Prediction of Drug-Target Binding Kinetics for Flexible Proteins by Comparative Binding Energy Analysis

Ariane Nunes-Alves,[*,†,‡] Fabian Ormersbach,[†] and Rebecca C. Wade[*,†,‡,¶]

†*Heidelberg Institute for Theoretical Studies (HITS), Schloß-Wolfsbrunnenweg 35, 69118 Heidelberg, Germany*

‡*Center for Molecular Biology (ZMBH), DKFZ-ZMBH Alliance, Heidelberg University, Im Neuenheimer Feld 282, 69120 Heidelberg, Germany*

¶*Interdisciplinary Center for Scientific Computing (IWR), Heidelberg University, Im Neuenheimer Feld 205, Heidelberg, Germany*

E-mail: ariane.nunes-alves@h-its.org; rebecca.wade@h-its.org

Abstract

There is growing consensus that the optimization of the kinetic parameters for drug-protein binding leads to improved drug efficacy. Therefore, computational methods have been developed to predict kinetic rates and to derive quantitative structure-kinetic relationships (QSKRs). Many of these methods are based on crystal structures of ligand-protein complexes. However, a drawback is that each protein-ligand complex is usually treated as having a single structure. Here, we present a modification of COMparative BINding Energy (COMBINE) analysis, which uses the structures of protein-ligand complexes to predict binding parameters. We introduce the option to use multiple structures to describe each ligand-protein complex into COMBINE analysis and apply this to study the effects of protein flexibility on the derivation of dissociation rate

1

constants (k_{off}) for inhibitors of p38 mitogen-activated protein (MAP) kinase, which has a flexible binding site. Multiple structures were obtained for each ligand-protein complex by performing docking to an ensemble of protein configurations obtained from molecular dynamics simulations. Coefficients to scale ligand-protein interaction energies determined from energy-minimized structures of ligand-protein complexes were obtained by partial least squares regression and allowed the computation of k_{off} values. The QSKR model obtained using single, energy minimized crystal structures for each ligand-protein complex had a higher predictive power than the QSKR model obtained with multiple structures from ensemble docking. However, the incorporation of protein-ligand flexibility helped to highlight additional ligand-protein interactions that lead to longer residence times, like interactions with residues Arg67 and Asp168, which are close to the ligand in many crystal structures. These results show that COMBINE analysis is a promising method to guide the design of compounds that bind to flexible proteins with improved binding kinetics.

1 Introduction

The realization that drug binding kinetic rates can be key determinants of drug efficacy in vivo[1–3] has led to high interest in the development of computational methods to estimate the drug-target dissociation rate constant (k_{off}) or its inverse, the residence time ($\tau = 1/k_{off}$). A range of computational methods have been proposed to estimate k_{off} values at low to moderate computational cost,[4] and these may facilitate rational drug design. On one hand, there are methods based on molecular dynamics (MD) simulations, such as τRAMD[5] and scaled MD,[6] in which a force or energy bias is added to the system, thereby allowing relative k_{off} values for a set of compounds to be estimated within a few days. Other methods, such as weighted ensemble,[7,8] allow absolute k_{off} values to be estimated at a higher computational cost. On the other hand, there are chemometric methods such as COMparative BINding Energy (COMBINE) analysis[9] and Volsurf,[10] which require an experimental dataset with at

least tens of ligands for training a system-specific model, but then allow one to estimate k_{off} values in seconds or minutes by using chemical descriptors of the ligand or of the structure of the ligand-protein complex.

Here, we used COMBINE analysis[11] to predict k_{off} values for ligand-protein complexes. Among chemometric methods, COMBINE analysis has the advantages of using information available from experimental structures of ligand-protein complexes and identifying specific protein residues that affect the computed parameters. In COMBINE analysis, the interaction energies (E) between the ligand and each protein residue computed from the structure of the complex using a classical molecular mechanics force field can be decomposed:

$$E = \sum_{k=1}^{n_p} E_k^{vdW} + \sum_{k=1}^{n_p} E_k^{elec} \tag{1}$$

where E^{vdW} and E^{elec} designate van der Waals and electrostatic interaction energies, respectively, n_p is the total number of protein residues and k is the index of residues. Eq. 1 includes only the terms from a classical force field; other terms can be added, such as changes in solvation energy computed with a continuum model or dependent on buried surface area, or changes in the intramolecular energy of the ligand or the protein upon (un)binding. Then, n interaction energies (E_i) that have standard deviations above a defined threshold and are therefore expected to contribute the most to the differences in k_{off} values are selected and scaled by coefficients or weights (w_i) obtained from partial least squares (PLS) regression for a training set:

$$log \ k_{off} \approx \sum_{i=1}^{n} w_i E_i + C \tag{2}$$

where i is the index of the selected interaction energies, and C is a constant that accounts for factors not included in E_i, such as covalent energy terms, entropy gain or solvation of the ligand or of the protein binding site during the (un)binding process. First, eq. 2 is used with computed E_i values and experimental k_{off} values to allow the assignment of w_i

and C values. Once known, the w_i and C values are employed to compute k_{off} values for other similar complexes. COMBINE analysis was recently used to predict k_{off} values for inhibitors of heat shock protein 90 (HSP90) and HIV-1 protease,[9] achieving good predictive power (coefficient of determination for cross-validation, Q^2, of 0.69 and 0.70, respectively). COMBINE analysis was used originally to predict inhibitory activities for complexes of the human synovial fluid phospholipase A2 with small molecules.[11] Moreover, it has already been applied to predict affinities or inhibitory activities for complexes of proteins with small molecules,[12–14] peptides[15] and other proteins.[16]

One of the potential drawbacks of COMBINE analysis is that the ligand-protein complex is represented by a single energy minimized structure, which limits the representation of protein flexibility and ligand binding modes. Sometimes ligands or side chains have more than one possible configuration in a single crystal structure. Moreover, ligand-protein complexes are flexible in in vivo conditions, so considering this flexibility might be expected to improve quantitative structure-kinetic relationship (QSKR) models or quantitative structure-activity relationship (QSAR) models derived by COMBINE analysis.

Different strategies can be used to represent protein-ligand complexes flexibility. For instance, a set of protein configurations from crystal structures[17] or nuclear magnetic resonance can be employed. Representative sets of protein configurations can also be obtained from computational methods, such as MD simulations,[18–21] machine learning[22,23] or elastic network models.[24] The representation of protein flexibility using ensembles of configurations was shown to be beneficial in docking, leading to improved results in the prediction of binding affinities[19] and in the identification of true binders in a large set of small molecules.[17,20]

The aim of this work was to extend the COMBINE analysis method to allow the use of an ensemble of structures to represent one ligand-protein complex, in order to incorporate the effects of structural flexibility in the predictions of binding parameters. Two COMBINE analysis models were built to predict k_{off} values, one using one energy minimized crystal structure to represent each ligand-protein complex, and one using 10 structures from ensem-

ble docking to represent each ligand-protein complex. The ensemble of protein configurations for docking was obtained from MD simulations. p38 mitogen-activated protein (MAP) kinase, for which k_{off} values are available for a number of complexes, was selected as a model system. This choice was motivated by the fact that there is a region (residues 170 to 184 in PDB 1KV2[25]) close to the binding site that is missing in many crystal structures, which suggests that it is disordered and flexible. Moreover, due to its location, this region may have transient interactions with inhibitors, playing a role in ligand binding kinetics. Type-II inhibitors, which occupy an allosteric pocket close to the ATP binding site,[26] were selected for this work due to the spread of experimental k_{off} values and the availability of crystal structures. They occupy a pocket created by a conformational change in the conserved amino acid sequence Asp-Phe-Gly (DFG motif, residues 168-170) in the activation loop, and bind to the out (inactive) conformation of the DFG motif of p38 MAP kinase.[26]

2   Computational methods

2.1   Dataset

The full dataset for the COMBINE analysis model was composed of 33 inhibitors of p38 MAP kinase (Figure S1, Table S1). These inhibitors were chosen based on the availability of (i) experimental k_{off} values[25,27–30] (Table S2); and (ii) experimental structures of the inhibitor bound to p38 MAP kinase or to other kinases in the DFG-out conformation (Table S2), or similarity to inhibitors with experimental structures available for the complex. The dataset was split into a training set and a test set, with 22 and 11 inhibitors, respectively. The dataset was ranked according to k_{off} values and every third ligand was assigned to the test set. Using this procedure, ligands 3, 5, 6, 12, 14, 16, 21, 24, 28, 34 and 35 were selected to be part of the test set (Table S2). Additional tests were performed with three other test sets selected similarly, namely: ligands 1, 2, 8, 15, 17, 22, 26, 27, 31, 32 and 36; ligands 7, 9, 11, 13, 18, 20, 23, 25, 29, 30 and 33; or ligands 3, 5, 12, 14, 16, 21, 24, 34 and 35. The

results were similar (data not shown).

2.2 Parametrization of COMBINE analysis model

First, two sets of structures for the dataset were modeled and used to derive COMBINE analysis models (sections 2.4 and 2.5). Then, structures were energy minimized (section 2.6). After energy minimization, interaction energy components for the training set and test set were obtained using the COMBINE software[31] and the AMBER ff14SB force field[32] to describe parameters for bonded and non-bonded interactions. Electrostatic interactions were computed using a uniform dielectric constant of 4. Interaction energy components that had a standard deviation greater than the cutoff value of 0.25 kcal/mol were selected for regression. Then, coefficients or weights to scale the ligand-protein interaction energies were obtained by COMBINE analysis[11] using partial least squares regression[33] (section 2.3). Up to 10 latent variables were tested to decompose the data in partial least squares regression.

2.3 Modification of COMBINE software to use multiple structures to represent one ligand-protein complex

Previously, COMBINE analysis has been performed using one structure to represent one protein-ligand complex and estimate one response (in this case, log I, where I can be k_{off}, binding affinity, inhibitory activity or any other parameter related to ligand-protein interactions). Here, the COMBINE software[31] was modified to retrieve an average response using N structures for each ligand-protein complex. Multiple structures for one complex were obtained using ensemble docking (details in section 2.5). Each structure is treated independently during the regression step to obtain weights for interaction energies. The responses obtained from the N structures are averaged, resulting in one final response for the complex. The average can be arithmetic or exponential.

Previously:[11]

$$(log\ I)^{comp} = log\ I^1 \tag{3}$$

where $(log\ I)^{comp}$ is the prediction for the response variable and $log\ I^1$ is the prediction made using 1 structure to represent the ligand-protein complex.

In the ensemble method introduced here, exponential (eq. 4) or arithmetic averages (eq. 5) can be used to obtain one response variable from N structures:

$$(log\ I)^{comp}_{exp} = -log\frac{1}{N}\sum_{j=1}^{N}e^{-log\ I^j} \tag{4}$$

$$(log\ I)^{comp}_{arit} = \frac{1}{N}\sum_{j=1}^{N}log\ I^j \tag{5}$$

where j is the index of the structure used. Eqs. 4 and 5 give the same result as eq. 3 for N=1 or for N identical structures representing one ligand-protein complex. The exponential average proposed is based on previous studies using protein conformational averaging for an adapted linear interaction model[19] and the implicit ligand theory.[34] Here, $(log\ I)^{comp}_{arit}$ and $(log\ I)^{comp}_{exp}$ are the total computed log I for one ligand-protein complex, while $log\ I^j$ is the log I computed from weighted interaction energy terms for structure j.

Two new parameters were added in the COMBINE analysis input file: N, the number of structures to describe one ligand-protein complex, and the averaging scheme to be used (arithmetic or exponential). The cross-validation part of the COMBINE software was also modified so that all structures of one ligand-protein complex are removed from the training set when necessary.

2.4 Modeling of protein-ligand structures using data from X-ray crystallography

Two sets of structures were used to derive COMBINE analysis models. In one set, each complex was represented using one structure, usually a crystal structure, described in this section. In the other set, each complex was represented using 10 structures from ensemble docking (section 2.5).

Crystal structures of protein-ligand complexes retrieved from the Protein Databank (PDB) were used after removal of water and other crystallization molecules. Ligands 1, 2, 16, 25, 32 and 33 had crystal structures available (PDBs 1KV1,[25] 1KV2,[25] 3HV6,[35] 3HEC,[36] 2YIW[28] and 2YIS,[28] respectively, Table S2). Incomplete regions were built using homology modeling with Modeller[37] and the structure from PDB file 1WBV[38] as a reference. This structure was chosen because it is fully resolved and has good resolution (2 Å). For each structure, 20 models were built and the one with the lowest Discrete Optimized Protein Energy (DOPE) score was selected. Incomplete side chains were built with Modeller. Hydrogens were added using pdb2pqr[39,40] at pH 7.2 for the protein and Babel[41] for the ligands. All protein structures had a net charge of -7 e.

Ligands 3-15, 17-24, 26-31 and 34-36 had no crystal structures in complex with p38 MAP kinase available. A ligand search in KLIFS,[42,43] a structural database for kinases, using SMILES showed that ligands 20-24 and 26-31 had crystal structures available in complexes with other kinases (PBDs 4RZW,[44] 3F82,[45] 3QRI,[46] 3LQ8,[47] 5IA5,[48] 3EFL, 3GP0, 6FNI,[49] 4U0I,[50] 4RT7[51] and 4ASE,[52] respectively, Table S2). Complexes with p38 MAP kinase for these ligands were built using the binding modes in the available crystal structures as references. Complexes with p38 MAP kinase for ligands 3-15, 17-18 and 34-36 were built using the binding mode of ligand 16 (PDB 3HV6) as a reference. Ligand superimposition was performed using a custom script built using Python 3.8 and the packages RDkit 2020.03.3[53] and MDAnalysis 1.0.0[54,55] (details and validation using ligRMSD[56] are given in section S2).

The ligand geometry was built using Molden[57] and energy minimized using AMBER 2016[58] and the General AMBER Force Field (GAFF)[59] if crystal structures were not available.

2.5   Modeling of protein-ligand structures using ensemble docking

A configurational ensemble was generated for p38 MAP kinase bound to ligand 16. An ensemble from the apo protein was not tested, as a previous study showed that ensembles generated from apoprotein structures can have occluded binding sites.[19] The structure of the ligand-protein complex was energy minimized, and a molecular dynamics simulation was run for 1200 nanoseconds (see details in section 2.6). The ensemble was built using 10 configurations (excluding the ligand) collected along the trajectory at time intervals of 80 nanoseconds for the last 800 nanoseconds of simulation. For each configuration in the ensemble, 20 ligand poses were generated by docking, and the top pose (irrespective of the binding mode) was selected, resulting in a total of 10 ligand-protein structures for each ligand.

Additionally, another configurational ensemble for p38 MAP kinase was built using clustering. 80 protein structures were collected along a 1200-nanosecond trajectory at time intervals of 10 nanoseconds for the last 800 nanoseconds of simulation. The 80 protein structures were grouped using hierarchical agglomerative clustering and an epsilon value of 1.5 Å. This procedure generated five clusters with fractional populations of 0.475, 0.15, 0.1, 0.087 and 0.075, and four clusters with fractional populations of 0.037, 0.037, 0.025 and 0.013. The clusters with populations equal to 0.037 or lower were discarded. The five remaining clusters were used for docking, with approximate populations of 0.5, 0.2, 0.1, 0.1 and 0.1.

Docking to the ensembles was performed with AutoDock Vina.[60] Parameters for docking were optimized by redocking and cross-docking (section 3.2), resulting in the choice of an exhaustiveness level of 8 and a cubic grid with a spacing of 0.375 Å and 70 points along each edge centered in the known binding site. Bond rotations were allowed in ligands, while the protein structures were kept rigid.

2.6 Force field and simulation details

AMBER 2016[58] was used for energy minimization and MD simulations. Parameters to describe the protein and the ligand were obtained from the AMBER ff14SB force field[32] and GAFF,[59] respectively. RESP[61,62] partial atomic charges for ligands were obtained using electrostatic potentials from quantum mechanical calculations performed in GAMESS[63] using single-point calculations (HF/6-31G**) after energy minimization of the ligand structure (HF/6-31G).

Energy minimization of the complexes was performed using implicit solvent prior to the use of the complexes to obtain interaction energies in COMBINE analysis, or using explicit solvent prior to performing MD simulations. Both procedures are described further below.

