFBuilder Default or Advanced mode
HIS tautomer/protonation states	HID, HIE, HIP

GLN and ASN orientation	Flipped orientation ON or OFF
Assay Top and Bottom	Upper and lower detection limit of experimental binding assay

Table 1. Active Learning optimization parameters

Feature Importance of the FEP Protocols

A benefit of running active learning over the FEP protocol space is that the individual protocols can then be run through a feature importance technique. This can give a deeper understanding of the relative importance of different parameters towards the overall accuracy of the FEP protocol. A popular technique for determining local and global feature contributions are Shapley values.[59] In this framework, one calculates the average marginal contribution of a feature across all possible feature coalitions. These marginal contributions on a sample-by-sample basis sum up to the difference between a sample's model prediction and the average prediction over the training set. While these Shapley feature contributions are on a sample basis, one can then average them across all samples to understand a feature's importance over a whole dataset.

Results and Discussions

MCL1

MCL1 is involved in the regulation of apoptosis and is considered a pro-survival protein. Orthosteric inhibitors of this target are a viable therapeutic strategy for various cancers (Figure 2).[60] This target has historically proven challenging for developing an accurate free-energy model.[7] This is reflected in Figure 5a where the default protocol generates an RMSE of 1.44 kcal/mol with several large outliers (RMSE > 2 kcal/mol) which is not suitable for prospective application.

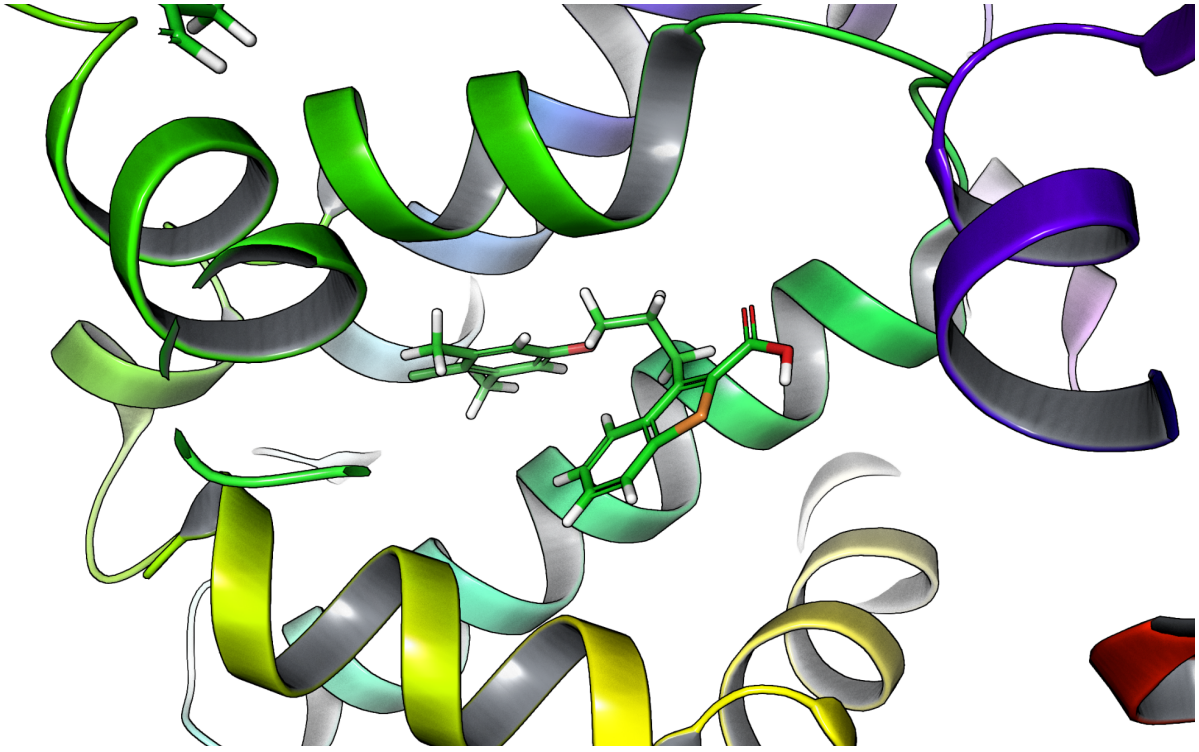

Figure 2. Inhibitor bound to the active site of MCL1 (PDB ID: 4HW3)

Traditionally, the aforementioned poor FEP performance would be followed up with labor-intensive manual debugging to generate a more predictive FEP protocol. Here we instead apply FEP-PB to more rapidly and rigorously generate a predictive model. The settings used for FEP-PB were 50 protocols per active-learning cycle, with three iterations of active learning. Figure 3 displays the RMSE values associated with each of the protocols generated by FEP-PB during the active learning portion of the workflow. Protocols 1 to 50 have a wide range of RMSE values, with some protocols returning RMSE values below 1.3 kcal/mol (colored in blue). A number of FEP protocols also showed poorer predictive power with RMSE above 1.3 kcal/mol (colored in red). This is expected since in the first iteration of active learning, the workflow generates a random set of 50 FEP protocols by randomly selecting parameter values from the defined parameter space (Table 1). Protocols 51 to 100 have a greater number of lower RMSE protocols when compared to the first iteration cycle. This improvement is due to the first round of active learning, which uses the output from the prior iteration to build an ML model for the subsequent iteration. The improved ML model is then able to identify the most relevant features from the FEP parameter space, and as a consequence, generate FEP protocols with higher predictive power. After another round of active learning, the workflow successfully generated many more protocols with RMSE values lower than 1.3 kcal/mol. This suggests that active learning is able to successfully discern which combination of parameters results in the FEP protocols with the lowest RMSE values.

