JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Leveraging Data Fusion Strategies in Multireceptor Lead Optimization MM/GBSA End-Point Methods

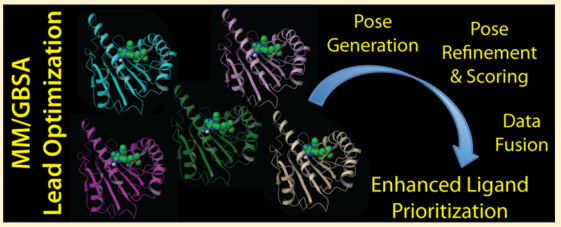
Jennifer L. Knight,*[,†] Goran Krilov,[†] Kenneth W. Borrelli,[†] Joshua Williams,[†] John R. Gunn,[†] Alec Clowes,[†] Luciano Cheng,[†] Richard A. Friesner,[‡] and Robert Abel[†]

[†]Schrödinger, 120 West 45th Street, 17th Floor, Tower 45, New York, New York 10036-4041, United States

[‡]Columbia University, Department of Chemistry, 3000 Broadway, MC 3110, New York, New York 10027, United States

Supporting Information

ABSTRACT: Accurate and efficient affinity calculations are critical to enhancing the contribution of in silico modeling during the lead optimization phase of a drug discovery campaign. Here, we present a large-scale study of the efficacy of data fusion strategies to leverage results from end-point MM/GBSA calculations in multiple receptors to identify potent inhibitors among an ensemble of congeneric ligands. The retrospective analysis of 13 congeneric ligand series curated from publicly available data across seven biological targets demonstrates that in 90% of the individual receptor structures MM/GBSA scores successfully identify subsets of inhibitors that are more potent than a random selection, and data fusion strategies that combine MM/GBSA scores from each of the receptors significantly increase the robustness of the predictions. Among nine different data fusion metrics based on consensus scores or receptor rankings, the SumZScore (i.e., converting MM/GBSA scores into standardized Z-Scores within a receptor and computing the sum of the Z-Scores for a given ligand across the ensemble of receptors) is found to be a robust and physically meaningful metric for combining results across multiple receptors. Perhaps most surprisingly, even with relatively low to modest overall correlations between SumZScore and experimental binding affinities, SumZScore tends to reliably prioritize subsets of inhibitors that are at least as potent as those that are prioritized from a "best" single receptor identified from known compounds within the congeneric series.



INTRODUCTION

Consensus scoring or data fusion strategies are widely used in the lead identification phase of drug discovery campaigns. During this phase, computational methods typically generate scores that describe the quality of modeled protein−ligand interactions[1] or the chemical or shape similarity between known actives and database compounds.[2] Ideally, these scores will identify new chemotypes that will be active against a given biological target. Recent studies have demonstrated that combining results from multiple methods generally improves the quality and robustness of the predictions compared with any individual method.[2b,3]

While these data fusion methods have been well documented for use in lead identification, their effectiveness in structure-based lead optimization has not yet been extensively validated. In silico contributions during the lead optimization stage tend to focus on identifying a handful of compounds within a congeneric series that are the most likely to be strong binders and should be designated as candidates for synthesis and biological testing; the most effective strategies are those that reliably identify potent compounds among the ensemble of idea molecules. In practice, these idea compounds could be culled from large ligand databases, identified from other in silico methods, or suggested by medicinal chemists. Ideally, in each design cycle, computational methods will tend to select those proposed compounds that have improved or comparable binding affinities relative to the best compounds synthesized to date. The results of the subsequent synthesis and biological testing of these prioritized compounds will then focus the next round of design ideas. Thus, through each iteration of design and synthesis in a given project, the chemical properties of the lead compound will be progressively optimized; that is, its binding affinity will be enhanced and/or its binding affinity will be maintained while its ADME/T (absorption/distribution/metabolism/excretion and toxicity) profile is ameliorated.

Within the limits of the force field accuracy, free energy methods provide the most accurate description of the thermodynamics of protein−ligand binding. However, the computational expense of free energy perturbation and thermodynamic integration simulations often limits their use in practice. More commonly employed are the more efficient, albeit more approximate, end-point approaches, such as Linear Interaction Energy calculations[4] or MM/GBSA and MM/PBSA (Molecular Mechanics with Generalized Born or Poisson−Boltzmann and Surface Area approximation)[5] approaches. MM/GBSA and MM/PBSA approaches are based on

$$\Delta G_{\text{bind}}^{\text{MM/GBSA}} = \Delta E_{\text{MM}} + \Delta G_{\text{solv}} + \Delta G_{\text{SA}}$$

Received: March 5, 2014
Published: July 9, 2014

Table 1. Data Set Characteristics

target for congeneric series	unique PDBs	no. receptors	no. xtal ligs	xtal ligs core RMSD(Å)[a]	actives: no. unique ligs	actives: ΔG_{bind} range (kcal/mol)	actives: R^2 (MW vs ΔG_{bind})	inactives: # ligs	SAR lab
bace1_A	3pi5, 3qbh, 3vf3, 3vg1, 4d8c, 4d83, 4d85, 4d88, 4d89	9	9	0.6	67	4.7	−0.43	3	Novartis[36]
bace1_B	4fsr, 4fs4, 4h1e, 4h3f, 4h3g, 4h3i, 4h3j, 4ha5	15	8	0.1	43	5.1	−0.32	0	Merck[37]
cdk2_A	1h1p, 1h1q, 1h1r, 1h1s, 1oi9, 1oiu, 1oiy, 2g9x	14	8	1.3	54	3.9	−0.02	13	U. Newcastle, U. Oxford, AstraZeneca[38]
cdk2_B	2c5n, 2c5o, 2c5v, 2uue, 2wev	14	5	0.7	26	2.8	−0.07	8	Cyclacel[39]
chk1_A	2hog, 2hxl, 2hxq, 2hy0	4	4	0.7	46	4.2	−0.27	0	Merck[40]
chk1_B	2e9n, 2e9o, 4fsm, 4fsn, 4fsq, 4fsr, 4fsu, 4ftk, 4ftl, 4fto	10	10	0.7	122	6.2	−0.03	5	Abbott[41]
chk1_C	4fsw, 4fsy, 4ft0, 4ftr, 4ftu	5	5	0.4	44	5.3	−0.51	6	Abbott[42]
chk1_D	2e9p, 2e9u, 2e9v, 2ymp, 4ft3, 4ft5, 4fta	7	7	0.5	87	3.6	0.00	2	Abbott[43]
hsp90[b]	1uy6, 1uy7, 1uy8, 1uy9, 1uyc, 1uyd, 1uye, 1uyf, 1uyg, 1uyh, 1uyi, 1uyk, 2fwy, 2fwz, 2h55	15	8	0.3	55	6.6	−0.37	4	Chiosis lab, MSKCC[44]
p38 MAP kinase	3fkl, 3fln, 3flq, 3fls, 3flw, 3fly, 3flz, 3fmk, 3fmh, 3fsk	10	10	1.1	37	3.2	−0.33	0	Roche[45]
ptp1b	2qbp, 2qbq, 2qbr, 2qbs, 2zmm, 2zn7	6	6	0.5	55	4.2	−0.63	0	Wyeth[46]
thrombin_A	1nt1, 1ta2, 1ta6, 1mu6, 1mu8, 1mue, 1nm6, 1sl3, 1z71	9	6	0.2	100	6.6	−0.31	2 LB[c]	Merck[47]
thrombin_B	1nt1, 1ta2, 1ta6, 1mu6, 1mu8, 1mue, 1nm6, 1sl3, 1z71	9	6	0.2	25	5.2	−0.52	2 LB[c]	Merck[47f]

[a]The RMSD of the ligand atoms in the common core after the crystallographic structures of the complexes were aligned. [b]The original SAR study reported EC50 values (ref 44a and 44b); this study uses IC50 values measured by the same SAR lab using the fluorescence polarization assay described in ref 44c. [c]Experimental measurements were reported as lower bounds.

where ΔE_{MM} is the physics-based energy difference between the free and bound states of the protein and ligand, ΔG_{solv} is the difference in the GB solvation energy between the free and bound states of the protein and ligand, and ΔG_{SA} is the difference in the surface area energy between the free and bound states of the protein and ligand. Recent studies have validated the use of MM/GBSA calculations for virtual screening enrichments[1b,6] and in predicting bound ligand poses.[6,7]

In recent years, different facets of MM/GBSA calculations have been explored to assess their impact on and improve the quality of the predicted relative binding affinity calculations. For example, studies have shown the importance of correct model preparation,[8] of using variable dielectric constants in the GB calculation,[9] of taking into account ligand strain, that is, the difference in the energy between the ligand in its relaxed free state conformation and its bound conformation[1b,9b,10] and of explicitly accounting for the contribution from the desolvation of the protein.[11] Other studies offer conflicting perspectives on or have been less conclusive about the extent of end-point sampling that is required,[6,9b,12] the advantage of incorporating explicit water molecules,[13] and inclusion of ligand conformational entropy, either in the form of empirical estimates,[9b,14] configurational integrals,[15] or normal-mode calculations.[16]

In this study, however, we examine a different aspect of MM/GBSA calculations, that is, how the robustness of ligand prioritizations in MM/GBSA calculations improves with the use of multiple crystallographic structures within a congeneric series. This retrospective study mimics the context of modeling congeneric series in lead optimization campaigns in which crystallographic structures of several chemical variations of the lead compound are available and where, from among a large set of candidate compounds that contain a common core, a handful of ligands would be prioritized for synthesis and subsequent testing. The primary measures of model quality, therefore, are the potency of the prioritized ligands and the reliability of obtaining high quality predictions.

To the best of our knowledge, this is the largest systematic study of end-point methods for scoring congeneric series to date and includes 13 data sets covering seven different biological targets with a combined total of 127 receptor structures and 761 ligands. For each of the data sets, the series of SAR compounds are processed through an automated lead optimization protocol which includes generating multiple three-dimensional conformations of each ligand, docking them into multiple receptor structures, refining and scoring the poses using hierarchical sampling techniques and MM/GBSA scoring strategies. This investigation demonstrates that data fusion strategies that combine scores and rankings from the ensemble of receptors tend to generate more robust predictions compared with those obtained from using a single receptor alone. While this strategy could be applied in virtual screening contexts,[2e,f] these curated data sets closely mimic those that would be encountered once a lead compound has already been identified and specific chemical modifications are being explored to improve its potency or maintain its potency while improving its ADME/T profile; thus, our discussion will specifically focus on lead optimization.