Energy minimization of ligand-protein complexes prior to their use in COMBINE analysis was performed in two steps. In the first step, a harmonic restraint with a force constant of 100 kcal/(mol Å^2) was applied to the positions of the heavy atoms. In the second step, no positional restraints were used. For each step of energy minimization, 1000 steps of the steepest descent algorithm followed by 4000 steps of the conjugate gradient algorithm were applied. Energy minimization was performed using implicit solvent. The generalized Born (GB) approximation as implemented in the GBn model[64] was used. The nonpolar contribution was calculated using the LCPO model[65] to compute the surface area and a surface tension of 0.005 kcal/(mol Å^2).[66]

MD simulations to obtain the configurational ensemble were carried out at 300 K with a 4 femtosecond time-step, which was possible due to the use of the hydrogen mass repartitioning method.[67] The structure of ligand 16 bound to p38 MAP kinase from PDB 3HV6[35] was used after removal of crystallization cosolute molecules (waters were maintained). Missing regions and residues were modeled as described previously (section 2.4). Hydrogens were added using pdb2pqr[39,40] at pH 7.2 for the protein and Babel[41] for the ligand. The complex was solvated in a periodic box of TIP3P water molecules[68] with a distance of 15 Å from the solute

to the box edge using tleap. 73 sodium ions and 66 chloride ions were added to achieve neutrality and an approximate ionic strength of 150 mM. The system was energy minimized with 1000 steps of the steepest descent algorithm followed by 1000 steps of the conjugate gradient algorithm, with restraints of 100 kcal/(mol Å^2) on the heavy atoms of the protein and the ligand. The system was heated in cycles of 100-picosecond simulations with gradual temperature increases (10, 50, 150 and 300 K) using the Berendsen thermostat.[69] Then, the pressure was adjusted to 1 atm with the Berendsen barostat[69] in a 400-ps simulation. The 100 kcal/(mol Å^2) restraints were gradually removed in cycles of 500-picosecond simulations (50, 10, 5 and 1 kcal/(mol Å^2)). Then, for production MD simulations, the Langevin thermostat was used to keep the temperature at 300 K, and the Monte Carlo barostat was used to keep the pressure at 1 bar (NPT ensemble). The configurational ensemble obtained from the structure of ligand 16 bound to p38 MAP kinase was used for docking all the ligands in the dataset.

Additionally, 50-nanosecond MD simulations were performed to generate one ensemble for each ligand-protein complex. The structures of the ligand-protein complexes modeled using data from X-ray crystallography, as described in section 2.4, were used as initial configurations. The procedure for system set up and equilibration was the same as the one described in the previous paragraph. The ensemble for each ligand-protein complex was composed of 10 structures collected along the trajectory at time intervals of 5 nanoseconds.

2.7 Sequence alignment

Sequence alignment of p38 MAP kinase with other human kinases was performed to validate the functional role of the residues identified in the COMBINE analysis. Protein sequences were obtained from UniProt[70] and pairwise sequence alignment was performed using needle.[71,72]

3 Results

3.1 COMBINE analysis using one crystal structure to represent one ligand-protein complex

We first made a model representing the protein as one structure to use it as a reference to assess the usefulness of incorporating protein flexibility in COMBINE analysis. Therefore, one energy-minimized structure from X-ray crystallography was used to represent each ligand-protein complex and obtain interaction energies. Using COMBINE analysis, 348 electrostatic and 348 van der Waals interaction energies for each of the 22 p38 MAP kinase inhibitors in the training set were calculated. These interaction energies correspond to the 348 residues in the protein. Then the interaction energies which had a standard deviation greater than the specified cutoff value were selected to remove energy terms that may not contribute to the differences in k_{off} values. A cutoff value of 0.25 kcal/mol was used, since it resulted in better models in a previous study,[9] resulting in the selection of 36 van der Waals interaction energies and 14 electrostatic interaction energies (Figure S2). Next, the weights of these 50 interaction energies for different numbers of latent variables were determined using PLS regression to correlate the interaction energies with the experimental log k_{off} values (Figure 1A). The coefficients of determination (R^2tr and Q^2 for the training and cross-validation sets, respectively), the average absolute errors (AAEtr and AAEv for the training and cross-validation sets, respectively) and the root mean squared errors (RMEtr and RMEv for the training and cross-validation sets, respectively) were calculated using different numbers of latent variables (Tables 1 and S4; equations given in section S1). The model with two latent variables was chosen, since it presented the highest Q^2 value, 0.75, when the leave-one-out method was used for cross-validation (Table S4). The AAEv and RMEv values were 0.68 log s^{-1} and 0.75 log s^{-1}, respectively, for the leave-one-out method for cross-validation. Moreover, the Q^2, AAEv and RMEv values were similar for the different cross-validation methods

used (Table 1), showing that the model with two latent variables is consistent. The decrease in Q^2 when moving from the leave-one-out method to the leave-two-out and leave-three-out methods was expected, since fewer structures were used in the training set in the latter. This model presented a R^2tr of 0.86, showing it could successfully fit the training set (Figure 1B, Table S4). The coefficient of determination for the test set (R^2te) was 0.79 (Figure 1B, Table 1), showing this model had high predictive power.

Table 1: Statistical measures of the validation and test sets for COMBINE analysis models derived for log k_{off} of p38 MAP kinase inhibitors. Q^2, R^2te: coefficient of determination of the cross-validation and test sets; AAEv, AAEte: average absolute error of the cross-validation and test sets (in log s^{-1}); RMEv, RMEte: root mean squared error of the cross-validation and test sets (in log s^{-1}).

|	single[a]			ensemble[b]		
cross-validation set	Q^2	AAEv	RMEv	Q^2	AAEv	RMEv
leave-one-out (LOO)	0.75	0.68	0.75	0.70	0.68	0.82
leave-two-out (L2O)	0.72	0.68	0.79	0.67	0.70	0.85
leave-three-out (L3O)	0.65	0.74	0.88	0.62	0.74	0.92
test set	R^2te	AAEte	RMEte	R^2te	AAEte	RMEte
	0.79	0.54	0.64	0.72	0.65	0.74

[a]One structure per ligand-protein complex, two latent variables. [b]Ten structures per ligand-protein complex, one latent variable.

The interaction energies with the highest weights were the van der Waals interaction energies of the inhibitors with residues Lys53, Arg70, Glu71 and Leu167 and the electrostatic interaction energies with residues Lys53 and Glu71, which surround the binding site (Figures 1A and 1C). The weight for the electrostatic interaction with Lys53 was negative, meaning that favorable interactions with this residue resulted in high k_{off} values. On the other hand, the weights for the electrostatic interaction with Glu71 and most van der Waals interactions were positive, indicating that favorable interactions with these residues result in low k_{off} values. Inhibitors with low k_{off} values usually make two hydrogen bonds with Glu71 and have two bulky hydrophobic groups, one making nonpolar interactions with Lys53 and the other making nonpolar interactions with the pocket formed by Arg70 and Glu71 (Figure 1C), in agreement with the interaction energies with highest weights identified in the model.

A van der Waals energies

A electrostatic energies

B single

B ensemble
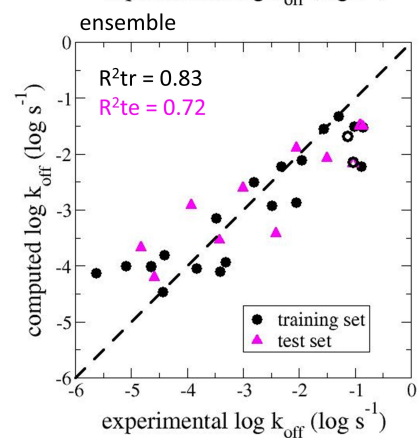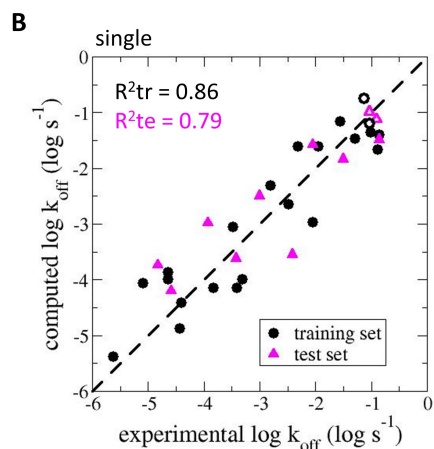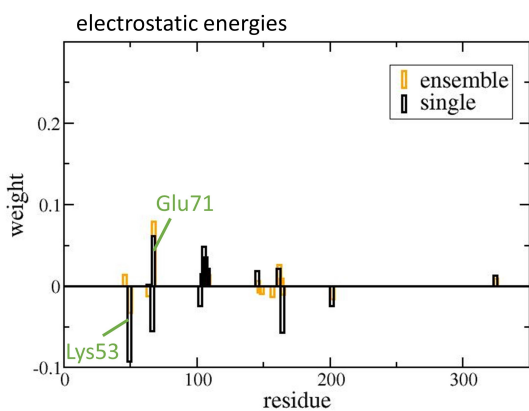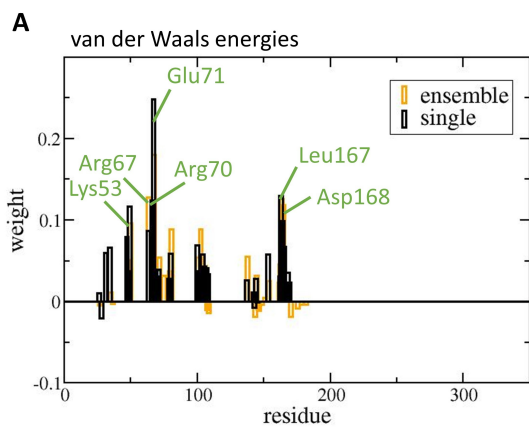
C Ligand 2 (BIRB 796), $k_{off} = 8.3 \cdot 10^{-6}$ s^{-1} (low k_{off})
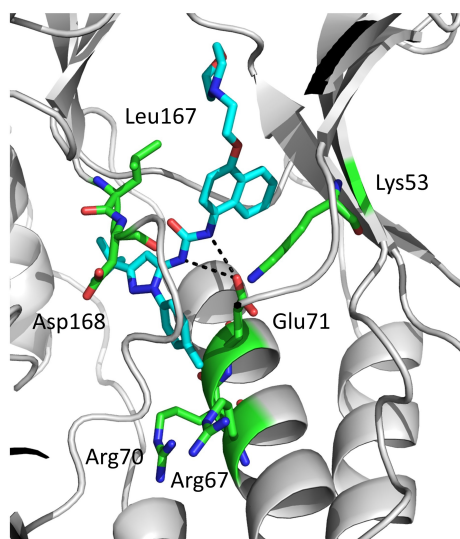
C Ligand 25 (Imatinib), $k_{off} = 1.3 \cdot 10^{-1}$ s^{-1} (high k_{off})
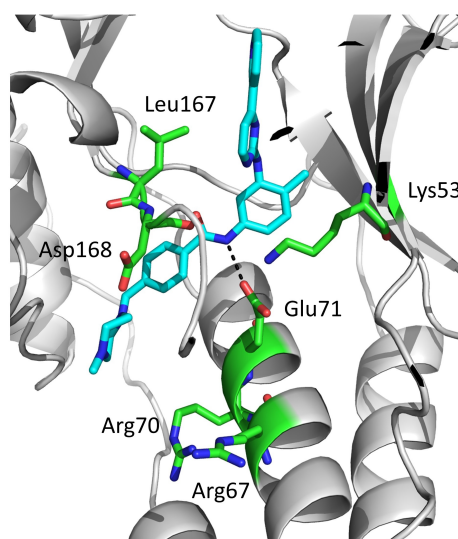
Figure 1: COMBINE analysis models to compute k_{off} values for p38 MAP kinase inhibitors. (A) Weights for different van der Waals and electrostatic interaction energies derived from partial least squares regression (PLS) using single structures (black) or ensemble docking (orange, bars displaced by 0.50 on the x axis to aid visualization). Results are shown for 2 or 1 latent variables for single structures or ensemble docking, respectively. Highlighted residues are shown in green in panel (C). (B) Plots of computed versus experimental log k_{off} values for the training and test sets (22 and 11 inhibitors, respectively) for PLS regression using single structures or ensemble docking. The straight line corresponds to y = x (ideal case). The empty symbols highlight the inhibitors with poor binding modes in ensemble docking: 23, 26 (training set), 21 and 24 (test set). (C) Comparison of binding modes and key interactions of a slow (ligand 2, BIRB 796, PDB 1KV2) and a fast dissociating inhibitor (ligand 25, Imatinib, PDB 3HEC). Protein backbone in gray, ligands and residues colored according to atom type, with carbon atoms in cyan and green, respectively. Hydrogen bonds are shown as black dotted lines.

3.1.1 Comparison of models using different energy contributions

A COMBINE model was tested including the solvation of the ligand as an extra term given by the change in solvent accessible surface area of the ligand upon unbinding. The model with the best Q^2 value (2 latent variables) had a performance similar to the model without solvation (Table S5).

Electrostatic interactions were computed using a uniform dielectric constant. Tests were also performed with a distance-dependent dielectric constant for which the predictive performance was less good (R^2te of 0.76 versus R^2te of 0.79 for the model with uniform dielectric constant, Table S6).

3.2 COMBINE analysis using multiple structures from ensemble docking to represent one ligand-protein complex

After making the reference model with one structure for each ligand, we made a COMBINE analysis model using a structural ensemble. A 1200-nanosecond MD simulation of p38 MAP kinase was performed to obtain different protein configurations. Protein configurations were collected every 80 nanoseconds for the last 800 nanoseconds of simulation, resulting in 10

protein configurations. Then, docking of the 33 p38 MAP kinase inhibitors to each of the 10 protein configurations was performed to obtain a total of 330 protein-ligand complexes. AutoDock Vina[60] was used for ligand-protein docking, since it was able to reproduce the crystallographic poses of ligands 2 and 16 in redocking and cross-docking tests (ligand root mean squared deviation, RMSD, lower than 2 Å in all cases, Table S3). For most ligands, ligand poses obtained from ensemble docking were similar (Figure S3) and had low RMSD from the ligand pose in the complex used in the COMBINE model with one structure to represent each complex. Ligands 21, 23, 24 and 26 are exceptions, as most ligand poses obtained from docking had high RMSD values from the ligand poses in the COMBINE model with one structure for each complex (Figure S4, Table S16 and discussion in section 4).

Ten energy-minimized structures from ensemble docking were used to represent each ligand-protein complex and obtain interaction energies. Using COMBINE analysis, 348 electrostatic and 348 van der Waals interaction energies for each of the 220 complexes (10 structures for each of the 22 p38 MAP kinase inhibitors) in the training set were calculated. Then the interaction energies that had a standard deviation greater than 0.25 kcal/mol were selected, resulting in 43 van der Waals interaction energies and 16 electrostatic interaction energies (Figure S2). Next, the weights of these 59 interaction energies for different numbers of latent variables were determined using PLS regression to correlate the interaction energies with the experimental log k_{off} values (Figure 1A). Independent log k_{off} values were computed for each structure describing one ligand-protein complex, and these values were later assembled in one computed k_{off} value for the complex using an exponential average (section 2.3). R^2, Q^2, AAEtr, AAEv, RMEtr and RMEv values were calculated using different numbers of latent variables (Tables 1 and S7). The model with one latent variable was selected, since it had the best compromise between high Q^2 value (0.70 when the leave-one-out method was used for cross-validation) and low number of latent variables (Table S7). The AAEv and RMEv values were 0.68 log s^{-1} and 0.82 log s^{-1}, respectively, for the leave-one-out method

for cross-validation. Moreover, the Q^2, AAEv and RMEv values were similar for the different cross-validation methods used (Table 1), showing that the model with one latent variable is consistent. The model with one latent variable presented a R^2tr of 0.83, showing it could fit the data of the training set (Figure 1B, Table S7). The R^2te value for the test set of 11 inhibitors was 0.72 (Figure 1B, Table 1), showing this model had good predictive power.

The interaction energies with the highest weights were the van der Waals interaction energies of the inhibitors with residues Arg67, Glu71, Leu167 and Asp168 and the electrostatic interaction energies with residues Lys53 and Glu71 (Figures 1A and 1C). Some of the major contributions identified using COMBINE analysis and single structures, like electrostatic interactions with Lys53 and Glu71 and van der Waals interactions with Glu71 and Leu167, were also captured in the model with ensemble docking. Moreover, the use of structural ensembles helped to highlight the contributions of interactions with residues Arg67 and Asp168 in the DFG motif, which came closer to the binding site during MD simulations (Figure 2). The contributions from these two residues were high in the model using single structures, and became dominant in the model using an ensemble of structures.

3.2.1 Comparison of models using different ensembles and averaging schemes

Tests with COMBINE analysis were also performed using 20 instead of 10 structures to represent each protein-ligand complex, or using one structure, the one with the highest Vina score, to represent each protein-ligand complex (Table 2). A higher number of interaction energies was selected when more structures were used (52, 59 and 68 for 1, 10 and 20 structures, respectively). For one structure, the model with two latent variables had the best compromise between high Q^2 value (0.65 when the leave-one-out method was used for cross-validation) and low number of latent variables (Tables 2 and S8). For 20 structures, the model with one latent variable had the best compromise between high Q^2 value (0.65) and low number of latent variables (Tables 2 and S9). The highest predictive power was achieved with the model with one structure, while the predictive power of models with
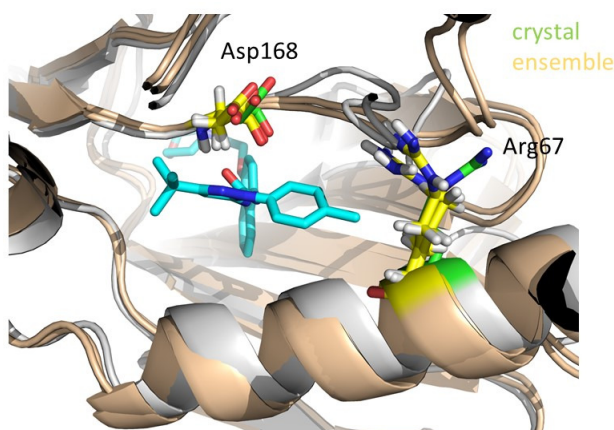
Figure 2: The use of ensemble docking to include protein and ligand flexibility in COMBINE analysis helped to highlight the contributions of interactions with residues Arg67 and Asp168. Arg67 and Asp168 side chains moved closer into the binding site during MD simulations. Ligand 2 (BIRB 796), Arg67 and Asp168 are shown in sticks. Protein backbone in gray (crystal structure, PDB 1KV2) or wheat (structure from MD simulation), ligand and residues colored according to atom type, with carbon atoms in cyan (ligand), green (residue in crystal structure) or yellow (residue in MD simulation).

multiple structures was similar (R^2te values of 0.80, 0.72 and 0.71 for 1, 10 and 20 structures, respectively). The maximum R^2te value achieved by the COMBINE analysis models with ensemble docking was 0.80, which is similar to the R^2te value achieved with the model which used single structures, 0.79, indicating that the structural information to achieve predictive models using single crystal structures can be recovered in ensemble docking.