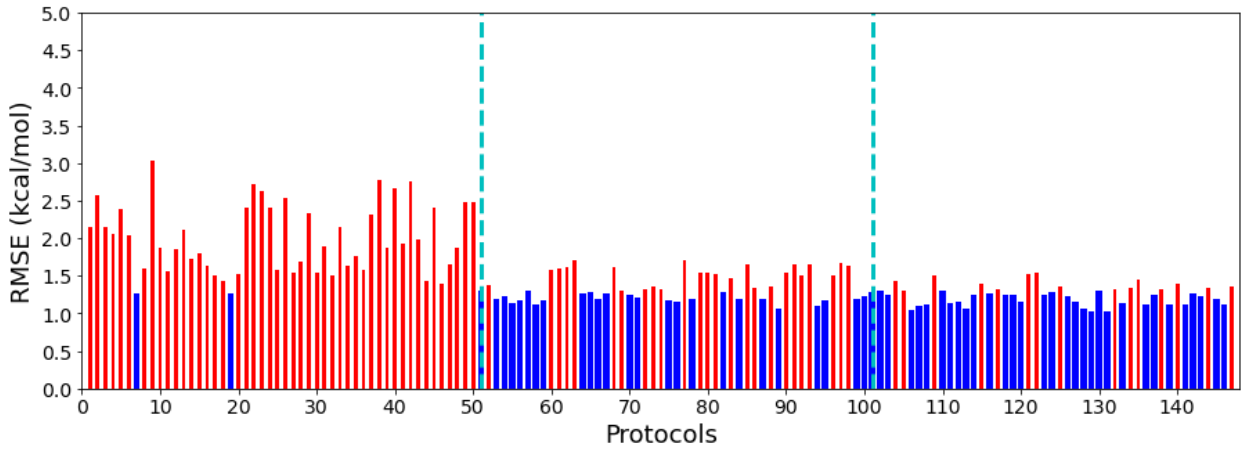

Figure 3. FEP Protocol Builder results for the MCL1 system after 3 iterations of active learning. RMSE values were calculated from predicted FEP and experimental free energy of binding. Blue bars represent protocols with RMSE values below 1.3 kcal/mol, while red bars represent protocols with RMSE values above 1.3 kcal/mol. Vertical cyan dashed lines mark each active learning iteration cycle, which is characterized by 50 FEP protocols.

Figure 4a illustrates the performance of the default FEP protocol on the training set (RMSE 1.62 kcal/mol), which is insufficient for prospective application. Figure 4b displays the performance of the best FEP-PB model on the training set from the three iterations of active learning (RMSE 1.10 kcal/mol), which is significantly better than the default protocol (Table S1). Extending this protocol out to 20 ns maintains good performance (Figure 4c). Moreover, the FEP-PB protocol displays a similar performance on the test set (Figure 4d). This increases confidence in the robustness of the FEP-PB protocol since these ligands were not seen by the active learning optimization.

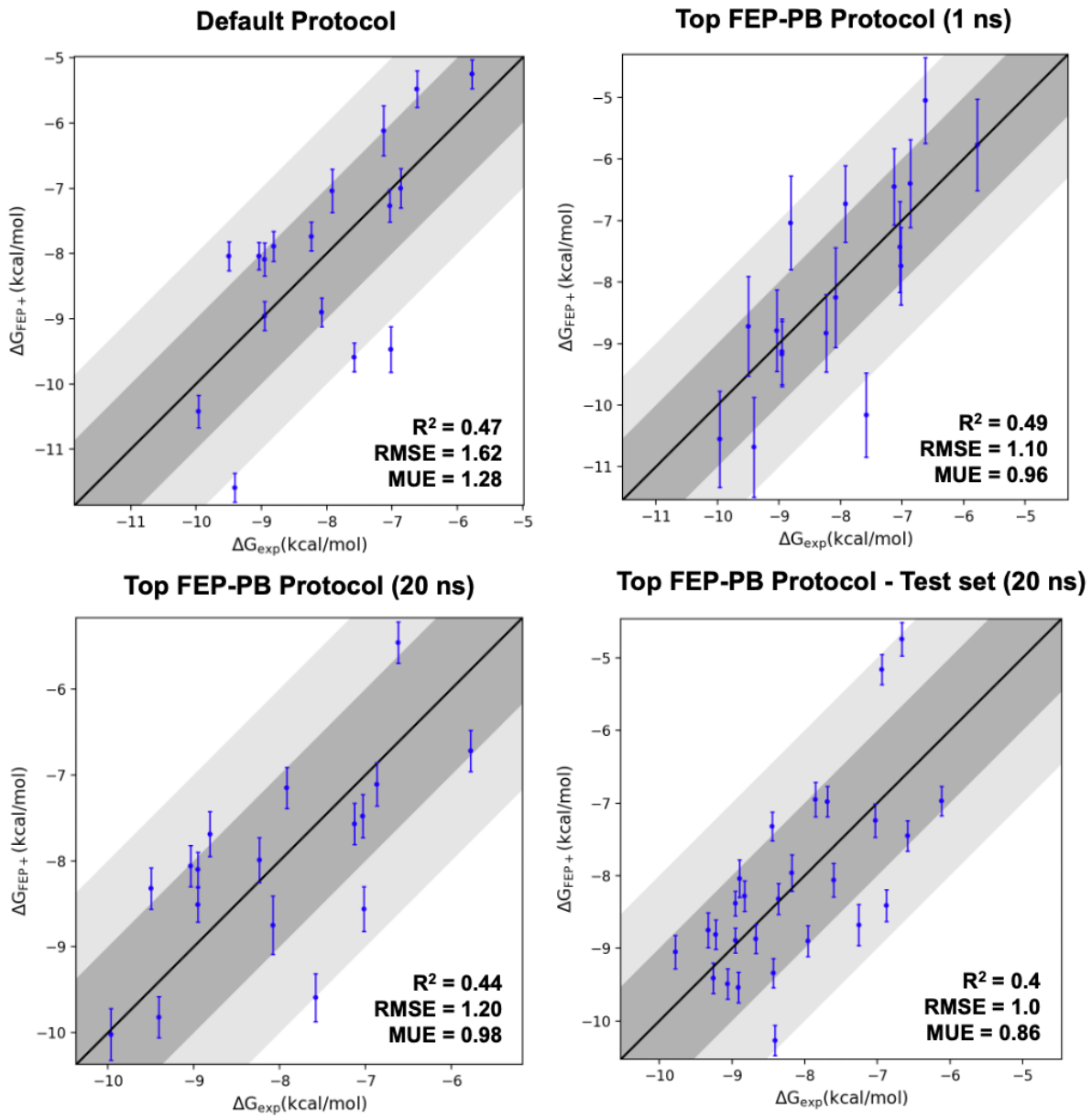

Figure 4. FEP performance of a) Default FEP Protocol on the training set b) Top FEP-PB protocol on the training set (1 ns simulation time) c) Top FEP-PB protocol on the training set (20 ns simulation time) d) Top FEP-PB protocol on the test set (20 ns simulation time). Predictions within 1 kcal/mol and 2 kcal/mol of the experimental affinity are highlighted by dark and light gray areas respectively.