METHODS

Data Set Preparation. Thirteen data sets were curated that had (i) at least four publicly available crystallographic structures with identical sequences and with cocrystallized ligands containing a common core, (ii) at least 25 unique ligands containing the common core for which experimental binding affinity data was available from the same laboratory and same assay, and (iii) experimental binding affinities that spanned at least 2.5 kcal/mol. Ligands with binding affinities that were

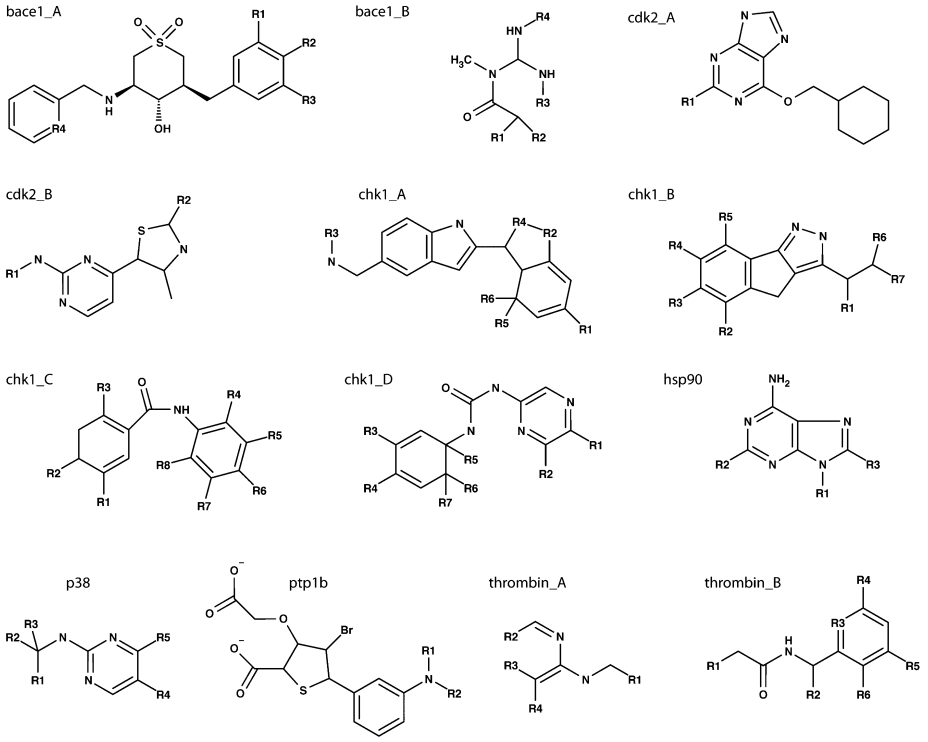


Figure 1. 2D schematic of common core for each data set.

reported as upper or lower bounds (e.g., with an $IC_{50} > 10\ \mu M$) were not included in the data sets. Table 1 summarizes the characteristics of these data sets and Figure 1 illustrates the common core for each data set. The 13 data sets include seven distinct biological targets: beta secretase 1 (bace1), cyclin dependent kinase 2 (cdk2), serine/threonine dependent kinase (chk1), heat shock protein 90 (hsp90), p38 MAP kinase (p38), protein/tyrosine phosphatase 1B (ptp1b), and thrombin and each data set includes between 4 and 15 receptor structures, between 25 and 122 unique ligands and experimental binding affinity ranges between 2.8 and 6.6 kcal/mol.

Structures of the complexes were obtained from the Protein Data Bank.[17] Each complex was processed through the Protein Preparation Wizard[18] in Maestro, version 9.3,[19] to remove crystallographic waters, optimize rotamer states for Gln and Asn side chains and protonation assignments for Glu, Asp, and His side chains. Models for loops that were missing near the binding pockets were constructed with the available X-ray reflection data using PrimeX, version 2.2.[20] Complexes associated with a given target were aligned using the Binding Site Alignment tool in Maestro in which residues within 5 Å of any crystallographic ligand were selected for alignment. Across the binding site residues, aligned complexes within each data set had backbone and heavy-atom RMSDs of up to 1.0 and 1.3 Å, respectively. The RMSD of the heavy-atoms within the common cores of the ligands in the aligned complexes were between 0.1 and 1.3 Å.

Force Field Builder. All ligands were processed through the Force Field Builder[21] in Maestro to derive parameters for torsional angles that were not explicitly represented in the OPLS2.1 force field.[22] These parameters were used in all subsequent calculations.

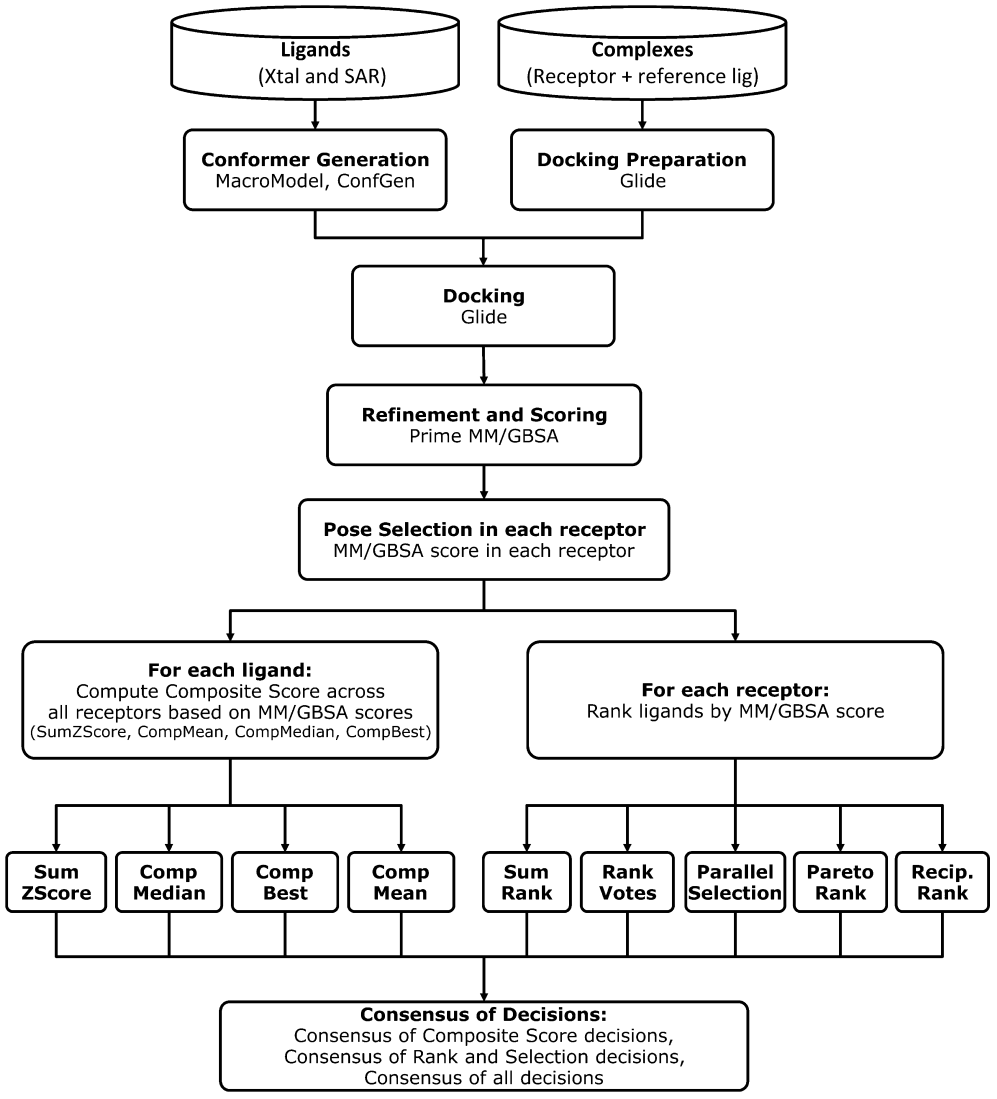
Lead Optimization Workflow. An overview of the lead optimization workflow is depicted in Scheme 1 and comprises a conformation generation stage, a pose generation stage, and a pose refinement and scoring stage. All calculations were performed on an in-house computer cluster comprised of dual six-core and 16-core AMD 2.2 GHz processors.

An ensemble of 7−40 structures for each ligand was constructed by sampling the ligand torsional angle space and accepting low-energy conformations in ConfGen, version 2.6,[23] and MacroModel, version 10.2.[24] Default settings were used for ConfGen to generate up to 50 conformations. MacroModel sampling was performed in solvent and in vacuum environments with each of three force fields (OPLS2.1, OPLS2005[25] and MMFF[26]). For each combination of environment and force field in MacroModel, the ligand conformation which had the lowest energy was retained. All MacroModel- and ConfGen-generated conformations were docked in each receptor structure using core-constraints in Glide SP, version 6.1[27] in which the grid of each receptor was generated by Glide using a 10 $Å^3$ box centered on the receptor's cocrystallized ligand.

MM/GBSA calculations were performed for each docked pose using the OPLS2.1 force field and VSGB2.1[9d] which is a surface-generalized Born model with a variable dielectric correction and incorporates additional physics-based terms designed to capture the effects of hydrogen bonding, $\pi-\pi$ interactions, self-contact corrections, and hydrophobic packing in Prime version 3.4.[28] Hierarchical sampling[29] was used to refine the translation and orientation of the ligand core and to identify the lowest-energy rotamer states of the ligand substituents decorating the congeneric core in the context of a rigid receptor. The energies of the receptor and ligand were then evaluated individually in solvent from the optimized bound conformation. The overall MM/GBSA binding affinity for a given ligand pose was estimated by the difference in the energy calculated for the complex and the receptor and ligand modeled in the solvent environment. For a given ligand, the final energy associated with its free state was taken to be the lowest energy reported for any input pose of the ligand. Within

dx.doi.org/10.1021/ct500189s | J. Chem. Theory Comput. 2014, 10, 3207−3220

Scheme 1. Lead Optimization Workflow and Data Fusion Strategies



a given receptor, the best pose for each ligand was identified as the pose that had the lowest binding affinity as estimated by the MM/GBSA score.