Another scheme to obtain protein configurations was also tested. 80 protein configurations from the 1200-nanosecond MD simulation of p38 MAP kinase were clustered using hierarchical agglomerative clustering, resulting in five clusters (details in section 2.5). These clusters had approximate relative weights of 5:2:1:1:1. The structures corresponding to the centers of these clusters were used for docking and the complexes were subsequently energy minimized. The resulting ligand-protein complexes, with weights proportional to the weig ht of the cluster the protein configuration belonged to, were used in COMBINE analysis. For example, a complex from a protein of weight 5 was represented 5 times in the training or test set. The resulting R^2te (Tables 2 and S10) was similar to the one obtained using

Table 2: Statistical measures of the validation and test sets for COMBINE analysis models derived for log k_{off} of p38 MAP kinase inhibitors using different ensembles and averaging schemes. Q^2, R^2te: coefficient of determination of the cross-validation and test sets; AAEv, AAEte: average absolute error of the cross-validation and test sets (in log s^{-1}); RMEv, RMEte: root mean squared error of the cross-validation and test sets (in log s^{-1}). LV: number of latent variables for the model with the best compromise between high Q^2 value and low number of latent variables.

ensemble type	size	cross-validation set			test set			LV
		Q^2	AAEv	RMEv	R^2te	AAEte	RMEte	
MD sim. (1)[a]	1	0.65	0.73	0.88	0.80	0.49	0.63	2
MD sim. (1)[b]	10	0.70	0.68	0.82	0.72	0.65	0.74	1
MD sim. (1)[c]	20	0.65	0.75	0.88	0.71	0.66	0.76	1
MD sim. (1) + cluster.[d]	5	0.70	0.64	0.82	0.70	0.65	0.77	3
MD sim. (33)[e]	10	0.80	0.59	0.67	0.80	0.52	0.63	3
MD sim. (1) + arith. av.[f]	10	0.78	0.58	0.70	0.71	0.61	0.76	1
MD sim. (1) + exp. av.[g]	10	0.71	0.70	0.81	0.66	0.71	0.82	2

[a]10 protein configurations for ensemble docking obtained from 1 MD simulation; the complex with the highest Vina score was selected. [b]10 protein configurations for ensemble docking obtained from 1 MD simulation. [c]20 protein configurations for ensemble docking obtained from 1 MD simulation. [d]5 protein configurations for ensemble docking obtained from clustering of structures from 1 MD simulation. [e]10 configurations of ligand-protein complexes obtained from 33 independent MD simulations (one for each inhibitor). [f]10 protein configurations for ensemble docking obtained from 1 MD simulation; arithmetic averages of individual interaction energies from the 10 different complexes were used as input. [g]10 protein configurations for ensemble docking obtained from 1 MD simulation; exponential averages of individual interaction energies from the 10 different complexes were used as input.

an ensemble of 10 protein configurations collected at regular intervals from the MD simulation, at the expenses of a higher number of latent variables. The relative weights of the clusters were also similar when the cluster composition was determined using the 10 protein configurations: 3:2:1:1:0 (the other 3 protein configurations belonged to smaller clusters, not considered in the analysis).

Previous studies have shown that short MD simulations to relax the structures of complexes obtained from docking can lead to better results in distinguishing active from decoy ligands.[73] Here, we also tested ensembles generated from MD simulations to relax the initial structures for each ligand-protein complex. One ensemble was generated for each

protein-ligand complex by collecting 10 structures at regular intervals of 5 nanoseconds from 50-nanosecond MD simulations performed for each ligand-protein complex. The structures of these ensembles were energy minimized and used in COMBINE analysis. The resulting R^2te (Tables 2 and S11) was better than that obtained using a single ensemble of 10 protein configurations to perform docking for all ligands. This was expected, as the single simulation sampled favorable protein configurations for the binding of one ligand, the one used in the simulation, whereas the simulations for each ligand could sample the favorable protein conformations for each particular ligand. However, it should be noted that this scheme can lead to high computational costs when a large number of ligands is present in the dataset as each protein-ligand complex has to be subjected to MD simulation.

COMBINE analysis was also performed using averages of the individual interaction energies obtained from the 10 structures of ligand-protein complexes obtained from ensemble docking using one protein ensemble. The R^2te of the best model using arithmetic averages (Tables 2 and S12) had a value similar to the models using averaging of 10 or 20 structures, while the R^2te of the best model using exponential averages (Tables 2 and S12) had a lower value. It should be noted that this kind of averaging can lead to unrealistic favorable interactions for one ligand-protein complex. For example, if favorable interactions between the ligand and two residues are mutually exclusive and only appear in different structures, in an exponential averaging scheme the interactions with these two residues can appear as simultaneously favorable.

3.2.2 Comparison of models using single or multiple structures

Comparisons between the different COMBINE analysis models obtained using a single crystal structure or structures from ensemble docking to represent one ligand-protein complex can help us to understand the weaknesses and strengths of each model. A similar R^2tr value was obtained from COMBINE analysis using one or multiple structures (0.86 vs. 0.83, respectively; Tables S4 and S7), indicating that both models can basically make good fits

for the training set. Both models could also reproduce the ranking for a small ligand series composed of ligands 13 (training set), 3 and 14 (test set), which differ by only one substitution (Figure 3). However, the overall R^2te value was higher for the model using one structure (0.79 and 0.72 for the models using one or multiple structures, respectively, Table 1), indicating that the group of interaction energies selected in the model with one structure had a higher predictive power. Possible reasons for the selection of interaction energies with lower predictive power in the model with multiple structures include worse signal to noise ratio, the lack of sampling of important protein configurations or poor ligand positions obtained from docking. The latter was the case for some ligands, as their RMSD value between docking and crystallographic positions was higher than 2 Å (Figure S4 and discussion in section 4). For instance, considering the test set, ligands 5 and 16 had a high prediction error (difference between experimental and computed log k_{off} higher than 1 log s^{-1}) for the model using one structure. These ligands also had a high prediction error for the model using multiple structures, together with ligand 21, which had a poor prediction of ligand positions from docking.

3.2.3 Comparison of models to predict k_{off} and K_d values

COMBINE analysis was also used to predict equilibrium dissociation constants (K_d) using a single crystal structure or structures from ensemble docking to represent each complex. The R^2tr and R^2te values obtained for these models (tables S13 and S14) were lower than the ones obtained from COMBINE analysis to predict k_{off} values, showing that model training was less successful and the predictive power was lower for models to predict K_d. The COMBINE analysis, as implemented in this work, considers only interaction energies between the ligand and the protein in the bound state. However, predictions for k_{off} and K_d values also depend on the energies of the transition and unbound states, respectively. The transition state is structurally not far from the bound state and can have interactions in common with it, while the bound state is less structurally related to the unbound state
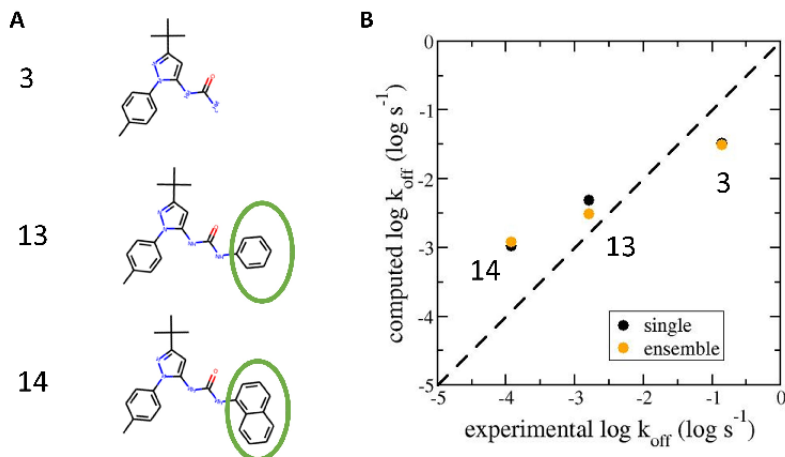
Figure 3: COMBINE analysis models to compute k_{off} values for p38 MAP kinase inhibitors can reproduce the ranking for a small series composed of ligands 13 (training set), 3 and 14 (test set). (A) Chemical structures of the ligands in the series. Green circles indicate the differences among the ligands. (B) Plot of computed versus experimental log k_{off} values for the ligands in the series. COMBINE analysis models were obtained using single structures (two latent variables) or ensemble docking (one latent variable). The straight line corresponds to y = x (ideal case). For ligand 13, k_{off} values were computed from training, while for ligands 3 and 14 k_{off} values were computed from testing.

in which there may be differences in solvation of the unbound ligands that contribute to K_d. COMBINE analysis was originally developed to predict binding affinities for series of related (congeneric) compounds. In the more diverse set of compounds studied here, differences in solvation are likely to be larger and make a larger contribution to the differences in K_d. This contribution is not considered explicitly in the COMBINE analysis described here. The structural similarity between transition and bound states, and the smaller influence of solvation effects may explain why good predictions were obtained for k_{off} values, but not for K_d values for this diverse set of compounds.

The interaction energies associated with the coefficients with the highest absolute values were compared for models used to predict k_{off} or K_d values (table S15). While four out of six interaction energies were different for models predicting k_{off} and K_d values using one structure to represent each ligand-protein complex, only one out of six interaction energies was different when multiple structures were used. This indicates that, in the case of models

22

using one structure to represent each ligand-protein complex, the predictions depend on different interaction energies, and K_d and k_{off} may be tuned independently.

4 Discussion

The present study is the second to apply COMBINE analysis to predict experimental k_{off} values. In the first study,[9] two independent COMBINE analyses were performed to predict k_{off} values using a training set of 53 inhibitors of the heat shock protein 90 (HSP90) and 33 inhibitors of HIV-1 protease, resulting in models with Q^2-LOO values of 0.69 and 0.70, respectively. The present study achieved a higher Q^2-LOO value, 0.75, despite the fact that less crystallographic data was available for the ligand-protein complexes (37/66, 12/33 and 6/33 complexes had crystal structures available for HSP90, HIV-1 protease and p38 MAP kinase, respectively), resulting in greater uncertainty in the modeling. A Q^2-LOO value similar to the previous study was achieved for the model with multiple structures.

Previous studies used MD simulations or regression methods to predict k_{off} values for p38 MAP kinase inhibitors (Table S17). A method combining steered molecular dynamics (SMD) and computation of free energies of dissociation obtained R^2te of 0.88 for a dataset comprised of eight inhibitors.[74] In another study, a model obtained using PLS regression, a training set of 14 inhibitors, a test set of 6 inhibitors and protein-inhibitor interaction fingerprints (IFPs) from simulations of inhibitor dissociation had R^2te of 0.56 and R^2tr of 0.72.[75] Negative binding energy integrals obtained from ligand dissociation using local-scaled MD achieved R^2te of 0.64 in the prediction of k_{off} values for 41 complexes with different kinases, including 12 complexes with p38 MAP kinase.[76] Amongst regression methods, Volsurf, a method which uses descriptors from the structures of inhibitors only, resulted in a model with R^2te of 0.82 and R^2tr of 0.82 for a training set of 18 inhibitors and a test set of 10 inhibitors.[77] A model obtained using random forest, ligand-protein interactions and information about the protein structure as descriptors achieved R^2te of 0.43 for a test set of 28 complexes with p38 MAP

kinase.[78] The study presented here had a R^2tr higher than the method with the highest R^2tr value, Volsurf, and R^2te value lower than the method with the highest predictive power, the method based on SMD. On one hand, the SMD method includes more details about the ligand-protein interactions and the unbinding process. On the other hand, it was tested using only 8 inhibitors. Despite only inhibitor structures being considered, Volsurf had a good fit for the training set. However, regression methods like Volsurf are usually not able to directly identify important protein residues acting in the modulation of inhibitor-protein binding kinetics.

Previous computational and experimental studies also identified inhibitor-protein interactions and inhibitor features that contribute to lower k_{off} values. Van der Waals interactions and hydrogen bonds of inhibitors with Glu71 were identified as important factors for lowering k_{off} values in SKR studies.[25,27] These features were also observed in the COMBINE analyses presented here. Weaker electrostatic interactions contributing to lower k_{off} values is a trend which was identified in a study using IFPs from simulations of inhibitor dissociation.[75] Here, this trend is also observed, as there is a negative weight for the electrostatic interaction between the inhibitors and Lys53, indicating that weaker electrostatic interactions with Lys53 contribute to lower k_{off} values. In the study using IFPs,[75] stronger van der Waals interactions were found to contribute to lower k_{off} values, and interactions with charged residues, such as Arg67 and Arg70, were involved in the modulation of k_{off} values. Interactions with these residues were also identified as modulators of k_{off} values in the COMBINE analyses presented here. Other residues identified by COMBINE analysis, such as Lys53, Glu71 and Leu167, were not identified by the study using IFPs, which may be related to differences in the training set used or in the methods, as the IFPs were obtained from simulations of inhibitor dissociation, while the interaction energies from COMBINE analysis were obtained from ligand-protein complexes in the bound state. Moreover, the contribution of larger molecular volumes and stronger van der Waals interactions to lower k_{off} values, found using Volsurf[77] and IFPs,[75] agrees with the overall results obtained in

COMBINE analyses, since most interaction energies with high weights are van der Waals interactions.

Computational predictions in this work are supported by data for mutations generating drug resistance in EGFR and Abl kinases. In the COMBINE analysis using one structure, interactions with residues Lys53, Arg70, Glu71 and Leu167 were identified as important for long residence times. Additionally, in the COMBINE analysis using an ensemble of structures, interactions with residues Arg67 and Asp168 were highlighted as important for long residence times. Mutations in positions 282[79,80] and 285[80] of Abl kinase (equivalent to positions 67 and 70 in p38 MAP kinase, respectively, Figure S5a) and in position 759 of EGFR[81] (equivalent to position 71 in p38 MAP kinase, Figure S5b) are associated with drug resistance, supporting the importance of residues Arg67, Arg70 and Glu71 in inhibitor binding.

As pointed out in section 3.2, the R^2te value was higher for the model using one structure than for the model using multiple structures (0.79 and 0.72 for the models using one or multiple structures, respectively, Table 1), indicating that the group of interaction energies selected in the model with one structure had a higher predictive power. One of the possible reasons for this is poor binding modes for ligands obtained from ensemble docking (Figure S4). Docking using a protein ensemble led to improved binding modes[24] and a higher area under the receiver operating characteristic (ROC) curve (AUC)[17] in previous studies. Here, there were discrepancies between the reference binding modes (the binding modes in the model using one structure) and the binding modes from ensemble docking for ligands 23, 26 (training set), 21 and 24 (test set), which had from 8/10 up to 10/10 binding modes with RMSD higher than 5 Å from the ligand in the reference structure (Figure S4, Table S16). Additionally, these ligands had low prediction errors in the COMBINE analysis model using one crystal structure (0.06 to 0.38 log s^{-1}), but average (0.55 to 0.58 log s^{-1} for ligands 24 and 26) to high (1.11 to 1.12 log s^{-1} for ligands 21 and 23) errors in the COMBINE analysis model using multiple structures.

Visual inspection of the crystal structures used to obtain the reference binding modes for ligands 21, 23 and 24 showed that a significant structural rearrangement happens for residues Lys53 and Glu71. Moreover, visual inspection of the crystal structure used to obtain the reference binding mode for ligand 26 shows a Val in position 84, while this position is occupied by a bulkier residue, Ile, in p38 MAP kinase. Docking was performed treating residues Lys53 and Glu71 (ligands 21, 23 and 24) or residue Ile 84 (ligand 26) as flexible, but this did not improve significantly the binding modes (Table S16). As the reference binding modes for these ligands are complexes with other kinases, it is possible that these ligands could have different binding modes with p38 MAP kinase, or that other residues of p38 MAP kinase, not explored in flexible docking, may need to reorganize for ligand binding.

Another possible reason for the higher predictive power of the model using single structures, compared to the model using an ensemble to represent each ligand-protein complex, is the diversity of structures. Regarding the disordered region after the DFG motif (residues 170 to 184 in PDB 1KV2), the six crystal structures used in the model with single structures presented a higher structural diversity than the 10 protein structures selected for ensemble docking (Figure S6c). As noted before, the region after the DFG motif was missing in the six crystal structures, and structural diversity came from modeling with Modeller (Section 2.4). The sampling of structures for the ensemble-based model could be improved by the use of enhanced sampling MD simulation techniques but further work would be necessary to investigate whether this results in improved predictive ability.