Table 2 outlines the performance of the top FEP-PB protocol on the test set at 1 ns and 20 ns simulation time. We observed a slight increase in RMSE upon extending the protocols from 1 ns to 20 ns, which is expected since many of the protocols at 1 ns are not fully converged. This is exemplified in Table 3, where we observed several convergence warnings for the top protocol at 1 ns, primarily due to the shorter timescale (see supporting information for a detailed explanation of convergence warnings). Upon extension of the top protocol to 20 ns, we see that the majority of these convergence warnings are resolved, suggesting that this model is suitable for prospective use. Similar trends are seen for all the top-performing models for this system (Tables S1, S2).

TimeScale	R2	RMSE (Edgewise)	MUE (Edgewise)	RMSE (Pairwise)	MUE (Pairwise)
1 ns	0.49	1.10	0.96	1.52	1.24
20 ns	0.44	1.20	0.98	1.37	1.12

Table 2. FEP-PB top protocol statistical performance on the test set at 1 ns and 20 ns simulation times

	Number of edges with convergence warnings				
TimeScale	FEP_ERROR	Energy convergence	REST Exchange	Cycle Closure Convergence	Cycles with High Hysteresis
1 ns	20	31	0	13	50
20 ns	0	0	0	3	3

Table 3. Convergence properties of the FEP-PB top protocol on the test set at 1 ns and 20 ns simulation times

Overall performance of the top FEP-PB model for all ligands (i.e. training and test set combined) further demonstrates that the best performing protocol is suitable for prospective application (Figure 5b) and significantly better than the default protocol (Figure 5a). We also observed a significant decrease in the number of large outliers (RMSE > 2 kcal/mol) in the FEP-PB top protocol relative to the default protocol (Figure 5).

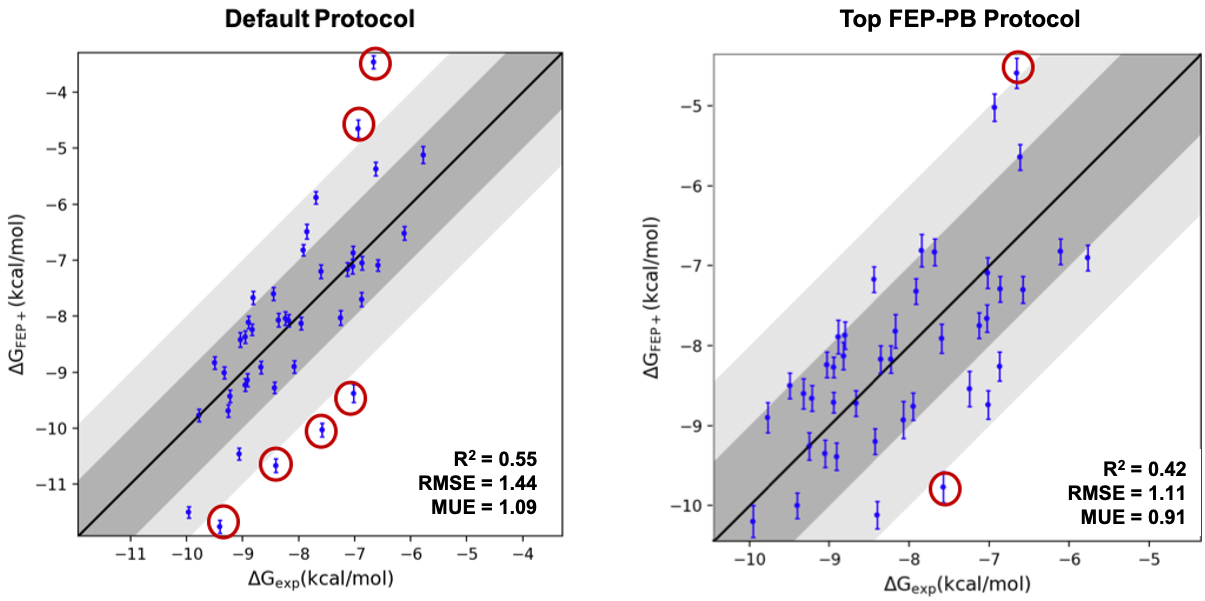

Figure 5: FEP performance for all ligands using a) The default FEP protocol b) The top FEP-PB protocol. Large outliers (RMSE > 2 kcal/mol) are circled in red. Predictions within 1 kcal/mol and 2 kcal/mol of the experimental affinity are highlighted by dark and light gray areas respectively.

Table 4 gives an overview of the parameters of the top FEP-PB protocol. Only parameters that vary from default values are shown. Interestingly, only a relatively small number of parameters have to be varied relative to default settings in order to obtain good performance. Nevertheless,

due to the complexity and interplay between molecular structures, simulations, and energetics, it is often not straightforward to intuit which parameters are important for a system of interest. FEP-PB offers a rapid and rigorous approach to identifying these key parameters and generating accurate FEP protocols.

MCL1 Top FEP-MB Protocol Parameters	
FEP+	**Input values**
Water models	TIP5P
Lambda schemes	(16, 20, 28)
Ligand REST	Entire ligand in REST solvent
Protein REST	Val 253, Leu 235
HIS tautomer/protonation states	HID 224, HID 252
Reference ligands	Ligand 37, ligand 45

Table 4: Parameters for the top FEP-PB protocol for the MCL1 system. All parameters not shown are set to default values.

Feature Importance for the MCL1 System

An additional benefit of running the FEP-PB workflow is scientists can look at trends in the data and gain insight into key FEP parameters of importance in their system of interest. In Figure 6 we show the distributions of the nine largest average Shapley values over the MCL1 dataset when calculated by training an XGBoost[61] model over all the FEP protocols which were run. The most important features are: sampling VAL253 in the REST region, not using reference ligand 64 and instead using reference ligands 26, 34, or 37, using the TIP5P water model, and not using GCMC water sampling.