Data Fusion. Results from modeling and scoring congeneric ligands in multiple receptors with the lead optimization workflow were fused in multiple ways and the definitions of the data fusion strategies are listed in Table 2. At each layer of data fusion, the five compounds that were predicted to be the best binders were identified for each data set.

In the first layer of data fusion, raw MM/GBSA scores were combined across multiple receptors to compute four composite scores for each unique ligand. The dimensionless SumZScore is a composite score for each ligand which is the sum of the ligand's Z-scores of its best poses within the individual receptors. The ligand Z-score within a receptor is derived by subtracting the population mean from the ligand's raw score and then dividing the difference by the population standard deviation. The CompMean score for each ligand is the mean of its MM/GBSA binding affinities evaluated across the receptors while CompMedian and CompBest scores for each ligand are its median and best MM/GBSA score, respectively, across the receptors.

In the second layer of data fusion, consensus receptor rankings were determined from the ligand rankings across multiple receptors. In each receptor j, the best ligand poses were first sorted by their MM/GBSA score and then assigned a rank based on this ordering from lowest to highest MM/GBSA score. These ligand ranks, $R(\text{rec } j)$, were used in each of the five consensus ranking and voting schemes. For a given ligand, the SumRank is the sum of its ranks across each of the receptors in the data set. RankVotes represents the number of times a ligand is ranked within the top five compounds in the ensemble of receptors. ParallelSelection selects the top ligand from each receptor in turn until an ensemble of five unique ligands is identified. If a ligand that would have been selected was already selected by a previous receptor then the next best ligand from the current receptor is chosen. ParetoRank scores a compound based on how many other compounds are ranked higher than it in all receptors. The top five ligands were identified in each scheme, and all ties were broken using SumZScore.

In the third layer of data fusion, the decision-level consensus ranking identifies the ligands that are most frequently identified to be potent among the composite score rankings and/or the consensus receptor rankings and selection strategies.

dx.doi.org/10.1021/ct500189s | J. Chem. Theory Comput. 2014, 10, 3207−3220

Table 2. Data Fusion Definitions

fusion rule	formula
Composite Scores: Combining Raw Scores from Multiple Receptors	
SumZScore	$\mathrm{SZS}_i = \sum_{j=1}^{N} Z_i(\mathrm{rec}\ j);\ Z_i(\mathrm{rec}\ j) = \frac{S_i(\mathrm{rec}\ j) - \langle S(\mathrm{rec}\ j)\rangle}{\sigma(S(\mathrm{rec}\ j))}$
CompMean	$\mathrm{CM}_i = \frac{1}{N}\sum_{j=1}^{N} S_i(\mathrm{rec}\ j)$
CompMedian	$\mathrm{CMed}_i = \mathrm{median}\{S_i(\mathrm{rec}\ 1)...S_i(\mathrm{rec}\ N)\}$
CompBest	$\mathrm{CB}_i = \min\{S_i(\mathrm{rec}\ 1)...S_i(\mathrm{rec}\ N)\}$
Consensus Receptor Rankings and Selections: in each receptor j, ligands are ordered by their Score, $S_i(\mathrm{rec}\ j)$, and then are ranked, $R_i(\mathrm{rec}\ j)$	
SumRank	$\mathrm{SR}_i = \sum_{j=1}^{N} R_i(\mathrm{rec}\ j)$
RankVotes	$\mathrm{RV}_i = \sum_{j=1}^{N} V_i(\mathrm{rec}\ j)\begin{cases} V_i(\mathrm{rec}\ j) = -1, & \text{if } R_i(\mathrm{rec}\ j) < \mathrm{cut} \\ V_i(\mathrm{rec}\ j) = 0, & \text{otherwise} \end{cases}$
ParallelSelection	$\mathrm{PS}_i = \sum_{k=1}^{K} V_k\begin{cases} V_k = 1, & \text{if for all } j = 1 \text{ to } N\text{: } R_i(\mathrm{rec}\ j) > R_k(\mathrm{rec}\ j) \text{ for all } j \\ V_k = 0, & \text{otherwise} \end{cases}$
ParetoRank	$\mathrm{PR}_i = \sum_{k=1}^{K} V_k\begin{cases} V_k = 1, & \text{if for all } j = 1 \text{ to } N\text{: } R_i(\mathrm{rec}\ j) > R_k(\mathrm{rec}\ j) \text{ for all } j \\ V_k = 0, & \text{otherwise} \end{cases}$
ReciprocalRank	$\mathrm{RR}_i = \sum_{j=1}^{N} \frac{1}{R_i(\mathrm{rec}\ j)}$

Measures of Model Quality for Individual Data Sets. The overall agreement between the experimental and estimated binding affinities was described by the Pearson correlation coefficient, R^2. Due to the importance of ligand prioritization during the lead optimization stage, the predictive ability of single receptors or data fusion strategies to identify potent inhibitors was evaluated by the mean experimental binding affinity (MEBA) among the N selected ligands:

$$\mathrm{MEBA} = \frac{1}{N}\sum_{i=1}^{N} \Delta G_{\mathrm{bind}}^{i}$$

For each data set, the MEBA of the five most and least potent ligands based on their experimental data as well as the MEBA of the five heaviest compounds were also computed. In addition, an estimate of the quality expected from a random selection of ligands was computed by randomly selecting a subset of five ligands from the ensemble, calculating their MEBA and repeating the procedure 10 000 times to obtain an overall average and standard deviation [i.e., MEBA(random) and σ(MEBA(random))], by which to compare with the MEBA of the five ligands predicted to be the best binders.

Measures of Classification Similarity. Fleiss' κ[30] was used to quantify the agreement between the ligands that are prioritized by each of the individual receptors or data fusion strategies and is defined as

$$\kappa = \frac{\overline{P} - \overline{P}_{\mathrm{e}}}{1 - \overline{P}_{\mathrm{e}}}$$

where the denominator represents the degree of agreement that is attained by chance and the numerator is the degree of agreement that is attained above chance. More specifically,

$$\overline{P} = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{n(n-1)} \sum_{j=1}^{k} n_{ij}(n_{ij} - 1)$$

and

$$\overline{P}_{\mathrm{e}} = \sum_{j=1}^{k} \frac{1}{Nn} \sum_{i=1}^{N} n_{ij}$$

in which N is the total number of ligands in the data set, n is the number of methods or metrics that generated a prioritized list, k is the number of categories into which ligands are assigned, and n_{ij} is the number of methods or metrics that assigned ligand i to be category j. In this case, $k = 2$ corresponding to the "prioritized" or "non-prioritized" ligand assignments. If the same five ligands are prioritized by each method under consideration then $\kappa = 1$ and if there is no agreement among the methods, other than what would be expected from chance selections, then $\kappa = 0$. As a first approximation, Landis and Koch suggest that κ values of 0.41−0.60 indicate moderate agreement, 0.61−0.80 indicate substantial agreement and 0.81−1.00 indicate near perfect agreement.[31]

RESULTS AND DISCUSSION

Data Set Characteristics. The distributions of binding affinities within each of the 13 data sets in this study varied considerably. As depicted in Figure 2, the difference between the MEBA for the five most and least potent inhibitors within any given data set ranged from 2.8 to 6.6 kcal/mol. In all but the cdk2_A and hsp90 data sets, the MEBA of the five heaviest compounds was lower than what would be obtained from a random selection of compounds. In fact, in two data sets, chk1_C and ptp1b, the MEBA of the five heaviest ligands were within 0.5 kcal/mol of the theoretical upper bound of the data set (i.e., the MEBA of the five most potent inhibitors). For nine

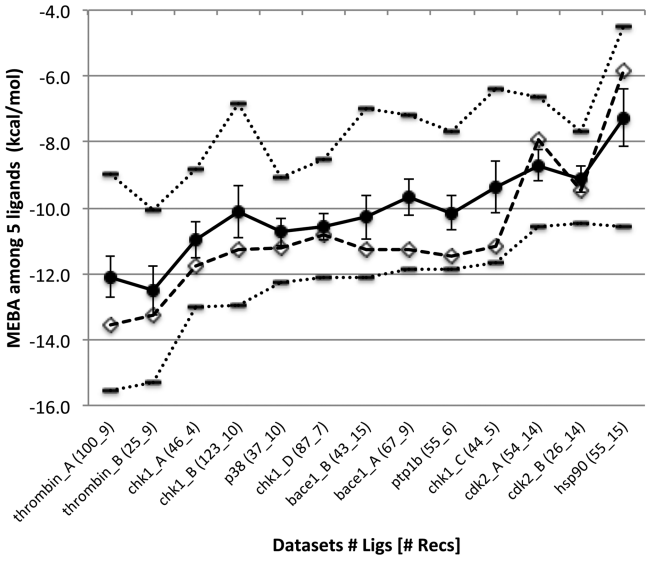


Figure 2. Data set characteristics. The theoretical maximum and minimum mean experimental binding affinities (MEBA) for the five most and least potent compounds, respectively (dotted lines, bars), in each data set. The average and standard deviations calculated from the average MEBA values calculated from 10 000 random selections of five compounds (solid line, circles) and the MEBA of the five heaviest compounds from the data set (dashed line, diamonds).

data sets, the correlation between the ligand molecular weights and experimental binding affinities was statistically significant (see Supporting Information Table 1 for p-values) and the corresponding R^2 values ranged from 0.27 for chk1_A to 0.63 for ptp1b. Even the weak compounds whose binding affinities were reported as upper limits (e.g., $IC_{50} > 10\ \mu M$) tended to have relatively low molecular weights so including these compounds did not significantly degrade the observed correlation between molecular weight and binding affinity. Four data sets in this study, cdk2_A, cdk2_B, chk1_B, and chk1_D, exhibited very weak or no significant correlation between molecular weight and experimental binding affinities with R^2 values less than 0.07. It has been our experience that publicly available data sets tend to show greater agreement between molecular weight and binding affinity than proprietary data sets in part due to the presence of heavier molecular weight compounds that are poor or nonbinders in the latter data sets. However, ensembles of SAR ligands such as these cdk2 and chk1 data sets that contain weak binders that have high-molecular weight are crucial to robust methods development to ensure that new computational strategies are able to effectively discriminate among compounds that have a similar mass but a wide range of potencies within a series.