In the present study, energy minimization of structures prior COMBINE analysis was performed using implicit solvation, and water molecules were not considered during collection of the energy contributions. Modeling the water implicitly, and not by explicit molecules, is an approximation to simplify the model, but it can lead to loss of important water-mediated interactions. However, such interactions may be captured implicitly in the COMBINE model. For instance, in a previous study by Ganotra and Wade,[9] the correlation between low k_{off} values and water-mediated interactions between HIV-1 protease and its inhibitors was cap-

tured implicitly by negative coefficients attributed to residues involved in hydrogen bonding. Among the six crystal structures available for complexes between inhibitors and p38 MAP kinase, in four of them no water-mediated interactions between inhibitor and protein were found. In two of them, two water molecules mediate interactions of the inhibitor with Arg149 (Figure 4). Interestingly, water-mediated protein-protein interactions in the binding site of four crystal structures were found. In these structures, water mediated interactions between the backbone of Asp168 and Lys53 (3 structures, Figure 4) or Glu71 (1 structure). The interaction involving Lys53 can form a lock, hindering inhibitor exit from the binding site, and explaining mechanistically why the electrostatic interaction with Lys53 has a negative coefficient in k_{off} predictions, as suggested by COMBINE analysis. Electrostatic interactions with Lys53 have a negative coefficient. When this negative coefficient is combined with favorable electrostatic interactions, it leads to positive contributions to the predicted log k_{off} values, resulting in high k_{off} values. When the inhibitor interacts with Lys53, it may prevent the formation of the lock between Lys53 and Asp168, resulting in facilitated dissociation and a high k_{off} value.

The current COMBINE analysis model presented here deals with one of the approximations of previous models, the treatment of protein flexibility. However, there are other approximations in the present COMBINE analysis model which could be improved in future work. One of them is the lack of consideration of the transition state ensemble. For instance, inhibitors could have similar bound states but different transition state ensembles, what would lead to different experimental k_{off} values. Describing such an ensemble requires knowledge of the dissociation path, which could make the preparation for COMBINE analysis computationally expensive. Moreover, the description of such an ensemble can be challenging if many dissociation pathways are available. This is the case, for instance, for the T4 lysozyme L99A mutant, for which eight dissociation pathways have been identified.[4] So far, three dissociation pathways were identified for inhibitors of p38 MAP kinase.[76,82,83]

Another assumption of the present COMBINE analysis model is that protein conforma-
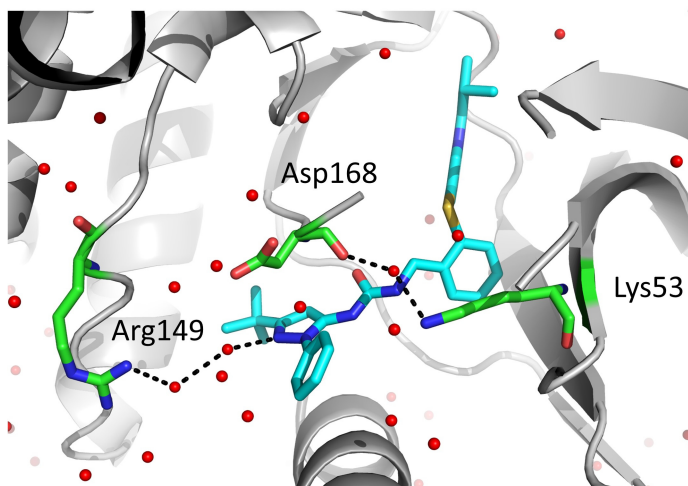
Figure 4: Protein-ligand and protein-protein interactions mediated by water molecules (PDB 2YIW). The water mediated interaction between Asp168 and Lys53 can form a lock, hindering inhibitor dissociation. Ligand 32, Lys53, Arg149 and Asp168 are shown in sticks. Crystallographic waters are shown as red spheres. Hydrogen bonds are shown as black dotted lines, the protein backbone in gray, and the ligand and residues are colored according to atom type, with carbon atoms in cyan (ligand) or green (residues).

tional changes during inhibitor dissociation do not need to be considered explicitly and their effects on the relative dissociation rates of the ligands are accounted for implicitly by considering the protein-ligand interactions in the bound state. Bulkier inhibitors may require larger conformational changes to dissociate and, therefore, show a higher energy penalty for leaving the protein binding site, leading to lower k_{off} values. A previous study[83] sampled dissociation paths for type I and type II inhibitors of p38 MAP kinase and identified by umbrella sampling that the dissociation of type II inhibitors presented larger energy barriers due to the larger protein motions required for inhibitor dissociation, which explained the lower k_{off} values for these inhibitors. The dataset in the present study contained type II inhibitors only, which may have reduced potential problems arising from the lack of explicit consideration of protein conformational changes during dissociation. Further terms can be included in COMBINE analysis, such as energy penalties for changes in the protein conformation or differences in the intramolecular energy terms in bound and unbound protein conformations.[11]

One of the motivations of performing COMBINE analysis to predict k_{off} values or other biological activity is the speed of the parametrization and application of the method, especially for large datasets with tens or hundreds of inhibitors. One of the advantages of using one structure over multiple structures in COMBINE analysis is the lower computational cost, as usually only a single crystal structure and no simulations, just energy minimizations, are required when one structure is used to represent the ligand-protein complex. For the model with multiple structures obtained here, the computational bottleneck of the method was running one MD simulation of 1200 nanoseconds for a ligand-protein complex to obtain the configurational ensemble. This simulation length allowed the collection of more diverse protein structures (RMSD values and protein structures in Figure S6). High root mean squared fluctuations and structural variation close to the DFG motif (Figure S6) suggest flexibility is important for ligand binding to p38 MAP kinase. While docking with Vina took seconds for each ligand-protein pair, running the MD simulation took approximately 6 days using 1 GPU (NVIDIA Tesla P40). Therefore, inclusion of protein and ligand flexibility in COMBINE analysis can be achieved with modest increases in computational time. Moreover, the increase in computational time can be reduced by running a shorter MD simulation, or obtaining a configurational ensemble from other sources, like NMR data, crystal structures or computational methods such as the iterative Anisotropic Network Model (iterANM).[24]

Despite the increase in computational time compared to the model with a single structure, COMBINE analysis using multiple structures for one ligand-protein complex can be useful in cases where flexibility cannot be neglected. Here, a region in p38 MAP kinase close to the binding site was not resolved in all crystal structures of ligand-protein complexes, suggesting this region was highly flexible. The use of COMBINE analysis with multiple structures helped to reveal more residues involved in the modulation of k_{off} values. COMBINE analysis with multiple structures can also be applied in cases where the ligand has multiple binding modes, or in cases where the protein shows multiple conformations or rotamers for residues, especially in the binding site.

5 Conclusions

We propose a method to incorporate protein and ligand flexibility in COMBINE analysis using arithmetic or exponential averaging of the predicted k_{off} values from multiple structures describing one ligand-protein complex. Here, such structures were obtained from docking the ligand to a protein ensemble from MD simulations, but protein ensembles can also be obtained from multiple crystal structures or from NMR. We obtained models for k_{off} values for inhibitors of p38 MAP kinase with good predictive power (R^2te = 0.79, R^2tr = 0.86 and R^2te = 0.72, R^2tr = 0.83 for one and multiple structures, respectively) and identified key ligand-protein interactions that contribute to variance in binding kinetics. COMBINE analysis identified known important residues, such as Arg70 and Glu71, and also additional relevant residues, such as Lys53 and Leu167. Moreover, the incorporation of protein-ligand flexibility helped to highlight more residues connected to long residence times, such as Arg67 and Asp168. These specific interaction energies provide insights into the mechanisms of action of slow and fast dissociating inhibitors.

6 Data and Software Availability

The software to apply ensemble COMBINE analysis is available, with an example (using the structures from ensemble docking describe here), at https://github.com/HITS-MCM/ensembleCOMBINE. The python script for ligand superimposition is available at https://github.com/HITS-MCM/LigSuperposition. The input files and the files with coordinates and topologies of the ligand-protein complexes which were used to obtain the models using single structures or an ensemble of structures to represent each ligand-protein complex can be obtained at: https://doi.org/10.5281/zenodo.4597725.

Acknowledgement

Supporting Information Available

Supporting Information is available free of charge on the ACS Publications website. The supporting information (PDF file) contains additional information about the inhibitors and COMBINE models tested (Figures S1-S6, Tables S1-S17, Sections S1 and S2).

References

(1) Copeland, R. A. The drug-target residence time model: a 10-year retrospective. *Nat. Rev. Drug Discov.* **2016**, *15*, 87–95.

(2) Schuetz, D. A.; de Witte, W. E. A.; Wong, Y. C.; Knasmueller, B.; Richter, L.; Kokh, D. B.; Sadiq, S. K.; Bosma, R.; Nederpelt, I.; Heitman, L. H.; Segala, E.; Amaral, M.; Guo, D.; Andres, D.; Georgi, V.; Stoddart, L. A.; Hill, S.; Cooke, R. M.; Graaf, C. D.; Leurs, R.; Frech, M.; Wade, R. C.; de Lange, E. C. M.; IJzerman, A. P.; Müller-Fahrnow, A.; Ecker, G. F. Kinetics for Drug Discovery: an industry-driven effort to target drug residence time. *Drug Discov. Today* **2017**, *22*, 896–911.

(3) Bernetti, M.; Cavalli, A.; Mollica, L. Protein-ligand (un)binding kinetics as a new

paradigm for drug discovery at the crossroad between experiments and modelling. *Med. Chem. Commun.* **2017**, *8*, 534–550.

(4) Nunes-Alves, A.; Kokh, D. B.; Wade, R. C. Recent progress in molecular simulation methods for drug binding kinetics. *Curr. Opin. Struct. Biol.* **2020**, *64*, 126–133.

(5) Kokh, D. B.; Amaral, M.; Bomke, J.; Grädler, U.; Musil, D.; Buchstaller, H.-P.; Dreyer, M. K.; Frech, M.; Lowinski, M.; Vallee, F.; Bianciotto, M.; Rak, A.; Wade, R. C. Estimation of Drug-Target Residence Times by τ-Random Acceleration Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2018**, *14*, 3859–3869.

(6) Mollica, L.; Decherchi, S.; Zia, S. R.; Gaspari, R.; Cavalli, A.; Rocchia, W. Kinetics of protein-ligand unbinding via smoothed potential molecular dynamics simulations. *Sci. Rep.* **2015**, *5*, 11539.

(7) Huber, G. A.; Kim, S. Weighted–ensemble Brownian dynamics simulations for protein association reactions. *Biophys. J.* **1996**, *70*, 97–110.

(8) Zuckerman, D. M.; Chong, L. T. Weighted Ensemble Simulation: Review of Methodology, Applications, and Software. *Annu. Rev. Biophys.* **2017**, *46*, 43–57.

(9) Ganotra, G. K.; Wade, R. C. Prediction of Drug-Target Binding Kinetics by Comparative Binding Energy Analysis. *ACS Med. Chem. Lett.* **2018**, *9*, 1134–1139.

(10) Qu, S.; Huang, S.; Pan, X.; Yang, L.; Mei, H. Constructing Interconsistent, Reasonable, and Predictive Models for Both the Kinetic and Thermodynamic Properties of HIV-1 Protease Inhibitors. *J. Chem. Inf. Model.* **2016**, *56*, 2061–2068.

(11) Ortiz, A. R.; Pisabarro, M. T.; Gago, F.; Wade, R. C. Prediction of Drug Binding Affinities by Comparative Binding Energy Analysis. *J. Med. Chem.* **1995**, *38*, 2681–2691.

(12) Pérez, C.; Pastor, M.; Ortiz, A. R.; Gago, F. Comparative Binding Energy Analysis of HIV-1 Protease Inhibitors: Incorporation of Solvent Effects and Validation as a Powerful Tool in Receptor-Based Drug Design. *J. Med. Chem.* **1998**, *41*, 836–852.

(13) Peters, M. B.; Merz, K. M. Semiempirical Comparative Binding Energy Analysis (SE-COMBINE) of a Series of Trypsin Inhibitors. *J. Chem. Theory Comput.* **2006**, *2*, 383–399.

(14) Liu, S.; Fu, R.; Cheng, X.; Chen, S.-P.; Zhou, L.-H. Exploring the binding of BACE-1 inhibitors using comparative binding energy analysis (COMBINE). *BMC Struct. Biol.* **2012**, *12*, 21.

(15) Wang, T.; Wade, R. C. Comparative Binding Energy (COMBINE) Analysis of OppA-Peptide Complexes to Relate Structure to Binding Thermodynamics. *J. Med. Chem.* **2002**, *45*, 4828–4837.

(16) Tomić, S.; Bertoša, B.; Wang, T.; Wade, R. C. COMBINE analysis of the specificity of binding of Ras proteins to their effectors. *Proteins* **2007**, *67*, 435–447.

(17) Cleves, A. E.; Jain, A. N. Structure- and Ligand-Based Virtual Screening on DUD-E+: Performance Dependence on Approximations to the Binding Pocket. *J. Chem. Inf. Model.* **2020**, *60*, 4296–4310.

(18) Arantes, G. M. Flexibility and inhibitor binding in Cdc25 phosphatases. *Proteins* **2010**, *78*, 3017–3032.

(19) Nunes-Alves, A.; Arantes, G. M. Ligand–receptor affinities computed by an adapted linear interaction model for continuum electrostatics and by protein conformational averaging. *J. Chem. Inf. Model.* **2014**, *54*, 2309–2319.

(20) Bhattarai, A.; Wang, J.; Miao, Y. Retrospective ensemble docking of allosteric modu-

lators in an adenosine G-protein-coupled receptor. *Biochim. Biophys. Acta Gen. Subj.* **2020**, *1864*, 129615.

(21) Olubiyi, O. O.; Olagunju, M.; Keutmann, M.; Loschwitz, J.; Strodel, B. High Through-put Virtual Screening to Discover Inhibitors of the Main Protease of the Coronavirus SARS-CoV-2. *Molecules* **2020**, *25*, 3193.

(22) Noé, F.; Olsson, S.; Köhler, J.; Wu, H. Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning. *Science* **2019**, *365*.

(23) Degiacomi, M. T. Coupling Molecular Dynamics and Deep Learning to Mine Protein Conformational Space. *Structure* **2019**, *27*, 1034–1040.e3.

(24) Wang, A.; Zhang, Y.; Chu, H.; Liao, C.; Zhang, Z.; Li, G. Higher Accuracy Achieved for Protein-Ligand Binding Pose Prediction by Elastic Network Model-Based Ensemble Docking. *J. Chem. Inf. Model.* **2020**, *60*, 2939–2950.

(25) Pargellis, C.; Tong, L.; Churchill, L.; Cirillo, P. F.; Gilmore, T.; Graham, A. G.; Grob, P. M.; Hickey, E. R.; Moss, N.; Pav, S.; Regan, J. Inhibition of p38 MAP kinase by utilizing a novel allosteric binding site. *Nat. Struct. Mol. Biol.* **2002**, *9*, 268–272.

(26) Wu, P.; Nielsen, T. E.; Clausen, M. H. FDA-approved small-molecule kinase inhibitors. *Trends Pharmacol. Sci.* **2015**, *36*, 422–439.

(27) Regan, J.; Pargellis, C. A.; Cirillo, P. F.; Gilmore, T.; Hickey, E. R.; Peet, G. W.; Proto, A.; Swinamer, A.; Moss, N. The kinetics of binding to p38 MAP kinase by analogues of BIRB 796. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 3101–3104.

(28) Millan, D. S.; Bunnage, M. E.; Burrows, J. L.; Butcher, K. J.; Dodd, P. G.; Evans, T. J.; Fairman, D. A.; Hughes, S. J.; Kilty, I. C.; Lemaitre, A.; Lewthwaite, R. A.; Mahnke, A.; Mathias, J. P.; Philip, J.; Smith, R. T.; Stefaniak, M. H.; Yeadon, M.; Phillips, C.

Design and Synthesis of Inhaled p38 Inhibitors for the Treatment of Chronic Obstructive Pulmonary Disease. *J. Med. Chem.* **2011**, *54*, 7797–7814.

(29) Redhead, M.; Satchell, R.; Morkūnaitė, V.; Swift, D.; Petrauskas, V.; Golding, E.; Onions, S.; Matulis, D.; Unitt, J. A combinatorial biophysical approach; FTSA and SPR for identifying small molecule ligands and PAINs. *Anal. Biochem.* **2015**, *479*, 63–73.

(30) Georgi, V.; Schiele, F.; Berger, B.-T.; Steffen, A.; Marin Zapata, P. A.; Briem, H.; Menz, S.; Preusse, C.; Vasta, J. D.; Robers, M. B.; Brands, M.; Knapp, S.; Fernández-Montalván, A. Binding Kinetics Survey of the Drugged Kinome. *J. Am. Chem. Soc.* **2018**, *140*, 15774–15782.