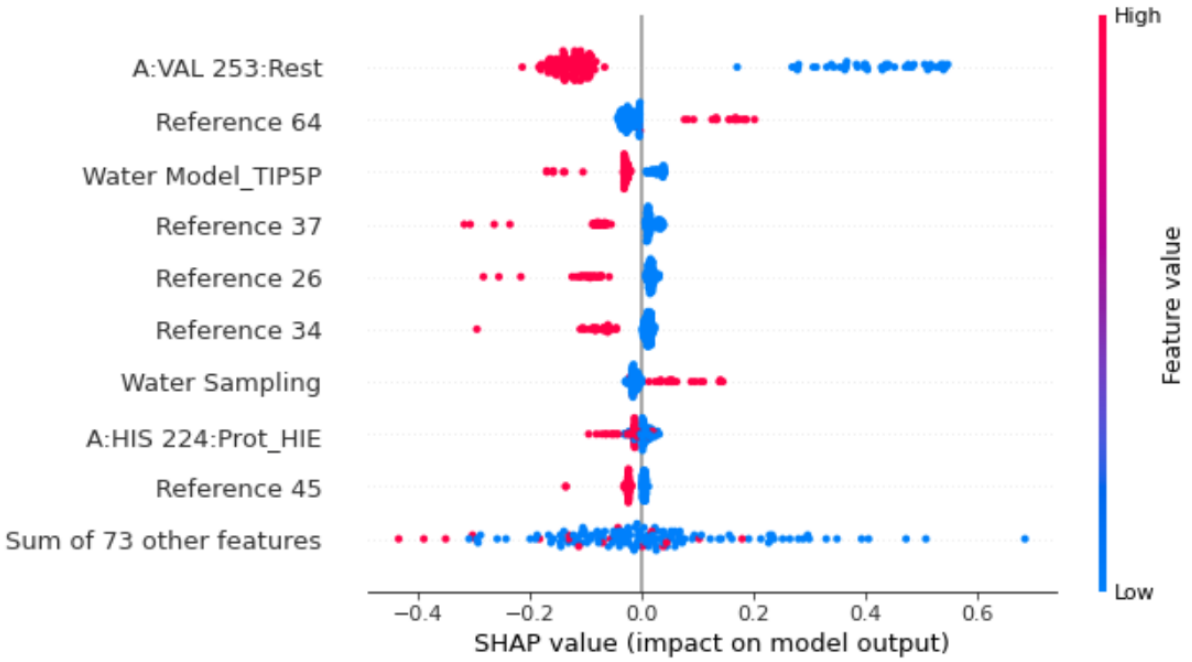

Figure 6: Distributions of the nine largest average absolute Shapley values over the MCL1 dataset.

P97

The p97 protein, also known as valosin-containing protein (VCP) plays a critical role in protein quality control, including degradation and turnover, as well as the removal of misfolded and damaged proteins from the endoplasmic reticulum (ER).[62] Inhibitors of p97 have therapeutic applications in cancer since p97 plays a key role in the degradation of oncoproteins and regulation of cell cycle progression (Figure 7).[63,64] Computationally validating this target first required developing an accurate FEP protocol. The default protocol was applied to the available literature SAR dataset (Figure S3), which resulted in relatively poor performance for prospective utilization (RMSE 1.33 kcal/mol) (Figure 8a). This is a relatively challenging SAR set for FEP since there are simultaneous perturbations to several regions of the molecules (Figure S3). Consequently, we enlisted an FEP expert user to manually generate a more accurate FEP protocol. At the same time, we also deployed FEP-PB to compare both approaches. Only one week was available for both approaches to attempt to generate a predictive FEP protocol. FEP-PB was run in a similar fashion to the MCL1 system and detailed results of the active learning optimization can be found in the supporting information (Figure S4, S5). Figure 8 outlines the performance of the (a) default protocol (b) the expert model, and (c) the top FEP-PB model on the entire p97 dataset. The expert was able to develop a protocol that slightly improved FEP performance (RMSE 1.25 kcal/mol) but still contained several significant outliers with RMSE > 1.3 kcal/mol. FEP-PB produced the most accurate protocol (RMSE 1.08 kcal/mol) and is suitable for prospective use. Interestingly, the receptor structure without the allosteric ligand gives the best FEP performance and is consistently selected by active learning in the top-performing FEP-PB protocols (Figure S4). Comparing the expert and FEP-PB protocols, the expert user was able to discern the importance of extending the lambda scheme and placing the entire ligand in the REST region but selected the suboptimal receptor structure and also did not

modify the equilibration time, protein REST, and ligand force field parameters (Table 5). Modifications to the aforementioned parameters resulted in the most accurate protocol from FEP-PB (Table 6). It is important to emphasize that these results do not reflect a lack of expertise of the human user, but instead, illustrate the intractability of manually exploring the FEP parameter space in a rigorous manner within standard drug discovery timelines. This is in agreement with the work of Schindler et al. who found that the relatively tight drug discovery timelines prevented them from carrying out extensive optimization of FEP protocols for systems where the default FEP protocol performed poorly.[13]

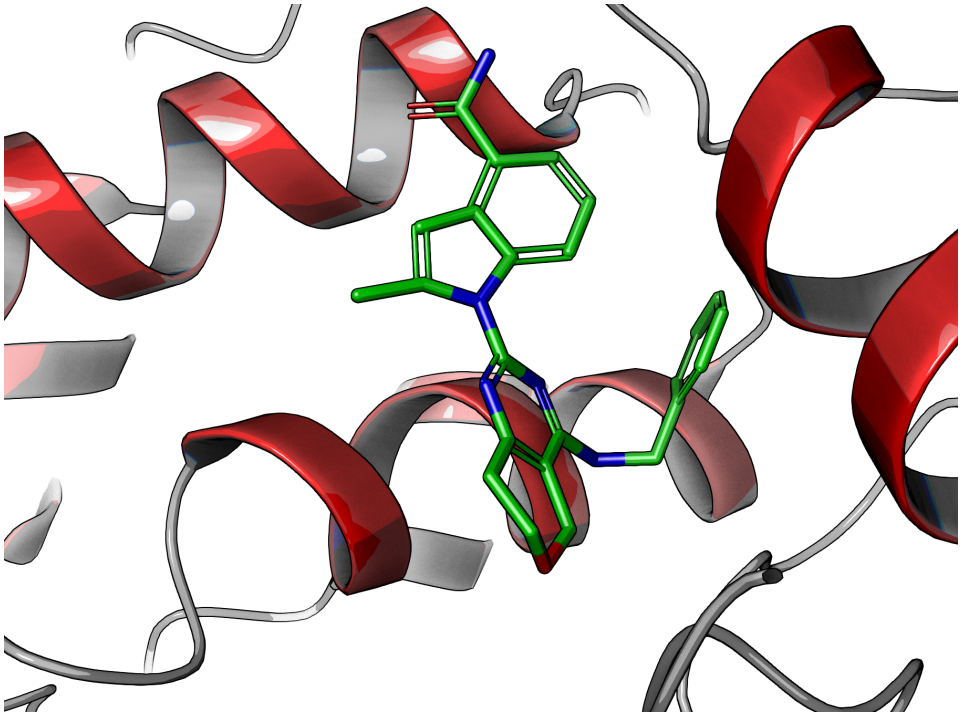

Figure 7. Inhibitor bound to the active site of p97 (PDB ID: 7RLI)