Single-Receptor Results. For each data set, the ensemble of congeneric ligands was docked into multiple receptor structures and their poses were refined and scored by the MM/GBSA method. Table 3 summarizes the resulting MEBA values for the top compounds selected by MM/GBSA score within individual receptors. For half of the data sets, the variation in the MEBA between the "best" and the "worst" receptor is within 1 kcal/mol with as little difference as 0.6 kcal/mol. The largest variation in the quality among a series of receptors is for the bace1_B data set in which the MEBA of the five prioritized ligands can differ by up 2.2 kcal/mol. By pooling the list of prioritized compounds from each of the receptors, there were as few as nine distinct ligands that were prioritized among the chk1_B and ptp1b inhibitors and as many as 23 distinct ligands that were prioritized among the cdk2_A inhibitors. With this variation in the consistency of results among structural models, for a given data set, it is not clear a priori how important the selection among receptors would be or which receptor would be the optimal one to select as the basis for modeling in blind studies. These trends in the MEBA values are relatively insensitive to the number of compounds that are selected for prioritization. Figure 1 in the Supporting Information shows the MEBA values that are obtained for selecting from five to 20 compounds within the receptors that have the best and poorest agreement with experimental data from each data set.

Many of the individual receptors were successfully able to identify potent inhibitors from the ensemble of congeneric ligands. In all but the chk1_D data set, the median single-receptor MEBA was better than the average MEBA obtained by selecting compounds at random and in 90% of the receptors (i.e., 113 out of a total of 127 receptors), the single-receptor MEBA values for the prioritized ligands were better than their

Table 3. Baseline Data Set Qualities Compared with Single-Receptor Results from the Lead Optimization Workflow[a]

		random[c]		single receptor MM/GBSA						
target	MW[b]	avg.	std. dev.	no. rec	no. unique priorities	min	median	max	avg.	std. dev.
bace1_A	−11.3	−9.7	0.6	9	16	−9.6	−10.0	−11.3	−10.2	0.5
bace1_B	−11.3	−10.3	0.7	15	21	−9.5	−11.5	−11.7	−11.4	0.5
cdk2_A	−7.9	−8.7	0.5	14	23	−8.4	−9.4	−9.9	−9.3	0.5
cdk2_B	−9.5	−9.1	0.4	14	15	−9.0	−9.6	−9.8	−9.4	0.3
chk1_A	−11.8	−11.0	0.5	4	15	−10.5	−11.3	−11.8	−11.2	0.5
chk1_B	−11.3	−10.1	0.8	10	15	−11.9	−12.4	−12.6	−12.3	0.2
chk1_C	−11.2	−9.4	0.8	5	9	−10.7	−11.2	−11.3	−11.1	0.2
chk1_D	−10.8	−10.6	0.4	7	13	−10.3	−10.5	−10.9	−10.6	0.2
hsp90	−5.9	−7.3	0.9	15	9	−9.1	−9.7	−9.7	−9.6	0.3
p38 MAP kinase	−11.2	−10.7	0.4	10	16	−11.3	−11.5	−11.9	−11.5	0.2
ptp1b	−11.5	−10.2	0.5	6	9	−11.0	−11.3	−11.6	−11.3	0.2
thrombin_A	−13.5	−12.1	0.6	9	14	−12.5	−13.1	−13.6	−13.1	0.3
thrombin_B	−13.3	−12.5	0.7	9	11	−13.3	−14.4	−15.2	−14.4	0.5
avg.	−11.0	−10.2	0.6		14.5	−10.5	−11.2	−11.6	−11.2	0.3

[a]MEBA values in kcal/mol among the five ligands identified to be "top binders" were computed for each receptor based on the MM/GBSA scores of the refined poses. [b]Of the five heaviest compounds in the data set. [c]The average of 10 000 random selections from the data set.

respective average MEBA from random ligand selections. For eight data sets, the median single-receptor MEBA was better than the MEBA of the heaviest ligands indicating that the MM/GBSA calculations are capturing the quality of specific interactions and are not simply providing a measure of the number of interactions. Six data sets had at least one receptor whose MEBA was within 0.5 kcal/mol of their respective maximum MEBA and the MEBA for 13% of all receptors were within 0.5 kcal/mol of their respective maximum MEBA values and 60% were within 1 kcal/mol. Thus, many individual receptors are approaching the optimal limit of what can be obtained given the inherent uncertainty in the experimental measurements. Specifically, if the uncertainty in each experimental measurement is 0.35 kcal/mol then the difference of 0.5 kcal/mol in the MEBA values of the five experimentally most potent and the five prioritized ligands would not be statistically significant.

Data Fusion Using Composite Scores Across Multiple Receptors. In the first layer of lead optimization data fusion, single-receptor MM/GBSA scores were combined into composite scores (i.e., SumZScore, CompMean, CompMedian, or CompBest) that were then used to identify the top five ligands. In the thrombin_B data set, the same five ligands were prioritized regardless of which of the four composite scores was used. Among the remaining data sets, there was substantial, though not perfect, overlap among the ligands that were selected by the different consensus scores, as indicated by Fleiss' κ values between 0.55 and 0.89 (see Table 2 in Supporting Information). For example, chk1_A, which exhibited the least agreement among the top ligands that were identified by different consensus scores, saw one ligand prioritized by all four consensus scores, four ligands prioritized by three of the consensus scores and four other ligands that were only identified to be a top binder by one consensus score.

Figure 3A depicts the differences between the average binding affinity of the subsets of five ligands identified by the MM/GBSA composite scores and the average single-receptor selections in terms of standardized Z-scores. The quality of predictions from SumZScore and CompMean scores were highly correlated with one another and tended to be in marginally more consistent agreement with the average single-receptor predictions than either CompMedian and CompBest predictions. Predictions based on SumZScore were improved over the average single-receptor predictions in three data sets and were of similar quality (i.e., < ±1 unit) to the average single-receptor predictions in eight data sets. Thus, in most data sets, the composite SumZScore score could robustly identify potent inhibitors without requiring any prior information about the quality of individual receptors to model the congeneric series.

Data Fusion Using Consensus Receptor Rankings Across Multiple Receptors. In the second layer of data fusion, five different consensus ranking or voting schemes (i.e., SumRank, RankVotes, ParallelSelection, ParetoRank, and ReciprocalRank) were used to combine rankings from the ensemble of receptors and identify the top five binders in each data set. All data sets demonstrated reasonable to substantial agreement among the ligands that were prioritized by the different consensus schemes with Fleiss' κ values between 0.53 and 0.87 (see Supporting Information Table 2).

As illustrated in Figure 3B, without any prior information about the quality of the receptors, in all but two to three data sets, the potency of the prioritized ligands obtained by the

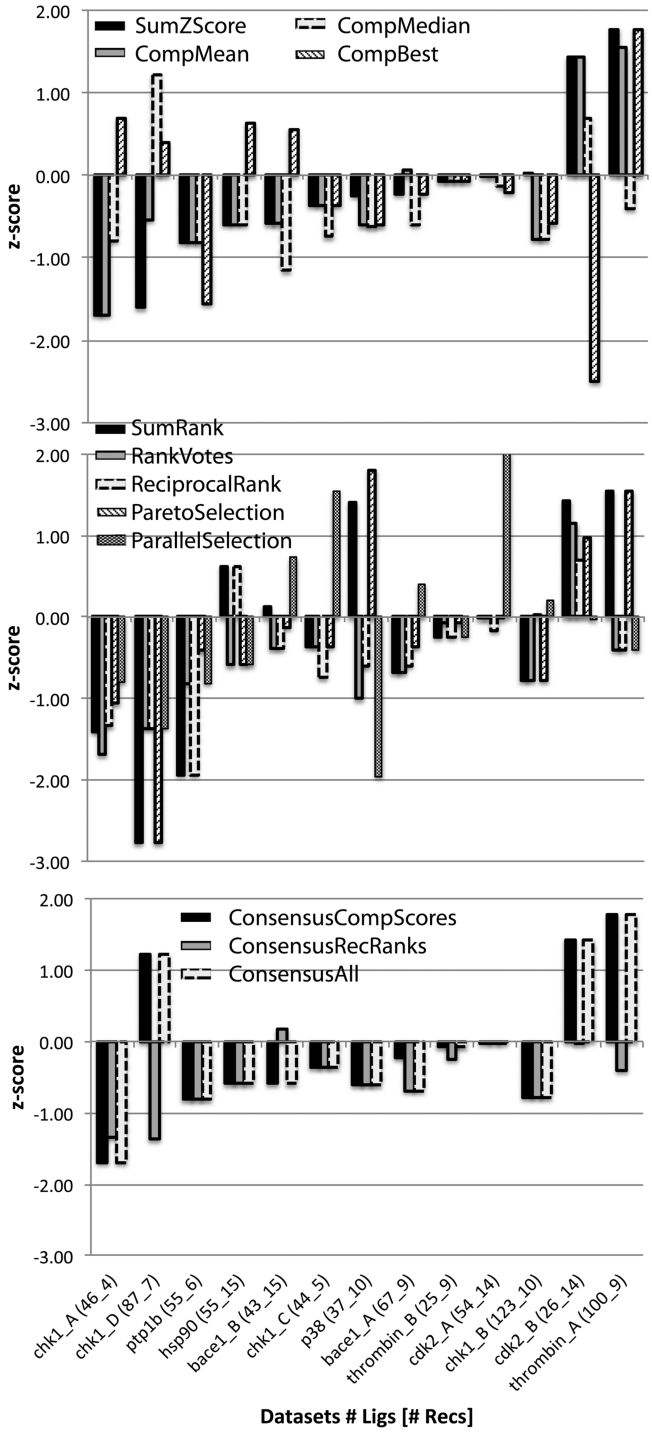


Figure 3. Comparison between the average MEBA of the top five ligands prioritized from modeling in single receptors and the MEBA of the top five ligands prioritized from (A) composite scores, (B) consensus receptor rankings across multiple receptors, and (C) the consensus of the decisions from composite scores and receptor rankings. The normalized z-scores plotted on the y-axis are computed by z-score = $(\mathrm{MEBA}_{\mathrm{datafusion}} - \mathrm{MEBA}_{\mathrm{single}})/(\sigma(\mathrm{MEBA}_{\mathrm{single}}))$.

receptor rankings was comparable to or better than what would be obtained on average by randomly selecting a single receptor to base one's predictions. The quality of the predictions by each consensus receptor ranking strategy improved over the corresponding average single-receptor predictions in three to four data sets, were of comparable quality in six to nine data

dx.doi.org/10.1021/ct500189s | J. Chem. Theory Comput. 2014, 10, 3207−3220

sets, and were of poorer quality in 0 to 3 data sets. ReciprocalRank and RankVotes demonstrated the largest overall increase in the potency of the top ligand predictions across the data sets while ParallelSelection tended to have marginally poorer quality predictions and had slightly less agreement with the ensemble of ligands that were prioritized by the other data fusion strategies, especially for data sets that contained more receptors.