(31) Gil-Redondo, R.; Klett, J.; Gago, F.; Morreale, A. gCOMBINE: A graphical user interface to perform structure-based comparative binding energy (COMBINE) analysis on a set of ligand-receptor complexes. *Proteins* **2010**, *78*, 162–172.

(32) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.

(33) Rännar, S.; Lindgren, F.; Geladi, P.; Wold, S. A PLS kernel algorithm for data sets with many variables and fewer objects. Part 1: Theory and algorithm. *J. Chemometrics* **1994**, *8*, 111–125.

(34) Minh, D. D. L. Implicit ligand theory: rigorous binding free energies and thermodynamic expectations from molecular docking. *J. Chem. Phys.* **2012**, *137*, 104106.

(35) Klüter, S.; Grütter, C.; Naqvi, T.; Rabiller, M.; Simard, J. R.; Pawar, V.; Getlik, M.; Rauh, D. Displacement Assay for the Detection of Stabilizers of Inactive Kinase Conformations. *J. Med. Chem.* **2010**, *53*, 357–367.

(36) Namboodiri, H. V.; Bukhtiyarova, M.; Ramcharan, J.; Karpusas, M.; Lee, Y.; Spring-
man, E. B. Analysis of Imatinib and Sorafenib Binding to p38α Compared with c-Abl
and b-Raf Provides Structural Insights for Understanding the Selectivity of Inhibitors
Targeting the DFG-Out Form of Protein Kinases. *Biochemistry* **2010**, *49*, 3611–3618.

(37) Šali, A.; Blundell, T. L. Comparative Protein Modelling by Satisfaction of Spatial
Restraints. *J. Mol. Biol.* **1993**, *234*, 779–815.

(38) Gill, A. L.; Frederickson, M.; Cleasby, A.; Woodhead, S. J.; Carr, M. G.; Wood-
head, A. J.; Walker, M. T.; Congreve, M. S.; Devine, L. A.; Tisi, D.; O'Reilly, M.;
Seavers, L. C. A.; Davis, D. J.; Curry, J.; Anthony, R.; Padova, A.; Murray, C. W.;
Carr, R. A. E.; Jhoti, H. Identification of Novel p38α MAP Kinase Inhibitors Using
Fragment-Based Lead Generation. *J. Med. Chem.* **2005**, *48*, 414–426.

(39) Dolinsky, T. J.; Nielsen, J. E.; McCammon, J. A.; Baker, N. A. PDB2PQR: an auto-
mated pipeline for the setup of Poisson–Boltzmann electrostatics calculations. *Nucleic
Acids Res.* **2004**, *32*, W665–W667.

(40) Unni, S.; Huang, Y.; Hanson, R. M.; Tobias, M.; Krishnan, S.; Li, W. W.; Nielsen, J. E.;
Baker, N. A. Web servers and services for electrostatics calculations with APBS and
PDB2PQR. *J. Comput. Chem.* **2011**, *32*, 1488–1491.

(41) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchi-
son, G. R. Open Babel: an open chemical toolbox. *J. Cheminform.* **2011**, *3*, 33.

(42) van Linden, O. P. J.; Kooistra, A. J.; Leurs, R.; de Esch, I. J. P.; de Graaf, C. KLIFS: A
Knowledge-Based Structural Database To Navigate Kinase–Ligand Interaction Space.
J. Med. Chem. **2014**, *57*, 249–277.

(43) Kooistra, A. J.; Kanev, G. K.; van Linden, O. P.; Leurs, R.; de Esch, I. J.; de Graaf, C.
KLIFS: a structural kinase-ligand interaction database. *Nucleic Acids Res.* **2015**, *44*,
D365–D371.

(44) Karoulia, Z.; Wu, Y.; Ahmed, T. A.; Xin, Q.; Bollard, J.; Krepler, C.; Wu, X.; Zhang, C.; Bollag, G.; Herlyn, M.; Fagin, J. A.; Lujambio, A.; Gavathiotis, E.; Poulikakos, P. I. An Integrated Model of RAF Inhibitor Action Predicts Inhibitor Activity against Oncogenic BRAF Signaling. *Cancer Cell* **2016**, *30*, 485–498.

(45) Schroeder, G. M.; An, Y.; Cai, Z.-W.; Chen, X.-T.; Clark, C.; Cornelius, L. A. M.; Dai, J.; Gullo-Brown, J.; Gupta, A.; Henley, B.; Hunt, J. T.; Jeyaseelan, R.; Kamath, A.; Kim, K.; Lippy, J.; Lombardo, L. J.; Manne, V.; Oppenheimer, S.; Sack, J. S.; Schmidt, R. J.; Shen, G.; Stefanski, K.; Tokarski, J. S.; Trainor, G. L.; Wautlet, B. S.; Wei, D.; Williams, D. K.; Zhang, Y.; Zhang, Y.; Fargnoli, J.; Borzilleri, R. M. Discovery of N-(4-(2-Amino-3-chloropyridin-4-yloxy)-3-fluorophenyl)-4-ethoxy-1-(4-fluorophenyl)-2-oxo-1,2-dihydropyridine-3-carboxamide (BMS-777607), a Selective and Orally Efficacious Inhibitor of the Met Kinase Superfamily. *J. Med. Chem.* **2009**, *52*, 1251–1254.

(46) Chan, W. W.; Wise, S. C.; Kaufman, M. D.; Ahn, Y. M.; Ensinger, C. L.; Haack, T.; Hood, M. M.; Jones, J.; Lord, J. W.; Lu, W. P.; Miller, D.; Patt, W. C.; Smith, B. D.; Petillo, P. A.; Rutkoski, T. J.; Telikepalli, H.; Vogeti, L.; Yao, T.; Chun, L.; Clark, R.; Evangelista, P.; Gavrilescu, L. C.; Lazarides, K.; Zaleskas, V. M.; Stewart, L. J.; Etten, R. A. V.; Flynn, D. L. Conformational control inhibition of the BCR-ABL1 tyrosine kinase, including the gatekeeper T315I mutant, by the switch-control inhibitor DCC-2036. *Cancer Cell* **2011**, *19*, 556–568.

(47) Qian, F.; Engst, S.; Yamaguchi, K.; Yu, P.; Won, K.-A.; Mock, L.; Lou, T.; Tan, J.; Li, C.; Tam, D.; Lougheed, J.; Yakes, F. M.; Bentzien, F.; Xu, W.; Zaks, T.; Wooster, R.; Greshock, J.; Joly, A. H. Inhibition of Tumor Cell Growth, Invasion, and Metastasis by EXEL-2880 (XL880, GSK1363089), a Novel Inhibitor of HGF and VEGF Receptor Tyrosine Kinases. *Cancer Res.* **2009**, *69*, 8009–8016.

(48) Heinzlmeir, S.; Kudlinzki, D.; Sreeramulu, S.; Klaeger, S.; Gande, S. L.; Linhard, V.;

Wilhelm, M.; Qiao, H.; Helm, D.; Ruprecht, B.; Saxena, K.; Médard, G.; Schwalbe, H.; Kuster, B. Chemical Proteomics and Structural Biology Define EPHA2 Inhibition by Clinical Kinase Drugs. *ACS Chem. Biol.* **2016**, *11*, 3400–3411.

(49) Tröster, A.; Heinzlmeir, S.; Berger, B.-T.; Gande, S. L.; Saxena, K.; Sreeramulu, S.; Linhard, V.; Nasiri, A. H.; Bolte, M.; Müller, S.; Kuster, B.; Médard, G.; Kudlinzki, D.; Schwalbe, H. NVP-BHG712: Effects of Regioisomers on the Affinity and Selectivity toward the EPHrin Family. *ChemMedChem* **2018**, *13*, 1629–1633.

(50) Garner, A. P.; Gozgit, J. M.; Anjum, R.; Vodala, S.; Schrock, A.; Zhou, T.; Serrano, C.; Eilers, G.; Zhu, M.; Ketzer, J.; Wardwell, S.; Ning, Y.; Song, Y.; Kohlmann, A.; Wang, F.; Clackson, T.; Heinrich, M. C.; Fletcher, J. A.; Bauer, S.; Rivera, V. M. Ponatinib Inhibits Polyclonal Drug-Resistant KIT Oncoproteins and Shows Therapeutic Potential in Heavily Pretreated Gastrointestinal Stromal Tumor (GIST) Patients. *Clin. Cancer Res.* **2014**, *20*, 5745–5755.

(51) Smith, C. C.; Zhang, C.; Lin, K. C.; Lasater, E. A.; Zhang, Y.; Massi, E.; Damon, L. E.; Pendleton, M.; Bashir, A.; Sebra, R.; Perl, A.; Kasarskis, A.; Shellooe, R.; Tsang, G.; Carias, H.; Powell, B.; Burton, E. A.; Matusow, B.; Zhang, J.; Spevak, W.; Ibrahim, P. N.; Le, M. H.; Hsu, H. H.; Habets, G.; West, B. L.; Bollag, G.; Shah, N. P. Characterizing and Overriding the Structural Mechanism of the Quizartinib-Resistant FLT3 "Gatekeeper" F691L Mutation with PLX3397. *Cancer Discov.* **2015**, *5*, 668–679.

(52) McTigue, M.; Murray, B. W.; Chen, J. H.; Deng, Y.-L.; Solowiej, J.; Kania, R. S. Molecular conformations, interactions, and properties associated with drug efficiency and clinical performance among VEGFR TK inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 18281–18289.

(53) RDKit: Open-Source Cheminformatics and Machine Learning, http://rdkit.org/.

(54) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: a

toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* **2011**, *32*, 2319–2327.

(55) Gowers, R. J.; Linke, M.; Barnoud, J.; Reddy, T. J. E.; Melo, M. N.; Seyler, S. L.; Domański, J.; Dotson, D. L.; Buchoux, S.; Kenney, I. M.; Beckstein, O. MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. Proceedings of the 15th Python in Science Conference. 2016; pp 98–105.

(56) Velázquez-Libera, J. L.; Durán-Verdugo, F.; Valdés-Jiménez, A.; Núñez-Vivanco, G.; Caballero, J. LigRMSD: a web server for automatic structure matching and RMSD calculations among identical and similar compounds in protein-ligand docking. *Bioinformatics* **2020**, *36*, 2912–2914.

(57) Schaftenaar, G.; Noordik, J. H. Molden: a pre– and post–processing program for molecular and electronic structures. *J. Comput.–Aided Mol. Des.* **2000**, *14*, 123–134.

(58) Case, D. A.; Betz, R. M.; Cerutti, D. S.; III, T. E. C.; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Homeyer, N.; Izadi, S.; Janowski, P.; Kaus, J.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Luchko, T.; Luo, R.; Madej, B.; Mermelstein, D.; Merz, K. M.; Monard, G.; Nguyen, H.; Nguyen, H. T.; Omelyan, I.; Onufriev, A.; Roe, D. R.; Roitberg, A.; Sagui, C.; Simmerling, C. L.; Botello-Smith, W. M.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Xiao, L.; Kollman, P. A. AMBER 2016, University of California, San Francisco. 2016.

(59) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.

(60) Trott, O.; Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010**, *31*, 455–461.

(61) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* **1993**, *97*, 10269–10280.

(62) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A. Application of RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation. *J. Am. Chem. Soc.* **1993**, *115*, 9620–9631.

(63) Gordon, M. S.; Schmidt, M. W. In *Theory and Applications of Computational Chemistry: the first forty years*; Dykstra, C. E., Frenking, G., Kim, K. S., Scuseria, G. E., Eds.; Elsevier: Amsterdam, 2005; pp 1167–1189.

(64) Mongan, J.; Simmerling, C.; McCammon, J. A.; Case, D. A.; Onufriev, A. Generalized Born Model with a Simple, Robust Molecular Volume Correction. *J. Chem. Theory Comput.* **2007**, *3*, 156–169.

(65) Weiser, J.; Shenkin, P. S.; Still, W. C. Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO). *J. Comput. Chem.* **1999**, *20*, 217–230.

(66) Sitkoff, D.; Sharp, K. A.; Honig, B. Accurate Calculation of Hydration Free Energies Using Macroscopic Solvent Models. *J. Phys. Chem.* **1994**, *98*, 1978–1988.

(67) Hopkins, C. W.; Le Grand, S.; Walker, R. C.; Roitberg, A. E. Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *J. Chem. Theory Comput.* **2015**, *11*, 1864–1874.

(68) 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*, 926–935.

(69) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R.

Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(70) Consortium, T. U. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **2019**, *47*, D506–D515.

(71) Needleman, S. B.; Wunsch, C. D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* **1970**, *48*, 443–453.

(72) Madeira, F.; Park, Y. M.; Lee, J.; Buso, N.; Gur, T.; Madhusoodanan, N.; Basutkar, P.; Tivey, A. R. N.; Potter, S. C.; Finn, R. D.; Lopez, R. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **2019**, *47*, W636–W641.

(73) Guterres, H.; Im, W. Improving Protein-Ligand Docking Results with High-Throughput Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2020**, *60*, 2189–2198.

(74) Braka, A.; Garnier, N.; Bonnet, P.; Aci-Sèche, S. Residence Time Prediction of Type 1 and 2 Kinase Inhibitors from Unbinding Simulations. *J. Chem. Inf. Model.* **2020**, *60*, 342–348.

(75) Zhang, D.; Huang, S.; Mei, H.; Kevin, M.; Shi, T.; Chen, L. Protein-ligand interaction fingerprints for accurate prediction of dissociation rates of p38 MAPK Type II inhibitors. *Integr. Biol.* **2019**, *11*, 53–60.

(76) Du, Y.; Wang, R. Revealing the Unbinding Kinetics and Mechanism of Type I and Type II Protein Kinase Inhibitors by Local-Scaled Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2020**, *16*, 6620–6632.

(77) Huang, S.; Chen, L.; Mei, H.; Zhang, D.; Shi, T.; Kuang, Z.; Heng, Y.; Xu, L.; Pan, X. In Silico Prediction of the Dissociation Rate Constants of Small Chemical Ligands by 3D-Grid-Based VolSurf Method. *Int. J. Mol. Sci.* **2020**, *21*.

(78) Amangeldiuly, N.; Karlov, D.; Fedorov, M. V. Baseline Model for Predicting Protein–Ligand Unbinding Kinetics through Machine Learning. *J. Chem. Inf. Model.* **2020**, *60*, 5946–5956.

(79) Azam, M.; Latek, R. R.; Daley, G. Q. Mechanisms of autoinhibition and STI-571/imatinib resistance revealed by mutagenesis of BCR-ABL. *Cell* **2003**, *112*, 831–843.

(80) Ray, A.; Cowan-Jacob, S. W.; Manley, P. W.; Mestan, J.; Griffin, J. D. Identification of BCR-ABL point mutations conferring resistance to the Abl kinase inhibitor AMN107 (nilotinib) by a random mutagenesis study. *Blood* **2007**, *109*, 5011–5015.

(81) Hu, R.; Xu, H.; Jia, P.; Zhao, Z. KinaseMD: kinase mutations and drug response database. *Nucleic Acids Res.* **2021**, *49*, D552–D561.

(82) Sun, H.; Tian, S.; Zhou, S.; Li, Y.; Li, D.; Xu, L.; Shen, M.; Pan, P.; Hou, T. Revealing the favorable dissociation pathway of type II kinase inhibitors via enhanced sampling simulations and two-end-state calculations. *Sci. Rep.* **2015**, *5*, 8457.

(83) You, W.; Chang, C. A. Role of Molecular Interactions and Protein Rearrangement in the Dissociation Kinetics of p38α MAP Kinase Type-I/II/III Inhibitors. *J. Chem. Inf. Model.* **2018**, *58*, 968–981.