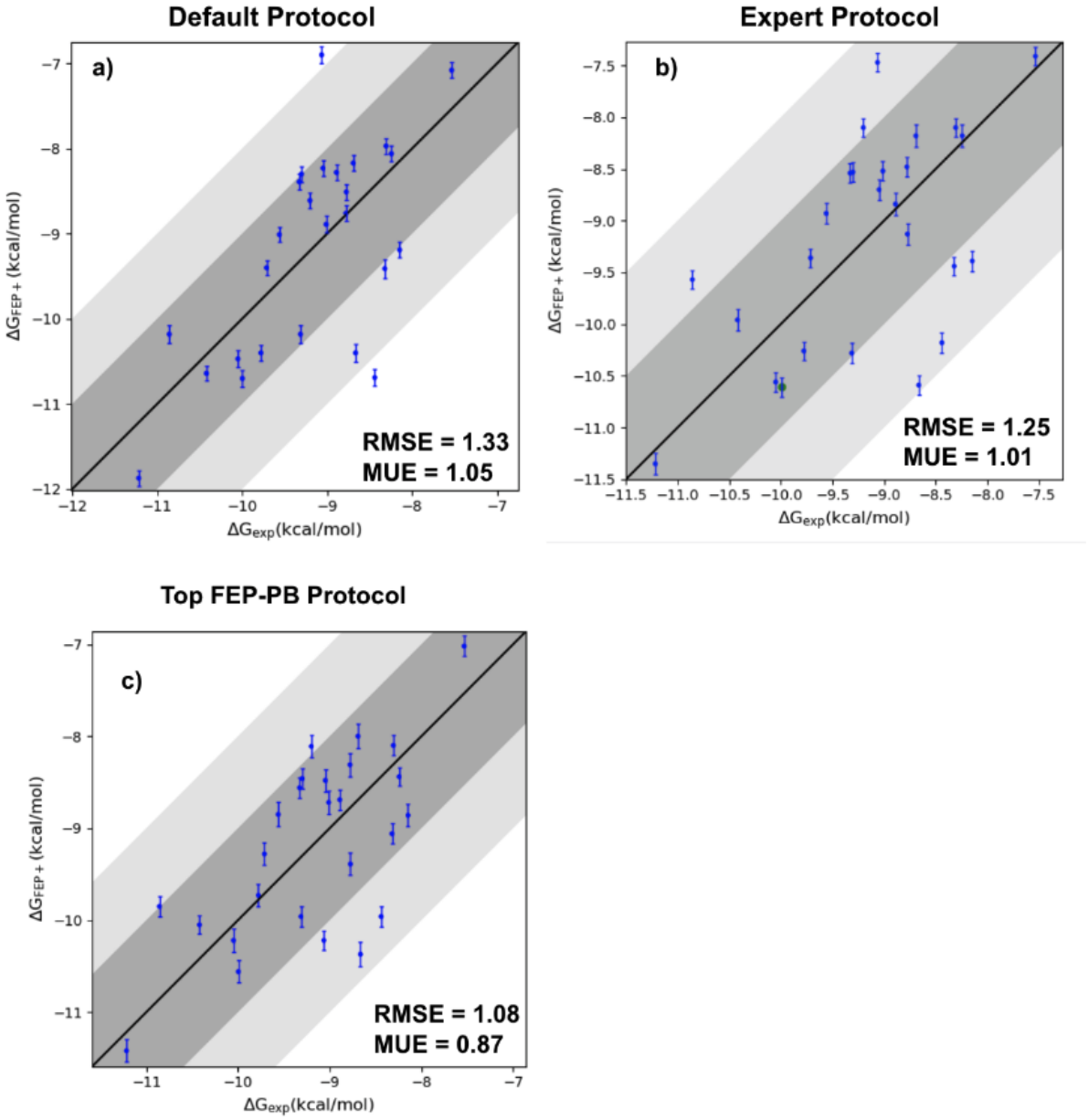

Figure 8. FEP performance of a) Default FEP Protocol on the entire dataset b) Expert protocol on the entire dataset c) Top FEP-PB protocol on the entire dataset. Predictions within 1 kcal/mol and 2 kcal/mol of the experimental affinity are highlighted by dark and light gray areas respectively.

p97 Expert User Protocol Parameters	
FEP+	**Input values**
Receptor	Receptor containing allosteric ligand
Lambda schemes	(20, 20, 24)
Ligand REST	Entire ligand in REST solvent
Reference ligands	Ligand 35, optimal map

Table 5: Parameters for the Expert FEP-PB protocol for the p97 system. All parameters not shown are set to default values.

p97 Top FEP-PB Protocol Parameters	
FEP+	**Input values**
Receptor	Receptor without allosteric ligand
Water model	TIP4PEW
Lambda schemes	(16, 20, 28)
Equilibration time	0.5 ns
Ligand REST	Entire ligand in REST solvent
Protein REST	Arg668, Trp475
Ligand Force Field	Force-Field Builder Advanced mode ON
Reference ligands	Ligand 28, ligand 72

Table 6: Parameters for the Top FEP-PB protocol for the p97 system. All parameters not shown are set to default values.

CONCLUSIONS

In the past few years, we have observed significant and simultaneous improvements in compute power, structural biology, and the predictive accuracy of free energy perturbation calculations.[65,66] To fully take advantage of these improvements, we developed FEP-PB, which employs active learning to rapidly develop accurate FEP protocols for biological targets of interest. We demonstrate the utility of FEP-PB by applying it to the optimization of FEP protocols for a pharmaceutically relevant target, MCL1, where the default settings do not produce suitable model accuracy. FEP-PB is able to generate protocols with good accuracy and convergence that can be applied prospectively. When applied head-to-head against the expert user, FEP-PB is able to generate a more accurate protocol for the p97 system, requiring only ~1 week to complete all calculations using cloud computing resources, which is in line with common drug discovery timelines. Importantly, this workflow is designed to be carried out with limited human intervention, which opens up the possibility of validating many more systems in a much shorter timeline. For example, we can now accelerate the ability to rapidly validate off-target models for modeling ligand selectivity with FEP.[67] We have used FEP-PB extensively in our internal drug discovery projects and found it able to generate low RMSE protocols for several difficult targets. These FEP protocols also perform well when applied prospectively for these targets. With the advent of AlphaFold,[68] we have also regularly been incorporating these protein receptor models into the active-learning workflow, further widening the pool of targets amenable to FEP. Furthermore, we are collecting FEP-PB active learning data from all these projects to better decipher which settings consistently improve FEP model performance across many targets. By calculating the feature importance across these diverse targets, the goal is to eventually arrive at a more robust default protocol. Overall, we expect that FEP-PB will greatly increase the

number of targets that are amenable to free-energy perturbation calculations, allowing rapid acceleration of many more small molecule drug discovery projects.