Theoretical Differences among Data Fusion Strategies. Data fusion strategies assume a certain level of noise in each of the contributing data sources, but proposed strategies in the literature differ in the sensitivity of their combined results to these errors. For example, sum-based data fusion rules that more evenly distribute the weight of the contributions from the different data sources tend to be more resilient to large individual errors than product-based rules whose results may be dominated by contributions from a single data source.[32] In this work, ligands that were prioritized by favorable SumZScore, SumRank, CompMean, and ParetoRank data fusion metrics tended to score reasonably well in every receptor since poor binding scores or ranks would explicitly degrade the score. By contrast, poor binding scores or ranks do not contribute to CompBest, CompMedian, RankVotes, ParallelRank, and (only weakly to) ReciprocalRank, and thus, these metrics will tend to prioritize ligands that score exceptionally well in at least one receptor and can be more tolerant of poor docking in subsets of receptors, thus, avoid penalizing false negatives. As a result, SumZScore, SumRank, CompMean, and ParetoRank metrics may miss potent binders that require a specific tuning of the binding pocket, for example, larger compounds that cannot fit into narrower pockets. In this way, SumZScore, SumRank, CompMean, and ParetoRank will tend to avoid false positives by prioritizing ligands that are less sensitive to the structural details of the binding pocket. Arguably, these four metrics will provide robust overall predictions that will result in prioritized ligands that will more often be found to be potent in vitro and will tend to be more amenable to further development than ligands that are highly tuned to a single receptor structure. Thus, these predictions are likely to identify inhibitors that will be reliable stepping stones in subsequent iterations of the lead optimization process.

For example, in thrombin_A, different ensembles of ligands were prioritized depending on which data fusion scheme was used. The same three ligands were prioritized by each scheme; they were ranked within the top 2 ligands in most receptors and ranked no lower than 14th in every receptor. In addition, ligands thrombin_A_18 and thrombin_A_19 were prioritized by RankVotes, ParallelSelection, ReciprocalRank, and CompMedian; in at least one receptor, these two ligands were ranked within the top two ligands. However, because thrombin_A_18 and thrombin_A_19 were very poor binders in 1−2 receptors, ranking 86th or poorer, they were prioritized by neither SumRank nor ParetoRank. SumZScore prioritized thrombin_A_19, but replaced thrombin_A_18 with ligand thrombin_A_0, which ranked no lower than eighth in every receptor. Thrombin_A_11, thrombin_A_18, and thrombin_A_19 have similar experimental binding affinities, but thrombin_A_0 is 3.5 kcal/mol weaker, so in this case, RankVotes, ParallelSelection, ReciprocalRank, and CompMedian significantly outperformed the other data fusion methods.

As noted above, ParallelSelection seemed to be an outlier in the quality and consistency of its predictions compared with the other ranking schemes. In the context of lead optimization data fusion, this behavior is not unexpected. In this case, only five receptors are contributing information and so the ligands that are selected can be sensitive to the order in which the receptors are processed. By contrast, in virtual screening strategies in which ParallelSelection has been used successfully,[3c,33] the number of ligands selected is usually significantly larger than the number of pools from which ligands are selected. For example, instead of selecting five compounds from rankings within 5−15 receptors, virtual screening may select tens to hundreds of compounds from rankings from a handful of fingerprinting methods. Thus, ParallelSelection could be used to support ligand selections made by other data fusion approaches but should not be used as a primary strategy for identifying top binders in most lead optimization contexts.

Data Fusion Using Consensus among Decisions. In this third layer of data fusion, decisions from the individual composite scoring and/or consensus receptor rankings were combined to construct a single list of top five binders. As demonstrated in Figure 3C, for most data sets, there was significant agreement between the consensus scoring and the receptor ranking strategies and pooling all of the decisions ensured that the prioritized ligands were of comparable or improved quality relative to the average single-receptor predictions. The predictions based on the consensus of the receptor rankings decisions (ConsRecRank) are improved by more than one Z-Score unit compared with the average single-receptor predictions for two data sets and of comparable quality in the remaining data sets. For both the consensus of the consensus scores decisions (ConsRecScores) and the consensus of all nine data fusion decisions (ConsAll), there was one data set which demonstrated a systematic improvement of the average single-receptor results and three in which there was a degradation. The MEBA of the predictions based on the ConsAll were within 0.5 kcal/mol of the theoretical maximum for four data sets and within 1.0 kcal/mol for eight data sets.

Predictive Ability of SumZScore. Given its more direct connection with the MM/GBSA scoring function, as opposed to an integer rank from the consensus ranking or selection methods, and its robustness among the individual data fusion strategies and the decision-level data fusion strategies, the remaining analysis will focus on ligand prioritizations based on SumZScore. Figure 4 illustrates the quality of the top ligand predictions based on SumZScore. In two data sets, the MEBA for the SumZScore predictions were within 0.5 kcal/mol of the theoretical maximum MEBA and eight data sets were within 1.0 kcal/mol of the theoretical maximum. In ten data sets, the MEBA for the SumZScore predictions were the same or better than the MEBA of the five heaviest ligands. In most cases, the quality of the results are relatively insensitive to the number of compounds that are selected for prioritization as depicted in Figure 2 in the Supporting Information.

The correlation between the SumZScores and the experimental binding affinities for the ensemble of ligands is statistically significant for each data set with R^2 values ranging from 0.04 for chk1_D and thrombin_A to 0.58 for thrombin_B (see Figure 3 in the Supporting Information). The relatively low R^2 values for individual data sets, as depicted in Figure 5, does not intuitively convey the confidence that one can have in the ability of the binding affinity estimate to identify potent compounds among an ensemble of ligands. Specifically, there are seven data sets in which the R^2 value is between 0.09 and 0.29; in all but one of these data sets, the MEBA of the top ligand predictions are between 0.4 and 2.2 kcal/mol better than

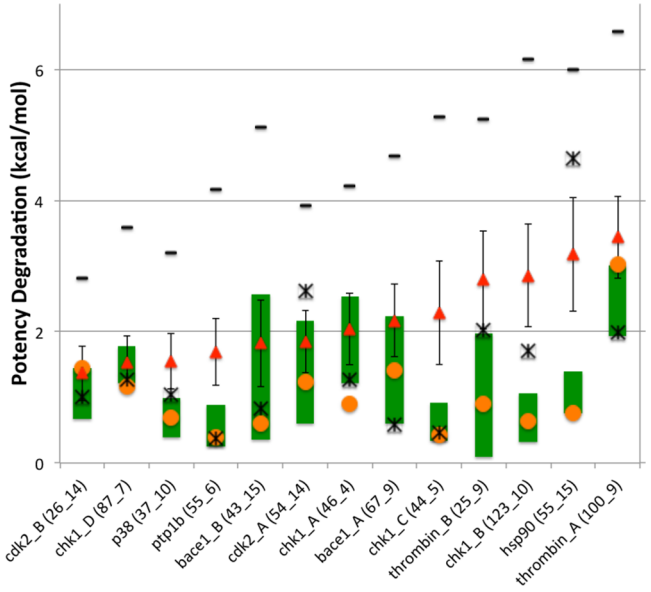


Figure 4. Quality of the five prioritized ligands identified by individual receptors (green bar indicates the range among individual receptors), data fusion SumZScore (orange circles), random selections (red triangles with black error bars), molecular weight (black stars), and five least potent inhibitors (black dashes) compared with the five most potent inhibitors, that is, potency degradation = MEBA(5 prioritized ligands) − MEBA(5 most potent inhibitors).

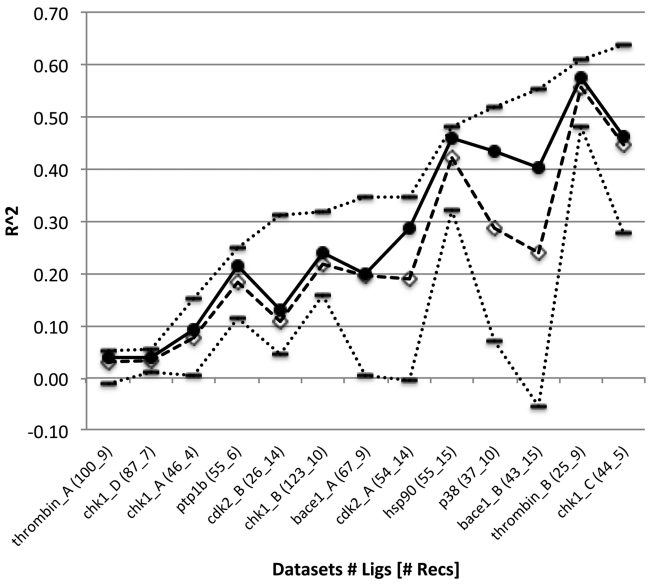


Figure 5. Correlation coefficients describing the overall agreement between experimental binding affinities and composite SumZScore (circles, solid line), MM/GBSA scores in individual receptors (bars, dotted lines), and the mean R^2 of the individual receptors (diamonds, dashed line).

the MEBA of the randomly selected compounds and within 1.2 kcal/mol of their respective theoretical maximum MEBA. Arguably, during the lead optimization phase of a drug discovery campaign, reliably identifying potent binders is of greater importance than ensuring that the detailed ordering among the mediocre candidates is correct. Thus, while R^2 is a commonly used metric to describe the predictive ability of modeling in a variety of contexts, we believe, in part based on this work, that it is not the best metric to be used in evaluating lead optimization strategies. Instead, metrics should be constructed that reflect the extent to which the computed results would facilitate good decisions, as captured by the relative difference between the MEBA of a set of prioritized compounds versus the MEBA of a random selection or an alternative selection scheme.