Prediction of Drug-Target Binding Kinetics for Flexible Proteins by Comparative Binding Energy Analysis

Ariane Nunes-Alves[1,2,‡]**, Fabian Ormersbach**[1]**, and Rebecca C. Wade**[1,2,3,‡]

[1]*Heidelberg Institute for Theoretical Studies (HITS), Schloß-Wolfsbrunnenweg 35, 69118*

Heidelberg, Germany

[2]*Center for Molecular Biology (ZMBH), DKFZ-ZMBH Alliance, Heidelberg University, Im*

Neuenheimer Feld 282, 69120 Heidelberg, Germany

[3]*Interdisciplinary Center for Scientific Computing (IWR), Heidelberg University, Im*

Neuenheimer Feld 205, Heidelberg, Germany

[‡]ariane.nunes-alves@h-its.org; rebecca.wade@h-its.org

Supporting Information
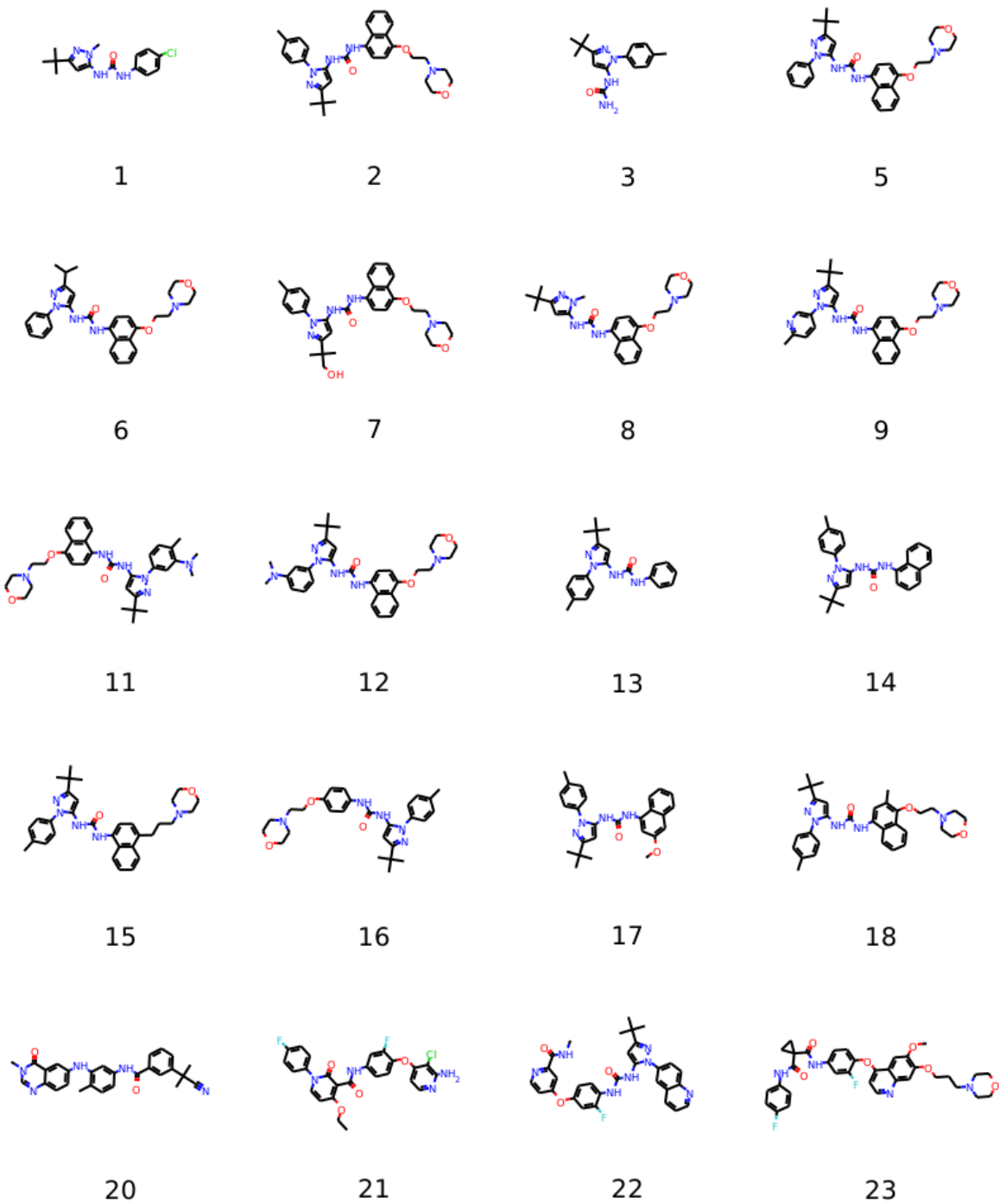
Figure S1: Chemical structures of the 33 inhibitors of p38 MAP kinase used for COMBINE analysis.
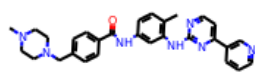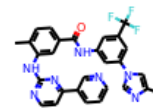
24   25   26   27
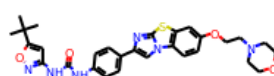
28   29   30   31
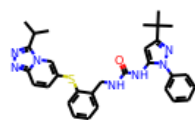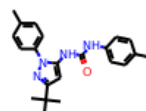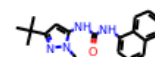
32   33   34   35
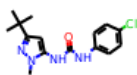
36

Figure S1: Continuation.
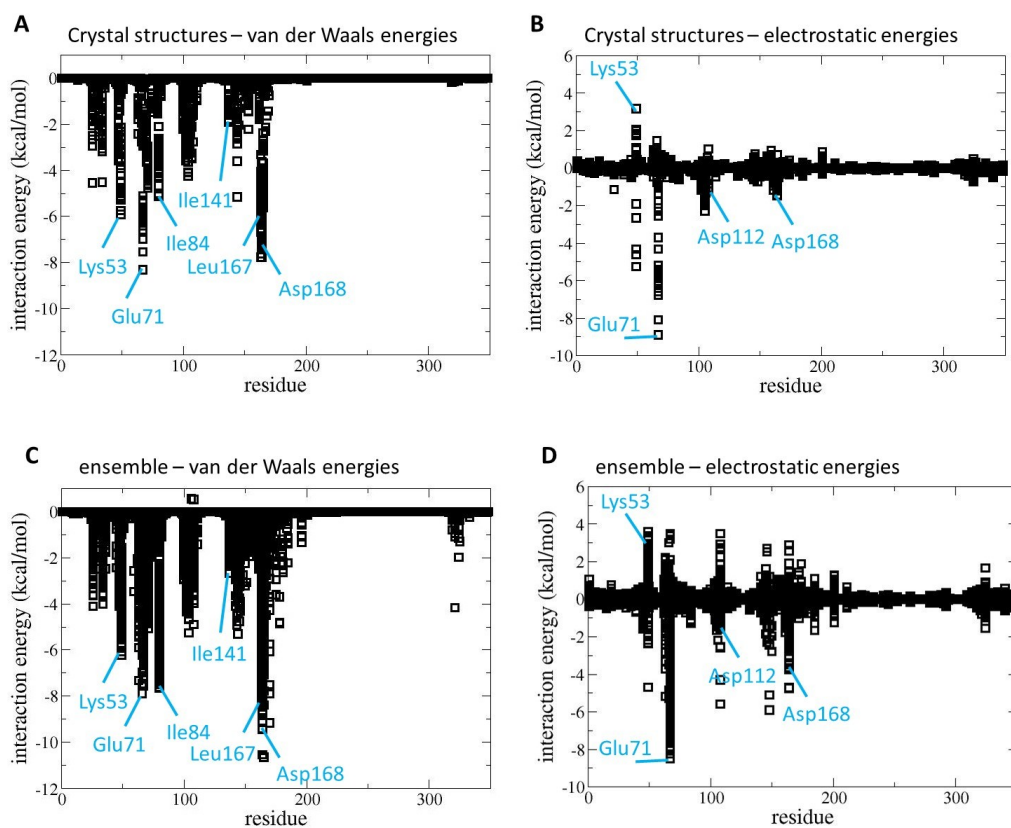
Figure S2: Interaction energies between p38 MAP kinase amino acid residues and inhibitors. Van der Waals (a, c) and electrostatic (b, d) interaction energies were computed between inhibitors and the 348 residues in p38 MAP kinase using the COMBINE program. Each column has 33 (a, b) data points corresponding to the 33 inhibitors used for COMBINE analysis applied to single crystal structures or 330 data points (c, d) corresponding to the 330 ligand-protein complexes (10 complexes per ligand) used for COMBINE analysis applied to structures from ensemble docking. Selected residues are labelled.
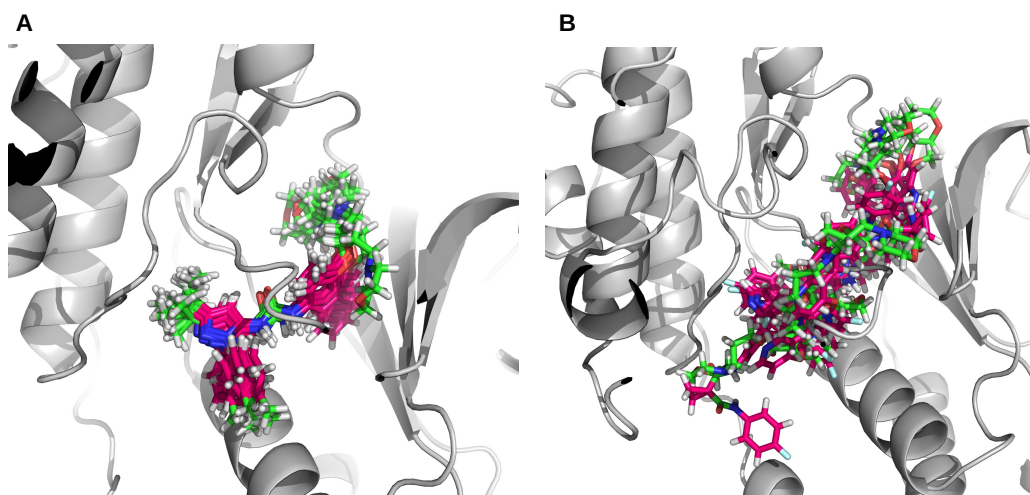
Figure S3: Ten binding modes obtained from ensemble docking for ligands 2 (a) and 23 (b). Protein backbone in gray (PDB 1KV2), ligand colored according to atom type. The binding modes were obtained from docking to ten different protein configurations using AutoDock Vina.
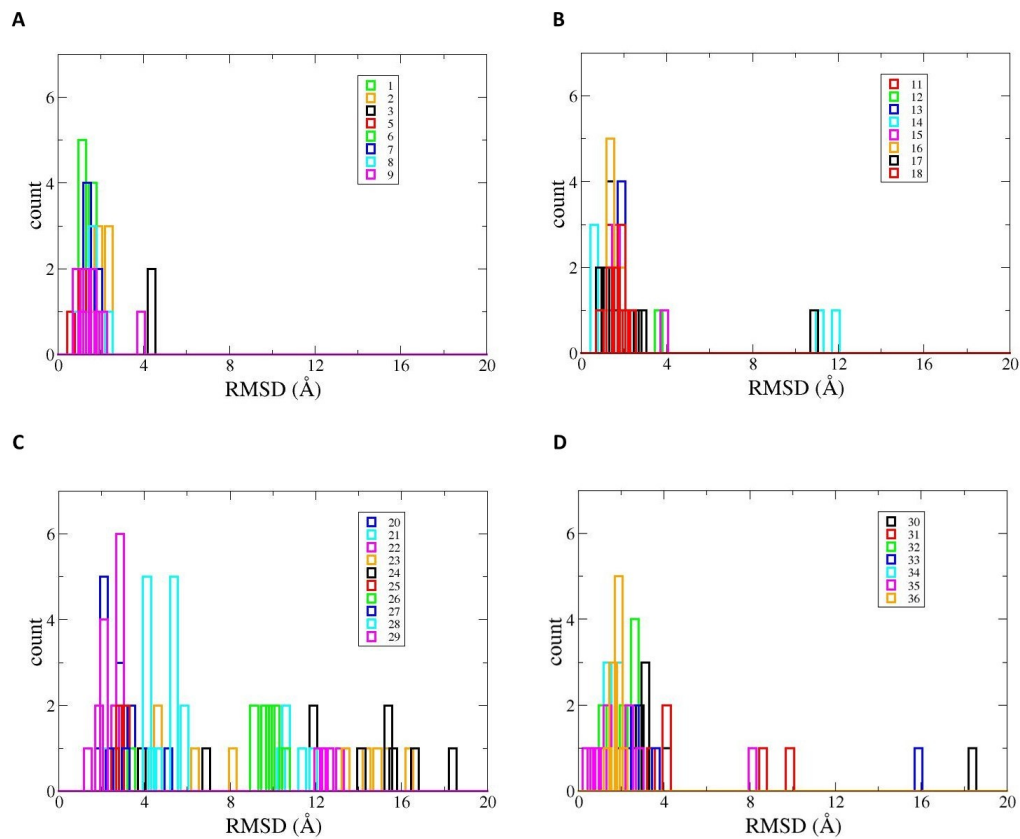
Figure S4: Histograms of root mean squared deviations (RMSD) between the ligand position in the crystal structure and the ligand position in the 10 complexes obtained from ensemble docking. (a) Ligands 1 to 9. (b) Ligands 11 to 18. (c) Ligands 20 to 29. (d) Ligands 30 to 36. Only heavy atoms were considered for RMSD calculation.

A

```
p38    33 GAYGSVCAAFDTKTGLRVAVKKLSRPFQSIIHAKRTYRELRLLKHMKHEN     82
          |.||.|......|..|.|||||.|.    :..:..:...:|..::|.:||.|
Abl   251 GQYGEVYEGVWKKYSLTVAVKTLK---EDTMEVEEFLKEAAVMKEIKHPN    297
```

B

```
p38    54 KLSRPFQSIIHAKRTYRELRLLKHMKHENVIGLLDVFTPARSLEEFNDVY    103
          :|  |...|....|....|..::..:.:.:|..||.:...      :.|.
EGFR  746 EL-REATSPKANKEILDEAYVMASVDNPHVCRLLGICLT-------STVQ    787
```

Figure S5: Alignment of p38 MAP kinase with the human Abl kinase (a) or with the human EGFR kinase (b). (a) Positions 67 and 70 of p38 MAP kinase are highlighted in bold. Mutations in the equivalent positions in Abl kinase are associated with drug resistance. (b) Position 71 of p38 MAP kinase is highlighted in bold. Mutations in the equivalent position in EGFR kinase are associated with drug resistance.
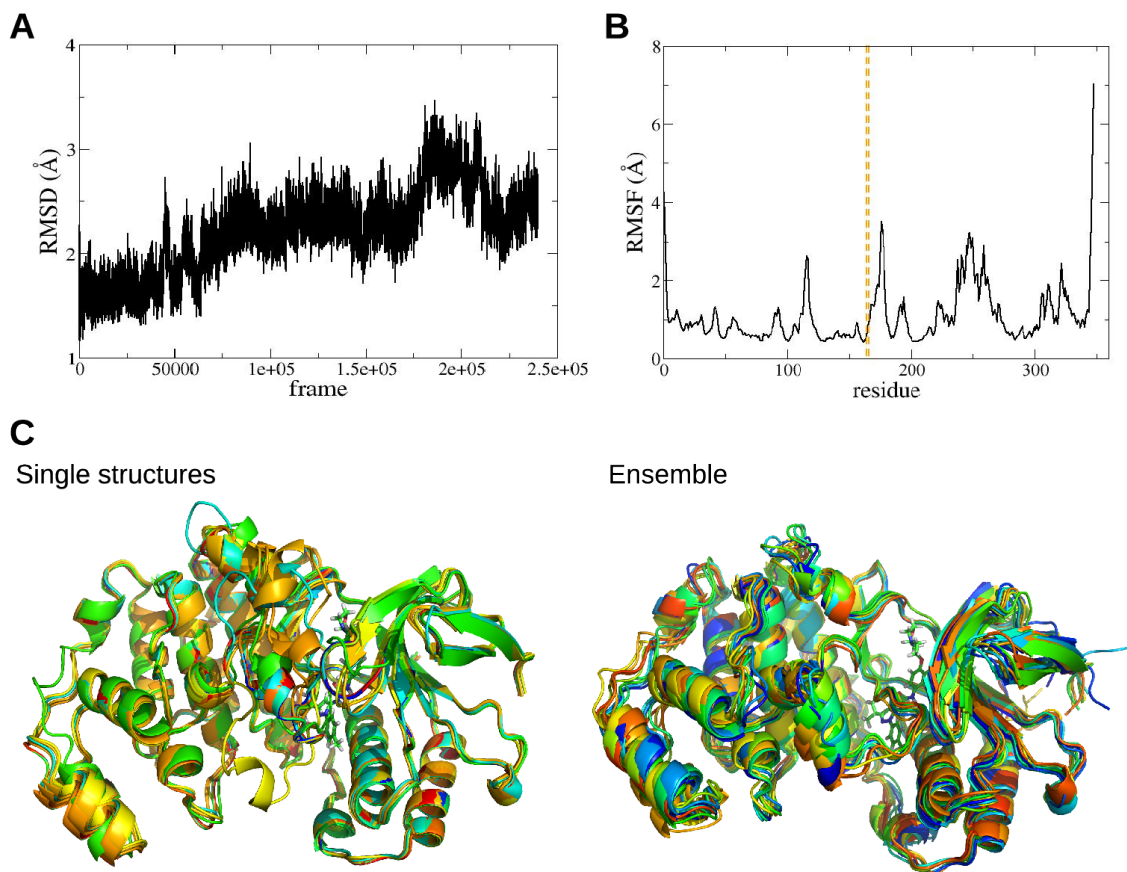
Figure S6: Structural variation of proteins in the models using single structures or an ensemble of structures to represent each ligand-protein complex. (a) Root mean squared deviation (RMSD) of p38 MAP kinase along the trajectory performed to obtain the ensemble. Only heavy atoms of the protein backbone were considered for RMSD calculation. (b) Root mean squared fluctuation (RMSF) of p38 MAP kinase in the trajectory performed to obtain the ensemble. Only Cα atoms were considered for RMSF calculation. Orange dashed lines highlight the position of the DFG motif. (c) Protein structures in the model using single structures or an ensemble to represent each ligand-protein complex. Ligand 16 is represented as sticks to show the position of the binding site.