Data and Software Availability

The software technologies used for this study are available at https://www.schrodinger.com/downloads/releases. Receptor and ligand structures described in the manuscript are available upon request.

Acknowledgments

The authors thank Pieter Bos for useful suggestions.

REFERENCES

(1) Doytchinova, I. Drug Design-Past, Present, Future. *Molecules* **2022**, *27* (5). https://doi.org/10.3390/molecules27051496.

(2) Wang, T.; Pulkkinen, O. I.; Aittokallio, T. Target-Specific Compound Selectivity for Multi-Target Drug Discovery and Repurposing. *Front. Pharmacol.* **2022**, *13*, 1003480.

(3) Cournia, Z.; Allen, B. K.; Beuming, T.; Pearlman, D. A.; Radak, B. K.; Sherman, W. Rigorous Free Energy Simulations in Virtual Screening. *J. Chem. Inf. Model.* **2020**, *60* (9), 4153–4169.

(4) Zhang, B.; D'Erasmo, M. P.; Murelli, R. P.; Gallicchio, E. Free Energy-Based Virtual Screening and Optimization of RNase H Inhibitors of HIV-1 Reverse Transcriptase. *ACS Omega* **2016**, *1* (3), 435–447.

(5) Cournia, Z.; Chipot, C.; Roux, B.; York, D. M.; Sherman, W. Free Energy Methods in Drug Discovery—Introduction. In *Free Energy Methods in Drug Discovery: Current State and Future Directions*; ACS Symposium Series; American Chemical Society, 2021; Vol. 1397, pp 1–38.

(6) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. The Statistical-Thermodynamic Basis for Computation of Binding Affinities: A Critical Review. *Biophys. J.* **1997**, *72* (3), 1047–1069.

(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.

(8) Wang, L.; Berne, B. J.; Friesner, R. A. On Achieving High Accuracy and Reliability in the Calculation of Relative Protein–ligand Binding Affinities. *Proceedings of the National Academy of Sciences* **2012**, *109* (6), 1937–1942.

(9) de Oliveira, C.; Yu, H. S.; Chen, W.; Abel, R.; Wang, L. Rigorous Free Energy Perturbation Approach to Estimating Relative Binding Affinities between Ligands with Multiple Protonation and Tautomeric States. *J. Chem. Theory Comput.* **2019**, *15* (1), 424–435.

(10) Xu, H. The Slow but Steady Rise of Binding Free Energy Calculations in Drug Discovery. *J. Comput. Aided Mol. Des.* **2023**, *37* (2), 67–74.

(11) Muegge, I.; Hu, Y. Recent Advances in Alchemical Binding Free Energy Calculations for Drug Discovery. *ACS Med. Chem. Lett.* **2023**, *14* (3), 244–250.

(12) *Website*. https://www.schrodinger.com/sites/default/files/s3/release/2022-4/Documentation/html/fep/fep_user_manual/fep_best_practices.htm?Highlight=RMSE.

(13) Schindler, C. E. M.; Baumann, H.; Blum, A.; Böse, D.; Buchstaller, H.-P.; Burgdorf, L.; Cappel, D.; Chekler, E.; Czodrowski, P.; Dorsch, D.; Eguida, M. K. I.; Follows, B.; Fuchß, T.; Grädler, U.; Gunera, J.; Johnson, T.; Jorand Lebrun, C.; Karra, S.; Klein, M.; Knehans, T.; Koetzner, L.; 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.; Unzue Lopez, A.; Viacava Follis, A.; Wegener, A.; Kuhn, D. Large-Scale Assessment of Binding Free Energy Calculations in Active Drug Discovery Projects. *J. Chem. Inf. Model.* **2020**, *60* (11), 5457–5474.

(14) Abel, R.; Wang, L.; Harder, E. D.; Berne, B. J.; Friesner, R. A. Advancing Drug Discovery through Enhanced Free Energy Calculations. *Acc. Chem. Res.* **2017**, *50* (7), 1625–1632.
(15) Liu, S.; Wu, Y.; Lin, T.; Abel, R.; Redmann, J. P.; Summa, C. M.; Jaber, V. R.; Lim, N. M.; Mobley, D. L. Lead Optimization Mapper: Automating Free Energy Calculations for Lead Optimization. *J. Comput. Aided Mol. Des.* **2013**, *27* (9), 755–770.
(16) Cournia, Z.; Allen, B.; Sherman, W. Relative Binding Free Energy Calculations in Drug Discovery: Recent Advances and Practical Considerations. *J. Chem. Inf. Model.* **2017**, *57* (12), 2911–2937.
(17) Tresadern, G.; Velter, I.; Trabanco, A. A.; Van den Keybus, F.; Macdonald, G. J.; Somers, M. V. F.; Vanhoof, G.; Leonard, P. M.; Lamers, M. B. A. C.; Van Roosbroeck, Y. E. M.; Buijnsters, P. J. J. A. [1,2,4]Triazolo[1,5-A]pyrimidine Phosphodiesterase 2A Inhibitors: Structure and Free-Energy Perturbation-Guided Exploration. *J. Med. Chem.* **2020**, *63* (21), 12887–12910.
(18) Pérez-Benito, L.; Casajuana-Martin, N.; Jiménez-Rosés, M.; van Vlijmen, H.; Tresadern, G. Predicting Activity Cliffs with Free-Energy Perturbation. *J. Chem. Theory Comput.* **2019**, *15* (3), 1884–1895.
(19) Bos, P. H.; Houang, E. M.; Ranalli, F.; Leffler, A. E.; Boyles, N. A.; Eyrich, V. A.; Luria, Y.; Katz, D.; Tang, H.; Abel, R.; Bhat, S. AutoDesigner, a Design Algorithm for Rapidly Exploring Large Chemical Space for Lead Optimization: Application to the Design and Synthesis of D-Amino Acid Oxidase Inhibitors. *J. Chem. Inf. Model.* **2022**, *62* (8), 1905–1915.
(20) Tang, H.; Jensen, K.; Houang, E.; McRobb, F. M.; Bhat, S.; Svensson, M.; Bochevarov, A.; Day, T.; Dahlgren, M. K.; Bell, J. A.; Frye, L.; Skene, R. J.; Lewis, J. H.; Osborne, J. D.; Tierney, J. P.; Gordon, J. A.; Palomero, M. A.; Gallati, C.; Chapman, R. S. L.; Jones, D. R.; Hirst, K. L.; Sephton, M.; Chauhan, A.; Sharpe, A.; Tardia, P.; Dechaux, E. A.; Taylor, A.; Waddell, R. D.; Valentine, A.; Janssens, H. B.; Aziz, O.; Bloomfield, D. E.; Ladha, S.; Fraser, I. J.; Ellard, J. M. Discovery of a Novel Class of D-Amino Acid Oxidase Inhibitors Using the Schrödinger Computational Platform. *J. Med. Chem.* **2022**, *65* (9), 6775–6802.
(21) Ciordia, M.; Pérez-Benito, L.; Delgado, F.; Trabanco, A. A.; Tresadern, G. Application of Free Energy Perturbation for the Design of BACE1 Inhibitors. *J. Chem. Inf. Model.* **2016**, *56* (9), 1856–1871.
(22) Gapsys, V.; Hahn, D. F.; Tresadern, G.; Mobley, D. L.; Rampp, M.; de Groot, B. L. Pre-Exascale Computing of Protein-Ligand Binding Free Energies with Open Source Software for Drug Design. *J. Chem. Inf. Model.* **2022**, *62* (5), 1172–1177.
(23) Yadav, S.; Kardam, V.; Tripathi, A.; T G, S.; Dubey, K. D. The Performance of Different Water Models on the Structure and Function of Cytochrome P450 Enzymes. *J. Chem. Inf. Model.* **2022**, *62* (24), 6679–6690.
(24) Aldeghi, M.; Heifetz, A.; Bodkin, M. J.; Knapp, S.; Biggin, P. C. Accurate Calculation of the Absolute Free Energy of Binding for Drug Molecules. *Chem. Sci.* **2015**, *7* (1), 207–218.
(25) Wang, L.; Friesner, R. A.; Berne, B. J. Replica Exchange with Solute Scaling: A More Efficient Version Of Replica Exchange with Solute Tempering (REST2). *J. Phys. Chem. B* **2011**, *115* (30), 9431–9438.
(26) Wang, L.; Deng, Y.; Knight, J. L.; Wu, Y.; Kim, B.; Sherman, W.; Shelley, J. C.; Lin, T.; Abel, R. Modeling Local Structural Rearrangements Using FEP/REST: Application To Relative Binding Affinity Predictions of CDK2 Inhibitors. *J. Chem. Theory Comput.* **2013**, *9* (2), 1282–1293.
(27) Cole, D. J.; Tirado-Rives, J.; Jorgensen, W. L. Enhanced Monte Carlo Sampling through Replica Exchange with Solute Tempering. *J. Chem. Theory Comput.* **2014**, *10* (2), 565–571.