The data sets that showed little correlation between molecular weight and binding affinity provide examples in which compounds with similar molecular weight have significantly different measured binding affinities and where SumZScore effectively discriminated between their potencies. For example, in cdk2_A, the two heaviest compounds, cdk2_A_5f (MW = 639 g/mol) and cdk2_A_5p (MW = 588 g/mol), have experimental binding affinities of −5.5 and −8.9 kcal/mol, respectively, and SumZScores based on the MM/GBSA calculations correctly identified the better binder of the two with scores of 10.9 and 1.8, respectively, where smaller numbers indicate relatively more favorable compounds. Similarly, in the cdk2_B data set, the three heaviest compounds have molecular weights of 368−370 g/mol and binding affinities of −8.3, −9.2, and −10.5 kcal/mol; their respective SumZScores also predicted the correct ordering by potency with scores of 10.3, −6.8, and −16.2, respectively.

However, most of these publicly available data sets exhibit significant correlations between molecular weight and experimental binding affinities and some may argue that the success of the MM/GBSA calculations for these data sets may result more from capturing the molecular weight correlation than an inherent ability to describe the physics of the underlying protein−ligand interactions. Thus, in order to assess the benefit of MM/GBSA predictions in a larger number of contexts in which molecular weight does not provide critical information, additional data sets were constructed such that ligands that would be prioritized by molecular weight are comparable in potency to ligands that would be selected at random. Specifically, for each of the 13 data sets, the N ligands were sorted by molecular weight and all sliding windows of $N/2$ ligands were identified for which the MEBA by molecular weight is within 0.25 kcal/mol of the average MEBA for the random selections among this subset of ligands. Between 1 and 15 such subsets were identified for each data set for a total of 80 constructed data sets that exhibit similar experimental binding affinities for ligands predicted to be potent based on molecular weight and those chosen at random. The quality of the results based on MEBAs for the ligands prioritized by SumZScore compared to random selections are summarized in Table 4. In all but nine of these constructed data sets (12%), the MEBA for the top ligands predicted by SumZScore was comparable to or better than the MEBA from randomly selected ligands and almost half of all constructed data sets had MEBAs for the prioritized based on SumZScores that were 0.5 kcal/mol more potent than the MEBA for the randomly selected compounds. This increased potency among the prioritized ligands confirms that MM/GBSA predictions are not simply identifying high molecular weight compounds, but rather, they are providing additional information that would enhance decision-making during the lead optimization stage of a structure-based drug discovery campaign.

Supervised Data Fusion. An alternative approach to data fusion when multiple crystallographic structures are available involves identifying a "representative" or preferred receptor structure and using that structure as the basis of modeling subsequent congeneric ligands. For example, consensus

Table 4. Quality of Predictions for Constructed Data Sets[a]

target	no. constructed data sets[a]	no. ligs ($N/2$)	% constructed data sets with ⟨MEBA(random)⟩ − MEBA(SumZScore) <−0.5 kcal/mol	<−0.25 kcal/mol	>0.25 kcal/mol	>0.5 kcal/mol	⟨⟨MEBA(random)⟩⟩[b] − ⟨MEBA(max)⟩ (kcal/mol)
bace1_A	6	33	0	0	50	0	1.4
bace1_B	4	21	0	0	100	25	1.1
cdk2_A	9	27	0	0	100	44	1.5
cdk2_B	4	13	25	100	0	0	1.1
chk1_A	4	23	0	0	50	25	1.2
chk1_B	15	61	0	0	100	100	2.4
chk1_C	1	22	0	0	100	100	1.0
chk1_D	13	44	31	31	8	0	1.5
hsp90	3	27	0	0	100	100	2.1
p38_MAP kinase	2	19	0	0	50	50	1.0
ptp1b	5	27	20	20	20	20	0.9
thrombin_A	9	50	0	0	78	78	1.9
thrombin_B	2	13	0	0	100	100	1.7
total	80		8	12	64	47	

[a]Data sets were constructed such that there are comparable potencies for the subset of ligands prioritized based on molecular weight and those selected at random. For each data set, the N ligands were sorted by molecular weight and all sliding windows of $N/2$ ligands were evaluated. Each sliding window for which the MEBA by molecular weight was within 0.25 kcal/mol of the average MEBA for the random selections among this subset of ligands became a "constructed data set"; that is, for each constructed data set of $N/2$ ligands, $|\langle \text{MEBA(random)} \rangle - \text{MEBA(MW)}| < 0.25$ kcal/mol. [b]The average of 10 000 random selections from the constructed data set.

receptors have been constructed for induced fit virtual screening protocols[34] and single structures have been identified that reliably reproduce relative binding affinities for a training set of ligands and then used for modeling new candidate compounds. To assess the latter approach, each of the 13 data sets were divided into training and test set ligands such that the former set contained all the crystallographic ligands and the latter set contained the remaining ligands in the congeneric series. As depicted in Figure 6, for most data sets, the "best" single receptor identified from the training set data (i.e., the receptor which has the best agreement between MM/GBSA score and experimental binding affinity) prioritizes ligands in the test set that are of comparable potency to those identified using SumZScore across all the receptors. For thrombin_A, the supervised selection of a "best" single receptor identified a subset of ligands that was more potent by 0.7 kcal/mol than the SumZScore predictions. However, for thrombin_B and chk1_A, the "best" single receptor yielded subsets of ligands that were 1.3 to 1.6 kcal/mol less potent than the corresponding predictions using SumZScore over all receptors. Thus, explicitly using information from all receptors ensures more robust predictions across the data sets than the computationally more expedient strategy of selecting a single "best" receptor from known ligands to model new compounds in the series.

Recent studies have demonstrated the utility of approximating binding pocket diversity for in silico calculations by selecting representative conformations from clustered molecular dynamics trajectories.[3e] Thus, in the absence of multiple crystallographic structures for a given biological target, this strategy provides a promising alternative for obtaining multiple MM/GBSA scores and rankings for use in data fusion methods. However, crystallographic structures represent experimentally validated low-energy conformations and so should be considered when available.

Comparison between Prioritizations Based on GlideSP and MM/GBSA. The average potency of prioritized ligands is marginally better for those ligands that are selected by their MM/GBSA-based SumZScore compared with those selected by their GlideSP-based SumZScore. In no data sets do GlideSP and MM/GBSA SumZScores identify the same set of top five ligands. In three data sets GlideSP and MM/GBSA SumZScores identify four of the same top ligands and in five data sets GlideSP and MM/GBSA SumZScores prioritize mutually exclusive sets of ligands. These findings are consistent

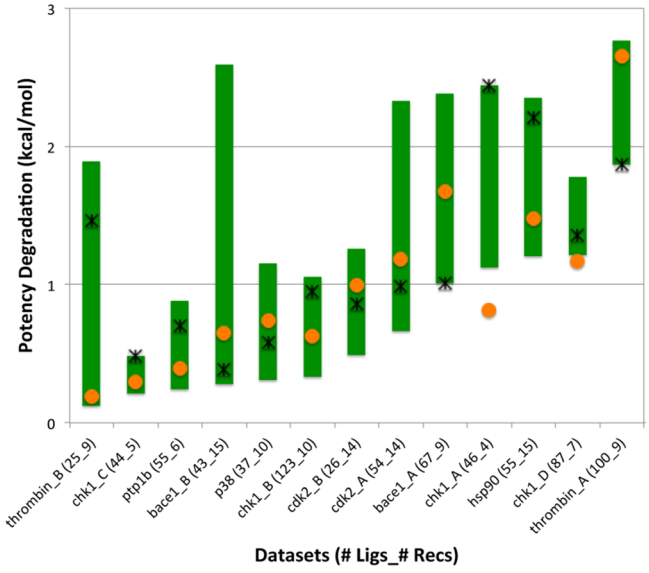


Figure 6. Quality of the five prioritized ligands among the test set compounds (i.e., the noncrystallographic ligands in the data sets) identified by individual receptors (green bar indicates the range among individual receptors), data fusion SumZScore (orange circles), and the single receptor that performed the best among the training set compounds (i.e., the crystallographic ligands) (black stars) compared with the five most potent inhibitors in the test set, that is, potency degradation = MEBA(5 prioritized ligands in test set) − MEBA(5 most potent inhibitors in test set).

with those reported by Brooijmans and Humblet,[1b] who recently showed in the context of virtual screening ensembles that GlideSP and MM/GBSA were complementary and identified different subsets of top ligands.

Rather than trying to reconcile the ligand prioritizations by these two scoring functions, combining these different prioritizations could serve to minimize the overall error of the predictions. For example, Tumer and Ghosh have shown that combining positively correlated classifiers only slightly reduces errors in the data-fused ensemble relative to the individual errors, uncorrelated classifiers diminish the ensemble error while negatively correlated classifiers reduce the ensemble error further.[35] Thus, in the current scenario, the complementary nature of these scoring functions opens up another avenue for data fusion and is the subject of ongoing investigation in our research group.

CONCLUSIONS

To the best of our knowledge, this study represents the largest systematic investigation to date into the advantages of applying data fusion strategies to end-point lead optimization calculations. Modeling 13 large and carefully curated retrospective data sets was facilitated by a recently developed automated Lead Optimization workflow that involved ligand conformational sampling, docking, pose refinement, and scoring. Across the data sets, most of the individual receptors successfully identified potent inhibitors among an ensemble of congeneric ligands based on MM/GBSA scores. However, data fusion strategies based on consensus scores and receptor rankings that combined information from each of the receptors increased the reliability of the predictions.