Table S1: List of the SMILES strings for the p38 MAP kinase inhibitors.

inhibitor ID	SMILES
1	C(=O)(Nc1ccc(Cl)cc1)Nc1cc(nn1C)C(C)(C)C
2	C(=O)(Nc1c2c(c(cc1)OCCN1CCOCC1)cccc2)Nc1cc(nn1c1ccc(cc1)C)C(C)(C)C
3	c1c(ccc(c1)C)n1c(cc(n1)C(C)(C)C)NC(=O)N
5	C1N(CCOC1)CCOc1c2c(c(cc1)NC(=O)Nc1cc(nn1c1ccccc1)C(C)(C)C)cccc2
6	C1N(CCOC1)CCOc1c2c(c(cc1)NC(=O)Nc1cc(nn1c1ccccc1)C(C)C)cccc2
7	c1c(ccc(c1)C)n1c(cc(n1)C(C)(C)CO)NC(=O)Nc1c2c(c(cc1)OCCN1CCOCC1)cccc2
8	C1N(CCOC1)CCOc1c2c(c(cc1)NC(=O)Nc1cc(nn1C)C(C)(C)C)cccc2
9	C1N(CCOC1)CCOc1c2c(c(cc1)NC(=O)Nc1cc(nn1c1cnc(cc1)C)C(C)(C)C)cccc2
11	C1N(CCOC1)CCOc1c2c(c(cc1)NC(=O)Nc1cc(nn1c1cc(c(cc1)C)N(C)C)C(C)(C)C)cccc2
12	C1N(CCOC1)CCOc1c2c(c(cc1)NC(=O)Nc1cc(nn1c1cc(ccc1)N(C)C)C(C)(C)C)cccc2
13	c1cc(ccc1)NC(=O)Nc1n(nc(c1)C(C)(C)C)c1ccc(cc1)C
14	c1c(ccc(c1)C)n1c(cc(n1)C(C)(C)C)NC(=O)Nc1c2c(ccc1)cccc2
15	C1N(CCOC1)CCCc1c2c(c(cc1)NC(=O)Nc1cc(nn1c1ccc(cc1)C)C(C)(C)C)cccc2
16	C1COCCN1CCOc1ccc(cc1)NC(=O)Nc1cc(C(C)(C)C)nn1c1ccc(C)cc1
17	c1c2c(ccc1)c(cc(c2)OC)NC(=O)Nc1n(nc(c1)C(C)(C)C)c1ccc(cc1)C
18	C1N(CCOC1)CCOc1c2c(c(cc1C)NC(=O)Nc1cc(nn1c1ccc(cc1)C)C(C)(C)C)cccc2
20	c12c(ccc(c1)Nc1c(ccc(c1)NC(=O)c1cc(ccc1)C(C#N)(C)C)C)ncn(c2=O)C
21	c1cc(ccc1n1ccc(c(c1=O)C(=O)Nc1ccc(c(c1)F)Oc1ccnc(c1Cl)N)OCC)F
22	c1(ccc2c(c1)cccn2)n1c(cc(n1)C(C)(C)C)NC(=O)Nc1ccc(cc1F)Oc1cc(ncc1)C(=O)NC
23	c1(c(cc2c(c1)c(ccn2)Oc1ccc(cc1F)NC(=O)C1(C(=O)Nc2ccc(cc2)F)CC1)OCCCN1CCOCC1)OC
24	C1CN(C)CCN1C1CCN(CC1)C(=O)Nc1cc(ccn1)Oc1cc(F)c(cc1)NC(=O)C1(CC1)C(=O)Nc1ccc(F)cc1
25	c1cc(cnc1)c1ccnc(n1)Nc1c(ccc(c1)NC(=O)c1ccc(cc1)CN1CCN(C)CC1)C

Table S1: Continuation.

inhibitor ID	SMILES
26	CC1(C)CNc2cc(ccc12)NC(=O)c1cccnc1NCc1ccncc1
27	O=C(Nc1cc(n2cc(C)nc2)cc(c1)C(F)(F)F)c1ccc(C)c(c1)Nc1nccc(n1)c1cccnc1
28	Cc1c(cc(cc1)C(=O)Nc1cccc(c1)C(F)(F)F)Nc1nc(nc2c1cnn2C)c1cnccc1
29	n1c2cccnn2c(c1)C#Cc1c(ccc(c1)C(=O)Nc1ccc(c(c1)C(F)(F)F)CN1CCN(CC1)C)C
30	N(C(=O)Nc1ccc(cc1)c1cn2c(n1)sc1c2ccc(c1)OCCN1CCOCC1)c1noc(c1)C(C)(C)C
31	c1(cc2c(cc1OC)c(ccn2)Oc1cc(c(cc1)NC(=O)Nc1cc(on1)C)Cl)OC
32	CC(C)c1nnc2ccc(Sc3ccccc3CNC(=O)Nc3cc(nn3c3ccccc3)C(C)(C)C)cn12
33	CC(C)(C)c1cc(NC(=O)NCc2ccccc2Sc2ccc3nnc(c4ccccc4SCCO)n3c2)n(n1)c1ccc(O)c(Cl)c1
34	c1c(ccc(c1)C)n1c(cc(n1)C(C)(C)C)NC(=O)Nc1ccc(cc1)C
35	c1c2c(ccc1)c(ccc2)NC(=O)Nc1n(nc(c1)C(C)(C)C)C
36	c1c(ccc(c1)Cl)NC(=O)Nc1n(nc(c1)C(C)(C)C)C

Table S2: List of the PDB identifiers of crystal structures and k_{off} values (in s^{-1}) for the p38 MAP kinase inhibitors.

inhibitor ID	PDB[a]	ref.[b] PDB	experimental k_{off}	ref. k_{off}
1	**1KV1**	25	1.40E-01	25
2	**1KV2**	25	8.30E-06	27
3*	-	-	1.40E-01	27
5*	-	-	1.50E-05	27
6*	-	-	3.80E-04	27
7	-	-	4.00E-04	27
8	-	-	3.30E-03	27
9	-	-	2.30E-05	27
11	-	-	3.80E-05	27
12*	-	-	2.60E-05	27
13	-	-	1.60E-03	27
14*	-	-	1.20E-04	27
15	-	-	2.30E-05	27
16*	**3HV6**	35	3.90E-03	27
17	-	-	3.40E-04	27
18	-	-	1.50E-04	27
20	4RZW	44	1.13E-02	30
21*	3F82	45	9.03E-02	30
22	3QRI	46	4.88E-04	30
23	3LQ8	47	9.10E-02	30
24*	5IA5	48	1.21E-01	30
25	**3HEC**	36	1.31E-01	30
26	3EFL	-	7.34E-02	30
27	3PG0	-	8.96E-03	30
28*	6FNI	49	3.18E-02	30
29	4U0I	50	4.82E-03	30
30	4RT7	51	5.20E-02	30
31	4ASE	52	9.68E-02	30
32	**2YIW**	28	4.00E-05	28
33	**2YIS**	28	2.40E-06	28
34*	-	-	1.00E-03	29
35*	-	-	9.00E-03	29
36	-	-	2.80E-02	29

*: inhibitors which were assigned to the test set. [a]Bold: ligand in complex with p38 MAP kinase; others: ligand in complex with other kinases (the binding mode was used as a reference). For ligands without crystal structures, the binding mode of ligand 16 was used as a reference. [b]ref.: reference.

Table S3: Root mean squared deviation (RMSD) between the ligand position in the crystal structure and the ligand position obtained from redocking or cross-docking to the p38 MAP kinase structures in PDBs 1KV2 or 3HV6 using Vina. Only heavy atoms were considered for the RMSD calculation.

	redocking		cross-docking	
ligand	PDB	RMSD (Å)	PDB	RMSD (Å)
2	1KV2	1.05	3HV6	1.20
16	3HV6	1.02	1KV2	1.80

Table S4: Summary of the models derived for different numbers of latent variables (LVs) for COMBINE analysis for log k_{off} of p38 MAP kinase inhibitors using crystal structures (one structure per ligand). The table lists the coefficient of determination for the training and test sets (R^2tr and R^2te), the coefficient of determination for leave-one-out cross-validation set (Q^2-LOO), average absolute errors (AAEtr, AAEv and AAEte, in log s^{-1}) and root mean squared errors (RMEtr, RMEv, and RMEte, in log s^{-1}) for the training set, leave-one-out cross-validation set, and test set, respectively. The model with 2 LVs (in bold) displayed the best balance between predictive performance and number of LVs.

LV	R^2tr	Q^2-LOO	AAEtr	AAEv	RMEtr	RMEv	R^2te	AAEte	RMEte
1	0.83	0.72	0.52	0.66	0.62	0.79	0.77	0.49	0.68
2	**0.86**	**0.75**	**0.49**	**0.68**	**0.55**	**0.75**	**0.79**	**0.54**	**0.64**
3	0.91	0.68	0.35	0.70	0.44	0.85	0.77	0.57	0.67
4	0.94	0.66	0.29	0.70	0.37	0.87	0.72	0.59	0.75
5	0.95	0.65	0.28	0.75	0.35	0.88	0.69	0.65	0.79
6	0.96	0.57	0.23	0.84	0.30	0.98	0.65	0.65	0.84
7	0.97	0.41	0.19	0.93	0.26	1.15	0.49	0.80	1.00
8	0.98	0.28	0.15	0.98	0.22	1.27	0.50	0.76	1.00
9	0.98	0.21	0.12	1.05	0.19	1.33	0.38	0.82	1.11
10	0.99	0.18	0.12	1.08	0.17	1.35	0.24	0.91	1.23

Table S5: Summary of the models derived for different numbers of latent variables (LVs) for COMBINE analysis for log k_{off} of p38 MAP kinase inhibitors using crystal structures (one structure per ligand). The model included as an extra term the solvation of the ligand, which was represented as the change in solvent accessible surface area upon unbinding. The table lists the coefficient of determination for the training and test sets (R^2tr and R^2te), the coefficient of determination for leave-one-out cross-validation set (Q^2-LOO), average absolute errors (AAEtr, AAEv and AAEte, in log s^{-1}) and root mean squared errors (RMEtr, RMEv, and RMEte, in log s^{-1}) for the training set, leave-one-out cross-validation set, and test set, respectively. The model with 2 LVs (in bold) displayed the best balance between predictive performance and number of LVs.

LV	R^2tr	Q^2-LOO	AAEtr	AAEv	RMEtr	RMEv	R^2te	AAEte	RMEte
1	0.60	0.52	0.80	0.87	0.95	1.04	0.20	1.15	1.26
2	**0.87**	**0.79**	**0.48**	**0.61**	**0.55**	**0.68**	**0.79**	**0.48**	**0.64**
3	0.90	0.73	0.38	0.65	0.46	0.77	0.75	0.58	0.71
4	0.93	0.67	0.30	0.70	0.38	0.86	0.75	0.58	0.71
5	0.94	0.66	0.28	0.74	0.35	0.87	0.70	0.64	0.78
6	0.96	0.60	0.22	0.81	0.30	0.95	0.64	0.70	0.84
7	0.97	0.46	0.19	0.92	0.26	1.10	0.50	0.79	0.99
8	0.98	0.35	0.15	0.97	0.23	1.21	0.49	0.80	1.01
9	0.98	0.27	0.12	1.02	0.19	1.28	0.33	0.93	1.16
10	0.99	0.20	0.11	1.08	0.18	1.34	0.21	0.98	1.25

Table S6: Summary of the models derived for different numbers of latent variables (LVs) for COMBINE analysis for log k_{off} of p38 MAP kinase inhibitors using crystal structures (one structure per ligand). Electrostatic interactions were computed using a distance-dependent dielectric constant. The table lists the coefficient of determination for the training and test sets (R^2tr and R^2te), the coefficient of determination for leave-one-out cross-validation set (Q^2-LOO), average absolute errors (AAEtr, AAEv and AAEte, in log s^{-1}) and root mean squared errors (RMEtr, RMEv, and RMEte, in log s^{-1}) for the training set, leave-one-out cross-validation set, and test set, respectively. The model with 3 LVs (in bold) displayed the best balance between predictive performance and number of LVs.

LV	R^2tr	Q^2-LOO	AAEtr	AAEv	RMEtr	RMEv	R^2te	AAEte	RMEte
1	0.44	0.26	0.86	0.97	1.12	1.29	0.45	0.81	1.05
2	0.86	0.77	0.49	0.60	0.55	0.71	0.72	0.61	0.75
3	**0.91**	**0.79**	**0.37**	**0.57**	**0.46**	**0.68**	**0.76**	**0.54**	**0.69**
4	0.94	0.75	0.30	0.61	0.38	0.74	0.69	0.63	0.79
5	0.96	0.75	0.25	0.61	0.31	0.75	0.68	0.65	0.80
6	0.98	0.74	0.19	0.64	0.24	0.77	0.65	0.69	0.83
7	0.98	0.73	0.15	0.65	0.20	0.78	0.69	0.64	0.79
8	0.99	0.73	0.11	0.64	0.15	0.77	0.73	0.57	0.73
9	0.99	0.76	0.09	0.62	0.12	0.74	0.70	0.60	0.78
10	1.00	0.74	0.08	0.64	0.10	0.76	0.71	0.59	0.77

Table S7: Summary of the models derived for different numbers of latent variables (LVs) for COMBINE analysis for log k_{off} of p38 MAP kinase inhibitors using structures from ensemble docking (10 structures per ligand). The table lists the coefficient of determination for the training and test sets (R^2tr and R^2te), the coefficient of determination for leave-one-out cross-validation set (Q^2-LOO), average absolute errors (AAEtr, AAEv and AAEte, in log s^{-1}) and root mean squared errors (RMEtr, RMEv, and RMEte, in log s^{-1}) for the training set, leave-one-out cross-validation set, and test set, respectively. The model with 1 LV (in bold) displayed the best balance between predictive performance and number of LVs.

LV	R^2tr	Q^2-LOO	AAEtr	AAEv	RMEtr	RMEv	R^2te	AAEte	RMEte
1	**0.83**	**0.70**	**0.56**	**0.68**	**0.69**	**0.82**	**0.72**	**0.65**	**0.74**
2	0.79	0.67	0.55	0.68	0.71	0.86	0.62	0.75	0.87
3	0.87	0.71	0.49	0.69	0.57	0.80	0.69	0.71	0.79
4	0.89	0.70	0.44	0.70	0.51	0.81	0.65	0.73	0.83
5	0.90	0.70	0.44	0.73	0.49	0.81	0.67	0.73	0.81
6	0.92	0.73	0.40	0.69	0.45	0.77	0.75	0.63	0.71
7	0.92	0.74	0.38	0.69	0.44	0.77	0.75	0.64	0.71
8	0.93	0.74	0.37	0.69	0.41	0.76	0.77	0.59	0.68
9	0.93	0.73	0.35	0.69	0.40	0.77	0.72	0.65	0.74
10	0.94	0.73	0.33	0.69	0.39	0.78	0.72	0.66	0.74

Table S8: Summary of the models derived for different numbers of latent variables (LVs) for COMBINE analysis for log k_{off} of p38 MAP kinase inhibitors using structures from ensemble docking (1 structure per ligand; the top structure was selected using Vina scores). The table lists the coefficient of determination for the training and test sets (R^2tr and R^2te), the coefficient of determination for leave-one-out cross-validation set (Q^2-LOO), average absolute errors (AAEtr, AAEv and AAEte, in log s^{-1}) and root mean squared errors (RMEtr, RMEv, and RMEte, in log s^{-1}) for the training set, leave-one-out cross-validation set, and test set, respectively. The model with 2 LVs (in bold) displayed the best balance between predictive performance and number of LVs.

LV	R^2tr	Q^2-LOO	AAEtr	AAEv	RMEtr	RMEv	R^2te	AAEte	RMEte
1	0.77	0.60	0.62	0.78	0.72	0.94	0.83	0.48	0.58
2	**0.83**	**0.65**	**0.53**	**0.73**	**0.62**	**0.88**	**0.80**	**0.49**	**0.63**
3	0.90	0.59	0.40	0.81	0.46	0.96	0.82	0.45	0.60
4	0.92	0.61	0.33	0.78	0.41	0.93	0.84	0.41	0.57
5	0.94	0.61	0.30	0.79	0.38	0.93	0.83	0.40	0.58
6	0.94	0.61	0.26	0.81	0.36	0.94	0.83	0.40	0.59
7	0.96	0.53	0.21	0.88	0.31	1.02	0.77	0.42	0.68
8	0.96	0.34	0.19	0.99	0.29	1.21	0.65	0.55	0.83
9	0.97	0.22	0.17	1.01	0.24	1.32	0.27	0.80	1.21
10	0.98	0.13	0.16	1.01	0.21	1.39	-0.15	0.96	1.51

Table S9: Summary of the models derived for different numbers of latent variables (LVs) for COMBINE analysis for log k_{off} of p38 MAP kinase inhibitors using structures from ensemble docking (20 structures per ligand). The table lists the coefficient of determination for the training and test sets (R^2tr and R^2te), the coefficient of determination for leave-one-out cross-validation set (Q^2-LOO), average absolute errors (AAEtr, AAEv and AAEte, in log s^{-1}) and root mean squared errors (RMEtr, RMEv, and RMEte, in log s^{-1}) for the training set, leave-one-out cross-validation set, and test set, respectively. The model with 1 LV (in bold) displayed the best balance between predictive performance and number of LVs.