(28) Fratev, F.; Sirimulla, S. An Improved Free Energy Perturbation FEP+ Sampling Protocol for Flexible Ligand-Binding Domains. *Sci. Rep.* **2019**, *9* (1), 16829.
(29) Steinbrecher, T. B.; Dahlgren, M.; Cappel, D.; Lin, T.; Wang, L.; Krilov, G.; Abel, R.; Friesner, R.; Sherman, W. Accurate Binding Free Energy Predictions in Fragment Optimization. *J. Chem. Inf. Model.* **2015**, *55* (11), 2411–2420.
(30) Matricon, P.; Ranganathan, A.; Warnick, E.; Gao, Z.-G.; Rudling, A.; Lambertucci, C.; Marucci, G.; Ezzati, A.; Jaiteh, M.; Dal Ben, D.; Jacobson, K. A.; Carlsson, J. Fragment Optimization for GPCRs by Molecular Dynamics Free Energy Calculations: Probing Druggable Subpockets of the A 2A Adenosine Receptor Binding Site. *Sci. Rep.* **2017**, *7* (1), 6398.
(31) Woo, H.-J.; Dinner, A. R.; Roux, B. Grand Canonical Monte Carlo Simulations of Water in Protein Environments. *J. Chem. Phys.* **2004**, *121* (13), 6392–6400.
(32) Ross, G. A.; Russell, E.; Deng, Y.; Lu, C.; Harder, E. D.; Abel, R.; Wang, L. Enhancing Water Sampling in Free Energy Calculations with Grand Canonical Monte Carlo.
(33) Jaques, N.; Gu, S.; Bahdanau, D.; Hernández-Lobato, J. M.; Turner, R. E.; Eck, D. Sequence Tutor: Conservative Fine-Tuning of Sequence Generation Models with KL-Control. **06--11 Aug 2017**, *70*, 1645–1654.
(34) Olivecrona, M.; Blaschke, T.; Engkvist, O.; Chen, H. Molecular de-Novo Design through Deep Reinforcement Learning. *J. Cheminform.* **2017**, *9* (1), 48.
(35) Friberg, A.; Vigil, D.; Zhao, B.; Daniels, R. N.; Burke, J. P.; Garcia-Barrantes, P. M.; Camper, D.; Chauder, B. A.; Lee, T.; Olejniczak, E. T.; Fesik, S. W. Discovery of Potent Myeloid Cell Leukemia 1 (Mcl-1) Inhibitors Using Fragment-Based Methods and Structure-Based Design. *J. Med. Chem.* **2013**, *56* (1), 15–30.
(36) 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.
(37) Bennett, C. H. Efficient Estimation of Free Energy Differences from Monte Carlo Data. *J. Comput. Phys.* **1976**, *22* (2), 245–268.
(38) Zhou, H.-J.; Wang, J.; Yao, B.; Wong, S.; Djakovic, S.; Kumar, B.; Rice, J.; Valle, E.; Soriano, F.; Menon, M.-K.; Madriaga, A.; Kiss von Soly, S.; Kumar, A.; Parlati, F.; Yakes, F. M.; Shawver, L.; Le Moigne, R.; Anderson, D. J.; Rolfe, M.; Wustrow, D. Discovery of a First-in-Class, Potent, Selective, and Orally Bioavailable Inhibitor of the p97 AAA ATPase (CB-5083). *J. Med. Chem.* **2015**, *58* (24), 9480–9497.
(39) Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; Kaus, J. W.; Cerutti, D. S.; Krilov, G.; Jorgensen, W. L.; Abel, R.; Friesner, R. A. OPLS3: A Force Field Providing Broad Coverage Of Drug-like Small Molecules and Proteins. *J. Chem. Theory Comput.* **2016**, *12* (1), 281–296.
(40) Head, M. L.; Holman, L.; Lanfear, R.; Kahn, A. T.; Jennions, M. D. The Extent and Consequences of P-Hacking in Science. *PLoS Biol.* **2015**, *13* (3), e1002106.
(41) Erasmus, A.; Holman, B.; Ioannidis, J. P. A. Data-Dredging Bias. *BMJ Evid Based Med* **2022**, *27* (4), 209–211.
(42) Pezoa, F.; Reutter, J. L.; Suarez, F.; Ugarte, M.; Vrgoč, D. Foundations of JSON Schema. Proceedings *of the 25th International Conference on World Wide Web*; WWW '16; International World Wide Web Conferences Steering Committee: Republic and Canton of Geneva, CHE, 2016; pp 263–273.
(43) Le, T. T.; Fu, W.; Moore, J. H. Scaling Tree-Based Automated Machine Learning to Biomedical Big Data with a Feature Set Selector. *Bioinformatics* **2020**, *36* (1), 250–256.