The consensus SumZScore provided a physically meaningful metric and demonstrated a quality of its predictions that were relatively consistent with other data fusion strategies, though will tend to offer more conservative prioritized list of compounds than RankVotes, ParallelSelection, CompMedian, and ReciprocalRank, for example. Perhaps most surprisingly, even with relatively low to modest correlations between the SumZScores and the experimental binding affinities in each of the data sets, SumZScore predictions reliably identify potent inhibitors. In all but one data set, SumZScore predictions were comparable to or better than the single-receptor predictions and random selections and within 1.2 kcal/mol of the theoretical maximum. Furthermore, in all but one data set, without any prior information, SumZScore prioritized a subset of ligands that were at least as potent as those that were prioritized from a single receptor that was identified to be the best performer on a known set of compounds.

Other more sophisticated supervised strategies may ensure that potent inhibitors from an ensemble of candidate molecules are identified efficiently and accurately in lead optimization. In addition, decomposing the MM/GBSA scores for known compounds and analyzing the contributions from individual energy terms using data fusion methods should provide additional insight into the target- and series-specific chemical driving forces of binding and guide medicinal chemists in their design strategies. Leveraging information from known data to judiciously select a subset of receptors, analyzing individual energy components, and combining weighted scores from complementary sampling methods and scoring functions are subjects of our ongoing investigations.

These data fusion methods that combine information from modeling congeneric ligands in multiple receptor provide a robust approach for ensuring high quality predictions of the most potent ligands among candidate compounds. Given the well-known limitations to end-point methods, the list of prioritized compounds could then be explored by more computationally intensive methods, such as free energy perturbation, to obtain more accurate relative binding affinity estimates within the congeneric series prior to determining the final list of compounds for synthesis and testing.

ASSOCIATED CONTENT

Supporting Information

Tables and figures describing the data set characteristics and analysis of the sensitivity of the data fusion ligand prioritizations. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Email: Jennifer.Knight@schrodinger.com.

Author Contributions

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

Notes

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

ACKNOWLEDGMENTS

We thank Prof. Gabriela Chiosis and Dr. Pallav D. Patel from the Memorial Sloan Kettering Cancer Center, who provided updated experimental binding affinity measurements from the fluorescence polarization assay for the hsp90 data set.

REFERENCES

(1) (a) Gilson, M. K.; Zhou, H.-X. Calculation of protein−ligand binding affinities. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 21−42. (b) Brooijmans, N.; Humblet, C. Chemical space sampling by different scoring functions and crystal structures. *J. Comput.-Aided Mol. Des.* **2010**, *24* (5), 433−447.

(2) (a) Kirchmair, J.; Distinto, S.; Markt, P.; Schuster, D.; Spitzer, G. M.; Liedl, K. R.; Wolber, G. How to optimize shape-based virtual screening: Choosing the right query and including chemical information. *J. Chem. Inf. Model.* **2009**, *49* (3), 678−692. (b) Sastry, G. M.; Dixon, S. L.; Sherman, W. Rapid shape-based ligand alignment and virtual screening method based on atom/feature-pair similarities and volume overlap scoring. *J. Chem. Inf. Model.* **2011**, *51* (10), 2455−2466. (c) Noha, S. M.; Atanasov, A. G.; Schuster, D.; Markt, P.; Fakhrudin, N.; Heiss, E. H.; Schrammel, O.; Rollinger, J. M.; Stuppner, H.; Dirsch, V. M.; Wolber, G. Discovery of a novel IKK-beta inhibitor by ligand-based virtual screening techniques. *Bioorg. Med. Chem. Lett.* **2011**, *21* (1), 577−583. (d) LaLonde, J.; Elban, M.; Courter, J.; Sugawara, A.; Soeta, T.; Madani, N.; Princiotto, A.; Kwon, Y.; Kwong, P.; Schön, A. Design, synthesis, and biological evaluation of small molecule inhibitors of CD4-gp120 binding based on virtual screening. *Bioorg. Med. Chem.* **2011**, *19* (1), 91−101. (e) Badrinarayan, P.; Sastry, G. N. Virtual high throughput screening in new lead identification. *Comb. Chem. High Throughput Screening* **2011**, *14* (10), 840−860. (f) Schneider, G. Virtual screening: An endless staircase? *Nat. Rev. Drug Discovery* **2010**, *9* (4), 273−276.

(3) (a) Feher, M. Consensus scoring for protein−ligand interactions. *Drug Discovery Today* **2006**, *11* (9−10), 421−428. (b) Willett, P. Combination of similarity rankings using data fusion. *J. Chem. Inf. Model.* **2013**, *53* (1), 1−10. (c) Svensson, F.; Karlén, A.; Sköld, C.

Virtual screening data fusion using both structure- and ligand-based methods. *J. Chem. Inf. Model.* **2012**, *52* (1), 225−232. (d) Plewczynski, D. Brainstorming: Weighted voting prediction of inhibitors for protein targets. *J. Mol. Model.* **2011**, *17* (9), 2133−2141. (e) Osguthorpe, D. J.; Sherman, W.; Hagler, A. T. Generation of receptor structural ensembles for virtual screening using binding site shape analysis and clustering. *Chem. Biol. Drug Des.* **2012**, *80* (2), 182−193. (f) Osguthorpe, D. J.; Sherman, W.; Hagler, A. T. Exploring protein flexibility: Incorporating structural ensembles from crystal structures and simulation into virtual screening protocols. *J. Phys. Chem. B* **2012**, *116* (23), 6952−6959.

(4) (a) Tominaga, Y.; Jorgensen, W. L. General model for estimation of the inhibition of protein kinases using Monte Carlo simulations. *J. Med. Chem.* **2004**, *47* (10), 2534−2549. (b) Aqvist, J.; Medina, C.; Samuelsson, J. E. A new method for predicting binding affinity in computer-aided drug design. *Protein Eng.* **1994**, *7* (3), 385−391. (c) Hansson, T.; Marelius, J.; Aqvist, J. Ligand binding affinity prediction by linear interaction energy methods. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 27−35. (d) Aqvist, J.; Marelius, J. The linear interaction energy method for predicting ligand binding free energies. *Comb. Chem. High Throughput Screening* **2001**, *4* (8), 613−626. (e) Wallin, G.; Nervall, M.; Carlsson, J.; Aqvist, J. Charges for large scale binding free energy calculations with the linear interaction energy method. *J. Chem. Theory Comput.* **2009**, *5* (2), 380−395.

(5) (a) Kollman, P.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W. Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Acc. Chem. Res.* **2000**, *33* (12), 889−897. (b) Massova, I.; Kollman, P. Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. *Perspect. Drug Discovery Des.* **2000**, *18* (1), 113−135.

(6) Kuhn, B.; Gerber, P.; Schulz-Gasch, T.; Stahl, M. Validation and use of the MM-PBSA approach for drug discovery. *J. Med. Chem.* **2005**, *48* (12), 4040−4048.

(7) Thompson, D. C.; Humblet, C.; Joseph-Mccarthy, D. Investigation of MM-PBSA rescoring of docking poses. *J. Chem. Inf. Model.* **2008**, *48* (5), 1081−1091.

(8) Greenidge, P. A.; Kramer, C.; Mozziconacci, J.-C.; Wolf, R. M. MM/GBSA binding energy prediction on the PDBbind data set: Successes, failures, and directions for further improvement. *J. Chem. Inf. Model.* **2013**, *53* (1), 201−209.

(9) (a) Ravindranathan, K.; Tirado-Rives, J.; Jorgensen, W. L.; Guimarães, C. R. W. Improving MM-GB/SA scoring through the application of the variable dielectric model. *J. Chem. Theory Comput.* **2011**, *7* (12), 3859−3865. (b) Hou, T.; Wang, J.; Li, Y.; Wang, W. Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *J. Chem. Theory Comput.* **2011**, *51* (1), 69−82. (c) Zhu, K.; Shirts, M.; Friesner, R. Improved methods for side chain and loop predictions via the protein local optimization program: Variable dielectric model for implicitly improving the treatment of polarization effects. *J. Chem. Theory Comput.* **2007**, *3* (6), 2108−2119. (d) Li, J.; Abel, R.; Zhu, K.; Cao, Y.; Zhao, S.; Friesner, R. A. The VSGB 2.0 model: A next generation energy model for high resolution protein structure modeling. *Proteins* **2011**, *79*, 2794−2812.

(10) Yang, C.-Y.; Sun, H.; Chen, J.; Nikolovska-Coleska, Z.; Wang, S. Importance of ligand reorganization free energy in protein−ligand binding-affinity prediction. *J. Am. Chem. Soc.* **2009**, *131* (38), 13709−13721.

(11) (a) Abel, R.; Young, T.; Farid, R.; Berne, B. J.; Friesner, R. A. Role of the active-site solvent in the thermodynamics of factor Xa ligand binding. *J. Am. Chem. Soc.* **2008**, *130* (9), 2817−2831. (b) Young, T.; Abel, R.; Kim, B.; Berne, B. J.; Friesner, R. A. Motifs for molecular recognition exploiting hydrophobic enclosure in protein-ligand binding. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104* (3), 808−813.

(12) Srivastava, H. K.; Sastry, G. N. Molecular dynamics investigation on a series of HIV protease inhibitors: Assessing the performance of MM-PBSA and MM-GBSA approaches. *J. Chem. Inf. Model.* **2012**, *52* (11), 3088−3098.

(13) (a) Maffucci, I.; Contini, A. Explicit ligand hydration shells improve the correlation between MM-PB/GBSA binding energies and experimental activities. *J. Chem. Theory Comput.* **2013**, *9*, 2706. (b) Checa, A.; Ortiz, A. R.; de Pascual-Teresa, B.; Gago, F. Assessment of solvation effects on calculated binding affinity differences: Trypsin inhibition by flavonoids as a model system for congeneric series. *J. Med. Chem.* **1997**, *40* (25), 4136−4145.

(14) Rastelli, G.; Rio, A.; Degliesposti, G.; Sgobba, M. Fast and accurate predictions of binding free energies using MM-PBSA and MM-GBSA. *J. Comput. Chem.* **2010**, *31* (4), 797−810.