LV	R^2tr	Q^2-LOO	AAEtr	AAEv	RMEtr	RMEv	R^2te	AAEte	RMEte
1	**0.84**	**0.65**	**0.63**	**0.75**	**0.75**	**0.88**	**0.71**	**0.66**	**0.76**
2	0.77	0.62	0.59	0.72	0.76	0.93	0.59	0.78	0.90
3	0.82	0.60	0.56	0.79	0.66	0.95	0.60	0.80	0.89
4	0.88	0.61	0.46	0.75	0.56	0.93	0.65	0.73	0.83
5	0.90	0.64	0.44	0.76	0.50	0.90	0.71	0.67	0.76
6	0.92	0.71	0.39	0.67	0.47	0.80	0.75	0.61	0.71
7	0.92	0.73	0.39	0.65	0.46	0.77	0.76	0.60	0.69
8	0.93	0.74	0.38	0.65	0.44	0.76	0.76	0.59	0.70
9	0.93	0.75	0.36	0.65	0.43	0.75	0.73	0.64	0.73
10	0.93	0.73	0.36	0.67	0.42	0.78	0.74	0.63	0.72

Table S10: Summary of the models derived for different numbers of latent variables (LVs) for COMBINE analysis for log k_{off} of p38 MAP kinase inhibitors using structures from ensemble docking (5 structures per ligand; protein structures obtained from hierarchical agglomerative clustering). The table lists the coefficient of determination for the training and test sets (R^2tr and R^2te), the coefficient of determination for leave-one-out cross-validation set (Q^2-LOO), average absolute errors (AAEtr, AAEv and AAEte, in log s^{-1}) and root mean squared errors (RMEtr, RMEv, and RMEte, in log s^{-1}) for the training set, leave-one-out cross-validation set, and test set, respectively. The model with 3 LV (in bold) displayed the best balance between predictive performance and number of LVs.

LV	R^2tr	Q^2-LOO	AAEtr	AAEv	RMEtr	RMEv	R^2te	AAEte	RMEte
1	0.80	0.60	0.62	0.75	0.79	0.94	0.64	0.70	0.84
2	0.74	0.53	0.56	0.74	0.76	1.02	0.59	0.74	0.90
3	**0.91**	**0.70**	**0.34**	**0.64**	**0.45**	**0.82**	**0.70**	**0.65**	**0.77**
4	0.92	0.73	0.32	0.61	0.42	0.77	0.66	0.65	0.83
5	0.95	0.80	0.27	0.55	0.35	0.67	0.68	0.61	0.79
6	0.95	0.82	0.24	0.52	0.33	0.63	0.66	0.64	0.82
7	0.96	0.84	0.23	0.50	0.31	0.61	0.63	0.71	0.86
8	0.96	0.85	0.20	0.48	0.29	0.58	0.60	0.72	0.89
9	0.97	0.84	0.19	0.49	0.26	0.59	0.65	0.69	0.84
10	0.97	0.83	0.19	0.50	0.26	0.61	0.69	0.64	0.79

Table S11: Summary of the models derived for different numbers of latent variables (LVs) for COMBINE analysis for log k_{off} of p38 MAP kinase inhibitors using structures from an ensemble (10 structures per ligand; structures obtained from 33 independent MD simulations, one for each inhibitor). The table lists the coefficient of determination for the training and test sets (R^2tr and R^2te), the coefficient of determination for leave-one-out cross-validation set (Q^2-LOO), average absolute errors (AAEtr, AAEv and AAEte, in log s^{-1}) and root mean squared errors (RMEtr, RMEv, and RMEte, in log s^{-1}) for the training set, leave-one-out cross-validation set, and test set, respectively. The model with 3 LV (in bold) displayed the best balance between predictive performance and number of LVs.

LV	R^2tr	Q^2-LOO	AAEtr	AAEv	RMEtr	RMEv	R^2te	AAEte	RMEte
1	0.86	0.73	0.47	0.61	0.57	0.78	0.85	0.44	0.55
2	0.90	0.76	0.43	0.64	0.49	0.74	0.78	0.59	0.66
3	**0.94**	**0.80**	**0.29**	**0.59**	**0.37**	**0.67**	**0.80**	**0.52**	**0.63**
4	0.95	0.84	0.25	0.49	0.34	0.60	0.76	0.56	0.69
5	0.96	0.84	0.23	0.49	0.29	0.61	0.67	0.63	0.81
6	0.97	0.84	0.22	0.51	0.28	0.61	0.64	0.65	0.84
7	0.97	0.82	0.21	0.53	0.26	0.63	0.65	0.62	0.83
8	0.97	0.82	0.20	0.54	0.25	0.64	0.65	0.62	0.84
9	0.97	0.81	0.19	0.55	0.25	0.65	0.67	0.60	0.81
10	0.97	0.80	0.19	0.54	0.25	0.66	0.68	0.58	0.80

Table S12: Summary of the models derived for different numbers of latent variables (LVs) for COMBINE analysis for log k_{off} of p38 MAP kinase inhibitors using structures from ensemble docking (10 structures per ligand; individual interaction energies averaged using an arithmetic or an exponential scheme). The table lists the coefficient of determination for the training and test sets (R^2tr and R^2te), the coefficient of determination for leave-one-out cross-validation set (Q^2-LOO), average absolute errors (AAEtr, AAEv and AAEte, in log s^{-1}) and root mean squared errors (RMEtr, RMEv, and RMEte, in log s^{-1}) for the training set, leave-one-out cross-validation set, and test set, respectively. The models which displayed the best balance between predictive performance and number of LVs are highlighted in bold.

LV	R^2tr	Q^2-LOO	AAEtr	AAEv	RMEtr	RMEv	R^2te	AAEte	RMEte
				arithmetic average					
1	**0.84**	**0.78**	**0.50**	**0.58**	**0.59**	**0.70**	**0.71**	**0.61**	**0.76**
2	0.88	0.78	0.44	0.59	0.51	0.70	0.72	0.56	0.75
3	0.92	0.81	0.34	0.57	0.41	0.66	0.72	0.57	0.75
4	0.93	0.81	0.32	0.58	0.39	0.65	0.74	0.54	0.72
5	0.95	0.82	0.26	0.56	0.35	0.63	0.79	0.48	0.65
6	0.96	0.83	0.22	0.54	0.31	0.63	0.78	0.50	0.66
7	0.97	0.82	0.19	0.55	0.26	0.63	0.78	0.52	0.66
8	0.97	0.85	0.17	0.49	0.24	0.58	0.75	0.61	0.70
9	0.98	0.85	0.16	0.45	0.21	0.57	0.63	0.77	0.86
10	0.99	0.79	0.14	0.53	0.18	0.68	0.52	0.86	0.97
				exponential average					
1	0.81	0.62	0.57	0.74	0.66	0.92	0.63	0.70	0.86
2	**0.86**	**0.71**	**0.50**	**0.70**	**0.56**	**0.81**	**0.66**	**0.71**	**0.82**
3	0.90	0.71	0.42	0.73	0.48	0.81	0.71	0.65	0.76
4	0.92	0.69	0.35	0.75	0.43	0.83	0.73	0.60	0.74
5	0.93	0.62	0.33	0.82	0.41	0.92	0.72	0.60	0.75
6	0.95	0.53	0.27	0.89	0.34	1.03	0.62	0.68	0.87
7	0.96	0.52	0.22	0.92	0.32	1.04	0.63	0.66	0.86
8	0.96	0.56	0.20	0.89	0.29	1.00	0.63	0.65	0.86
9	0.97	0.70	0.20	0.74	0.27	0.82	0.63	0.67	0.85
10	0.97	0.76	0.17	0.61	0.24	0.73	0.54	0.78	0.95

Table S13: Summary of the models derived for different numbers of latent variables (LVs) for COMBINE analysis for log K_d of p38 MAP kinase inhibitors using crystal structures (one structure per ligand). The table lists the coefficient of determination for the training and test sets (R^2tr and R^2te), the coefficient of determination for leave-one-out cross-validation set (Q^2-LOO), average absolute errors (AAEtr, AAEv and AAEte, in log s^{-1}) and root mean squared errors (RMEtr, RMEv, and RMEte, in log s^{-1}) for the training set, leave-one-out cross-validation set, and test set, respectively.

LV	R^2tr	Q^2-LOO	AAEtr	AAEv	RMEtr	RMEv	R^2te	AAEte	RMEte
1	0.64	0.38	0.75	1.01	0.97	1.26	0.37	0.93	1.22
2	0.78	0.46	0.67	0.99	0.75	1.17	0.39	1.05	1.21
3	0.87	0.57	0.50	0.89	0.57	1.05	0.44	0.95	1.16
4	0.90	0.60	0.42	0.90	0.50	1.02	0.27	1.00	1.33
5	0.92	0.53	0.37	0.98	0.45	1.09	0.08	1.06	1.49
6	0.94	0.50	0.29	1.01	0.39	1.14	-0.19	1.17	1.69
7	0.96	0.47	0.24	1.02	0.33	1.16	-0.34	1.21	1.79
8	0.98	0.35	0.19	1.12	0.25	1.29	-0.93	1.38	2.15
9	0.98	0.31	0.16	1.17	0.21	1.33	-1.15	1.43	2.27
10	0.99	0.31	0.13	1.20	0.19	1.33	-1.36	1.48	2.38

Table S14: Summary of the models derived for different numbers of latent variables (LVs) for COMBINE analysis for log K_d of p38 MAP kinase inhibitors using structures from ensemble docking (10 structures per ligand). The table lists the coefficient of determination for the training and test sets (R^2tr and R^2te), the coefficient of determination for leave-one-out cross-validation set (Q^2-LOO), average absolute errors (AAEtr, AAEv and AAEte, in log s^{-1}) and root mean squared errors (RMEtr, RMEv, and RMEte, in log s^{-1}) for the training set, leave-one-out cross-validation set, and test set, respectively.

LV	R^2tr	Q^2-LOO	AAEtr	AAEv	RMEtr	RMEv	R^2te	AAEte	RMEte
1	0.56	0.34	0.88	1.05	1.10	1.31	0.27	1.06	1.32
2	0.57	0.27	0.86	1.09	1.06	1.37	0.19	1.10	1.39
3	0.71	0.24	0.75	1.13	0.90	1.40	0.35	1.04	1.25
4	0.74	0.19	0.71	1.15	0.86	1.45	0.32	1.02	1.28
5	0.76	0.16	0.67	1.14	0.82	1.47	0.32	1.01	1.28
6	0.82	0.20	0.60	1.12	0.73	1.44	0.30	1.04	1.30
7	0.84	0.16	0.55	1.14	0.68	1.47	0.28	1.06	1.32
8	0.86	0.14	0.52	1.14	0.63	1.48	0.34	1.03	1.26
9	0.87	0.19	0.50	1.11	0.61	1.44	0.32	1.06	1.27
10	0.88	0.22	0.50	1.12	0.59	1.41	0.30	1.07	1.29

Table S15: Interaction energies associated with the coefficients with the highest absolute values in COMBINE models to predict log k_{off} or log K_d of p38 MAP kinase inhibitors. *: interactions in common in models to predict log k_{off} and log K_d values.

	single				ensemble		
	k_{off}^a		K_d^b		k_{off}^c		K_d^c	
residue	coefficient	residue	coefficient	residue	coefficient	residue	coefficient
van der Waals interactions							
Glu71*	0.248	Lys53*	0.382	Glu71*	0.180	Glu71*	0.154
Leu167	0.130	Leu171	0.285	Arg67*	0.128	Leu167*	0.116
Arg70	0.124	Hie148	0.215	Leu167*	0.126	Lys53	0.103
Lys53*	0.117	Glu71*	0.212	Asp168	0.119	Arg67*	0.097
electrostatic interactions							
Lys53	-0.092	Arg70	-0.158	Glu71*	0.079	Glu71*	0.043
Glu71	0.062	Met109	0.122	Lys53*	-0.033	Lys53*	-0.037

[a]Two latent variables. [b]Three latent variables. [c]One latent variable.

Table S16: Number of binding modes from ensemble docking with root mean squared deviation (RMSD) from the binding mode in the reference structure higher than 5 Å. Only heavy atoms of the ligand were considered for the RMSD calculation. No flex.: docking with rigid protein. Flex.: docking with flexible side chains (Ly53 and Glu71 for ligands 21, 23 and 24; Ile84 for ligand 26).

inhibitor ID	binding modes with RMSD > 5 Å	
	no flex.	flex.
1	0\10	-[a]
2	0\10	-
3	0\10	-
5	0\10	-
6	0\10	-
7	0\10	-
8	0\10	-
9	0\10	-
11	0\10	-
12	0\10	-
13	0\10	-
14	2\10	-
15	0\10	-
16	0\10	-
17	1\10	-
18	0\10	-
20	1\10	-
21	10\10	7\10
22	1\10	-
23	8\10	9\10
24	9\10	10\10
25	2\10	-
26	9\10	8\10
27	0\10	-
28	2\10	-
29	3\10	-
30	2\10	-
31	2\10	-
32	0\10	-
33	1\10	-
34	0\10	-
35	1\10	-
36	0\10	-

[a]Ligand not tested in flexible docking.

Table S17: Statistical measures of previous studies which predicted k_{off} values for inhibitors of p38 MAP kinase. The table lists the coefficient of determination for the training and test sets (R^2tr and R^2te), the number of inhibitors in the test set (N) and the computational method used.

reference	R^2tr	R^2te	N	method
74	-[a]	0.88	8	steered MD[b]
77	0.82	0.82	10	Volsurf
this work	0.86	0.79	11	COMBINE
76	-	0.64	41[c]	local-scaled MD
75	0.72	0.56	6	MD simulations + IFP[d] + PLS[e]
78	0.60[f]	0.43	28	random forest

[a]No training set was used. [b]MD: molecular dynamics. [c]The test set included complexes with different types of kinases, not only p38 MAP kinase. [d]Interaction fingerprints. [e]Partial least squares regression. [f]The R^2tr value reported is for a training set including the 28 complexes used in the test set.

S1 Equations for coefficient of determination, average absolute error and root mean squared error

The coefficient of determination (R^2, or Q^2 for cross validation) was calculated as:

$$R^2 = \frac{[\sum_{i=1}^{N}(y_i - \langle y \rangle)(x_i - \langle x \rangle)]^2}{\sum_{i=1}^{N}(y_i - \langle y \rangle)^2 \sum_{i=1}^{N}(x_i - \langle x \rangle)^2} \tag{S1}$$

where N is the number of data points, y_i is the experimental value for point i, x_i is the predicted value for point i, $\langle y \rangle$ is the average of experimental values and $\langle x \rangle$ is the average of predicted values.

The average absolute error (AAE) was calculated as:

$$AAE = \frac{1}{N} \sum_{i=1}^{N} |x_i - y_i| \tag{S2}$$

The root mean squared error (RME) was calculated as:

$$RME = \sqrt{\frac{\sum_{i=1}^{N}(x_i - y_i)^2}{N}} \tag{S3}$$

S2   Script for ligand superimposition

For ligand superimposition, a custom script was written using Python 3.8 and the packages RDkit 2020.03.3 [53] and MDAnalysis 1.0.0. [54,55]

"Reference system" refers to the coordinates of the protein of interest bound to a ligand, which will be used as a reference for the superimposition, and "new ligand" refers to the small molecule that will be superimposed onto the structure. The user must supply the reference system (variable "topology"), which should contain the fully prepared protein of interest with the reference ligand onto which the new ligand will be superimposed.

The reference system is parsed from a topology or coordinate file in a format supported by MDAnalysis (e.g. PDB format). The new ligand must be supplied as a RDkit mol object by the user (variable "ligand_new"). Hydrogens need to be present and the atoms in the ligand must have coordinates specifying their positions.

The reference ligand is selected from the universe object by using its residue name (variable "ligand_old"), saved as a PDB file and parsed to a mol object. RDkit is then used to calculate the Tanimoto similarity, find the maximum common fragment between the reference ligand and the new ligand, and extract it. The user can supply either additional atom indices or a completely new set of indices if the maximum common fragment is unsatisfactory. Both fragments are centered by subtracting the coordinates of their respective centers of mass from their atom coordinates. MDAnalysis is then used to calculate the rotation matrix that aligns the fragment in the new ligand optimally with the fragment in the reference ligand. Atoms are weighted according to their mass.

The new ligand is centered, translated by the center of mass of the reference ligand fragment and finally rotated using the calculated rotation matrix. The final output is the coordinates of the new ligand in PDB format, after superimposition.

In order to validate the ligand superimposition procedure, RMSDs between the superimposed ligands (ligands 3-15, 17-18 and 34-36) and the reference ligand (ligand 16) were

computed using the heavy atoms of the maximum common substructure for the ligands compared and the web server ligRMSD.[56] Table S18 shows that the RMSD values are low, as expected.

Table S18: Root mean squared deviations (RMSDs) between the superimposed ligands (ligands 3-15, 17-18 and 34-36) and the reference ligand (ligand 16). RMSDs were computed using the heavy atoms of the maximum common substructure for the ligands compared and the web server ligRMSD.

ligand ID	RMSD (Å)
3	0.16
5	0.30
6	0.31
7	0.30
8	0.31
9	0.31
11	0.30
12	0.30
13	0.18
14	0.18
15	0.35
17	0.18
18	0.30
34	0.18
35	0.20
36	0.20