(44) Shih, A. Y.; Hack, M.; Mirzadegan, T. Impact of Protein Preparation on Resulting Accuracy of FEP Calculations. *J. Chem. Inf. Model.* **2020**, *60* (11), 5287–5289.
(45) *Daylight Theory: SMARTS - A Language for Describing Molecular Patterns*. https://www.daylight.com/dayhtml/doc/theory/theory.smarts.html (accessed 2023-04-18).
(46) Walton, E. B.; Vanvliet, K. J. Equilibration of Experimentally Determined Protein Structures for Molecular Dynamics Simulation. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **2006**, *74* (6 Pt 1), 061901.
(47) Geschwindner, S.; Ulander, J. The Current Impact of Water Thermodynamics for Small-Molecule Drug Discovery. *Expert Opin. Drug Discov.* **2019**, *14* (12), 1221–1225.
(48) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79* (2), 926–935.
(49) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. Interaction Models for Water in Relation to Protein Hydration. In *Intermolecular Forces: Proceedings of the Fourteenth Jerusalem Symposium on Quantum Chemistry and Biochemistry Held in Jerusalem, Israel, April 13–16, 1981*; Pullman, B., Ed.; Springer Netherlands: Dordrecht, 1981; pp 331–342.
(50) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The Missing Term in Effective Pair Potentials. *J. Phys. Chem.* **1987**, *91* (24), 6269–6271.
(51) Nutt, D. R.; Smith, J. C. Molecular Dynamics Simulations of Proteins: Can the Explicit Water Model Be Varied? *J. Chem. Theory Comput.* **2007**, *3* (4), 1550–1560.
(52) Bortolato, A.; Tehan, B. G.; Bodnarchuk, M. S.; Essex, J. W.; Mason, J. S. Water Network Perturbation in Ligand Binding: Adenosine A(2A) Antagonists as a Case Study. *J. Chem. Inf. Model.* **2013**, *53* (7), 1700–1713.
(53) Zia, S. R.; Gaspari, R.; Decherchi, S.; Rocchia, W. Probing Hydration Patterns in Class-A GPCRs via Biased MD: The A2A Receptor. *J. Chem. Theory Comput.* **2016**, *12* (12), 6049–6061.
(54) Huang, D.; Rossini, E.; Steiner, S.; Caflisch, A. Structured Water Molecules in the Binding Site of Bromodomains Can Be Displaced by Cosolvent. *ChemMedChem* **2014**, *9* (3), 573–579.
(55) Clark, M.; Guarnieri, F.; Shkurko, I.; Wiseman, J. Grand Canonical Monte Carlo Simulation of Ligand-Protein Binding. *J. Chem. Inf. Model.* **2006**, *46* (1), 231–242.
(56) Kollman, P. Free Energy Calculations: Applications to Chemical and Biochemical Phenomena. *Chem. Rev.* **1993**, *93* (7), 2395–2417.
(57) Wang, L.; Chambers, J.; Abel, R. Protein–Ligand Binding Free Energy Calculations with FEP+. In *Biomolecular Simulations: Methods and Protocols*; Bonomi, M., Camilloni, C., Eds.; Methods in Molecular Biology; Springer: New York, NY, 2019; pp 201–232.
(58) Robertson, M. J.; Tirado-Rives, J.; Jorgensen, W. L. Improved Peptide and Protein Torsional Energetics with the OPLS-AA Force Field. *J. Chem. Theory Comput.* **2015**, *11* (7), 3499–3509.
(59) Lundberg, S.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. *arXiv [cs.AI]*, 2017. https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf (accessed 2023-04-17).
(60) Bolomsky, A.; Vogler, M.; Köse, M. C.; Heckman, C. A.; Ehx, G.; Ludwig, H.; Caers, J. MCL-1 Inhibitors, Fast-Lane Development of a New Class of Anti-Cancer Agents. *J. Hematol. Oncol.* **2020**, *13* (1), 173.
(61) Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the*

22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; KDD '16; Association for Computing Machinery: New York, NY, USA, 2016; pp 785–794.
(62) Huryn, D. M.; Kornfilt, D. J. P.; Wipf, P. p97: An Emerging Target for Cancer, Neurodegenerative Diseases, and Viral Infections. *J. Med. Chem.* **2020**, *63* (5), 1892–1907.
(63) Tang, W. K.; Odzorig, T.; Jin, W.; Xia, D. Structural Basis of p97 Inhibition by the Site-Selective Anticancer Compound CB-5083. *Mol. Pharmacol.* **2019**, *95* (3), 286–293.
(64) Chapman, E.; Maksim, N.; de la Cruz, F.; La Clair, J. J. Inhibitors of the AAA+ Chaperone p97. *Molecules* **2015**, *20* (2), 3027–3049.
(65) Cathomen, A.; Maier, D.; Kriz, J.; Abel, R.; Röhrich, F.; Baumberger, M.; Scivoletto, G.; Weidner, N.; Rupp, R.; Jutzeler, C. R.; Steeves, J. D.; Curt, A.; Bolliger, M.; EMSCI study group. Walking Outcome After Traumatic Paraplegic Spinal Cord Injury: The Function of Which Myotomes Makes a Difference? *Neurorehabil. Neural Repair* **2023**, 15459683231166937.
(66) Frye, L.; Bhat, S.; Akinsanya, K.; Abel, R. From Computer-Aided Drug Discovery to Computer-Driven Drug Discovery. *Drug Discov. Today Technol.* **2021**, *39*, 111–117.
(67) Albanese, S. K.; Chodera, J. D.; Volkamer, A.; Keng, S.; Abel, R.; Wang, L. Is Structure-Based Drug Design Ready for Selectivity Optimization? *J. Chem. Inf. Model.* **2020**, *60* (12), 6211–6227.
(68) 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.