(15) (a) Ruvinsky, A. M.; Kozintsev, A. V. New and fast statistical-thermodynamic method for computation of protein−ligand binding entropy substantially improves docking accuracy. *J. Comput. Chem.* **2005**, *26* (11), 1089−1095. (b) Ruvinsky, A. M. Role of binding entropy in the refinement of protein−ligand docking predictions: Analysis based on the use of 11 scoring functions. *J. Comput. Chem.* **2007**, *28* (8), 1364−1372. (c) Chang, C.-E.; Gilson, M. K. Free energy, entropy, and induced fit in host-guest recognition: Calculations with the second-generation mining minima algorithm. *J. Am. Chem. Soc.* **2004**, *126* (40), 13156−13164. (d) Chen, W.; Gilson, M. K.; Webb, S. P.; Potter, M. J. Modeling protein−ligand binding by mining minima. *J. Chem. Theory Comput.* **2010**, *6* (11), 3540−3557.

(16) (a) Genheden, S.; Kuhn, O.; Mikulskis, P.; Hoffmann, D.; Ryde, U. The normal-mode entropy in the MM/GBSA method: Effect of system truncation, buffer region, and dielectric constant. *J. Chem. Inf. Model.* **2012**, *52* (8), 2079−2088. (b) Kongsted, J.; Ryde, U. An improved method to predict the entropy term with the MM/PBSA approach. *J. Comput.-Aided Mol. Des.* **2009**, *23* (2), 63−71.

(17) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.

(18) (a) *Protein Preparation Wizard*, Schrödinger Suite; Epik version 2.5; Impact version 6.0; Prime version 3.3; Schrödinger, LLC: New York, 2013; (b) Sastry, G. M.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J. Comput.-Aided Mol. Des.* **2013**, *27* (3), 221−234.

(19) *Maestro, version 9.3*; Schrödinger, LLC: New York, 2013.

(20) (a) Bell, J. A.; Ho, K. L.; Farid, R. Significant reduction in errors associated with nonbonded contacts in protein crystal structures: Automated all-atom refinement with PrimeX. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **2012**, *68* (Pt 8), 935−952. (b) *Small-Molecule Drug Discovery Suite 2013−3: PrimeX, version 2.2*, Schrödinger, LLC, New York, NY, 2013;. (c) Bell, J. A.; Cao, Y.; Gunn, J. R.; Day, T.; Gallicchio, E.; Zhou, Z.; Levy, R.; Farid, R. PrimeX and the Schrödinger Computational Chemistry Suite of Programs. *International Tables for Crystallography* **2012**, *F* (18), 534−538.

(21) *Schrödinger Release 2013-3: Force Field Builder, version 1.3*; Schrödinger, LLC: New York, 2013.

(22) (a) Shivakumar, D.; Williams, J.; Wu, Y.; Damm, W.; Shelley, J.; Sherman, W. Prediction of Absolute Solvation Free Energies using Molecular Dynamics Free Energy Perturbation and the OPLS Force Field. *J. Chem. Theory Comput.* **2010**, *6* (5), 1509−1519. (b) *Schrödinger Release 2013-3: OPLS2.1*; Schrödinger, LLC: New York, 2013.

(23) (a) *Schrödinger Release 2013-3: ConfGen, version 2.6*, Schrödinger, LLC, New York, 2013; (b) Watts, K. S.; Dalal, P.; Murphy, R. B.; Sherman, W.; Friesner, R. A.; Shelley, J. C. ConfGen: A conformational search method for efficient generation of bioactive conformers. *J. Chem. Inf. Model.* **2010**, *50* (4), 534−546.

(24) *Schrödinger Release 2013-3: MacroModel, version 10.2*; Schrödinger, LLC: New York, 2013.

(25) (a) Jorgensen, W. L.; Tirado-Rives, J. The OPLS force field for proteins. Energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* **1988**, *110*, 1657−1666. (b) Kaminski, G.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation and reparametrization of the OPLS-AA force field for proteins via

comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B* **2001**, *105*, 6474−6487.
(26) (a) Halgren, T. A. Merck molecular force field. II. MMFF94 van der Waals and electrostatic parameters for intermolecular interactions. *J. Comput. Chem.* **1998**, *17* (5−6), 520−552. (b) Halgren, T. A. MMFF VII. Characterization of MMFF94, MMFF94s, and other widely available force fields for conformational energies and for intermolecular-interaction energies and geometries. *J. Comput. Chem.* **1999**, *20* (7), 730−748.
(27) (a) *Small-Molecule Drug Discovery Suite 2013-3: Glide, version 6.1*, Schrödinger, LLC: New York, 2013; (b) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: A new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* **2004**, *47*, 1750−1759. (c) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shaw, D. E.; Shelley, M.; Perry, J. K.; Francis, P.; Shenkin, P. S. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004**, *47*, 1739−1749.
(28) (a) *Schrödinger Release 2013-3: Prime, version 3.4*, Schrödinger, LLC: New York, 2013; (b) Jacobson, M. P.; Pincus, D. L.; Rapp, C. S.; Day, T. J.; Honig, B.; Shaw, D. E.; Friesner, R. A. A hierarchical approach to all-atom protein loop prediction. *Proteins* **2004**, *55*, 351−367. (c) Jacobson, M. P.; Friesner, R. A.; Xiang, Z.; Honig, B. On the role of crystal packing forces in determining protein sidechain conformations. *J. Mol. Biol.* **2002**, *320*, 597−608.
(29) Borrelli, K. W.; Cossins, B.; Guallar, V. Exploring hierarchical refinement techniques for induced fit docking with protein and ligand flexibility. *J. Comput. Chem.* **2010**, *31* (6), 1224−1235.
(30) Fleiss, J. L. Measuring nominal scale agreement among many raters. *Psychol. Bull.* **1971**, *76* (5), 378−382.
(31) Landis, J. R.; Koch, G. G. The measurement of observer agreement for categorical data. *Biometrics* **1977**, *33* (1), 159−174.
(32) Kittler, J. Combining classifiers: A theoretical framework. *Pattern Anal. Appl.* **1998**, *1* (1), 18−27.
(33) Sastry, G. M.; Inakollu, V. S. S.; Sherman, W. Boosting virtual screening enrichments with data fusion: Coalescing hits from two-dimensional fingerprints, shape, and docking. *J. Chem. Inf. Model.* **2013**, *53* (7), 1531−1542.
(34) Kalid, O.; Toledo Warshaviak, D.; Shechter, S.; Sherman, W.; Shacham, S. Consensus Induced Fit Docking (cIFD): Methodology, validation, and application to the discovery of novel Crm1 inhibitors. *J. Comput.-Aid. Mol. Des.* **2012**, *26* (11), 1217−1228.
(35) Kuncheva, L.; Whitaker, C. Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. *Mach. Learn.* **2003**, *51* (2), 181−207.
(36) (a) Rueeger, H.; Lueoend, R.; Rogel, O.; Rondeau, J.-M.; Möbitz, H.; Machauer, R.; Jacobson, L.; Staufenbiel, M.; Desrayaud, S.; Neumann, U. Discovery of cyclic sulfone hydroxyethylamines as potent and selective β-site APP-cleaving enzyme 1 (BACE1) inhibitors: Structure-based design and in vivo reduction of amyloid β-peptides. *J. Med. Chem.* **2012**, *55* (7), 3364−3386. (b) Rueeger, H.; Rondeau, J.-M.; Mccarthy, C.; Möbitz, H.; Tintelnot-Blomley, M.; Neumann, U.; Desrayaud, S. Structure based design, synthesis, and SAR of cyclic hydroxyethylamine (HEA) BACE-1 inhibitors. *Bioorg. Med. Chem. Lett.* **2011**, *21* (7), 1942−1947.
(37) (a) Stamford, A. W.; Scott, J. D.; Li, S. W.; Babu, S.; Tadesse, D.; Hunter, R.; Wu, Y.; Misiaszek, J.; Cumming, J. N.; Gilbert, E. J.; Huang, C.; Mckittrick, B. A.; Hong, L.; Guo, T.; Zhu, Z.; Strickland, C.; Orth, P.; Voigt, J. H.; Kennedy, M. E.; Chen, X.; Kuvelkar, R.; Hodgson, R.; Hyde, L. A.; Cox, K.; Favreau, L.; Parker, E. M.; Greenlee, W. J. Discovery of an orally available, brain penetrant BACE1 inhibitor that affords robust CNS Aβ reduction. *ACS Med. Chem. Lett.* **2012**, *3* (11), 897−902. (b) Mandal, M.; Zhu, Z.; Cumming, J. N.; Liu, X.; Strickland, C.; Mazzola, R. D.; Caldwell, J. P.; Leach, P.; Grzelak, M.; Hyde, L.; Zhang, Q.; Terracina, G.; Zhang, L.; Chen, X.; Kuvelkar, R.; Kennedy, M. E.; Favreau, L.; Cox, K.; Orth, P.; Buevich, A.; Voigt, J.; Wang, H.; Kazakevich, I.; Mckittrick, B. A.; Greenlee, W.; Parker, E. M.; Stamford, A. W. Design and validation of bicyclic iminopyrimidinones as β amyloid cleaving enzyme-1 (BACE1) inhibitors: conformational constraint to favor a bioactive conformation. *J. Med. Chem.* **2012**, *55* (21), 9331−9345.
(38) (a) Griffin, R. J.; Henderson, A.; Curtin, N. J.; Echalier, A.; Endicott, J. A.; Hardcastle, I. R.; Newell, D. R.; Noble, M. E. M.; Wang, L.-Z.; Golding, B. T. Searching for cyclin-dependent kinase inhibitors using a new variant of the cope elimination. *J. Am. Chem. Soc.* **2006**, *128* (18), 6012−6013. (b) Hardcastle, I. R.; Arris, C. E.; Bentley, J.; Boyle, F. T.; Chen, Y.; Curtin, N. J.; Endicott, J. A.; Gibson, A. E.; Golding, B. T.; Griffin, R. J.; Jewsbury, P.; Menyerol, J.; Mesguiche, V.; Newell, D. R.; Noble, M. E. M.; Pratt, D. J.; Wang, L.-Z.; Whitfield, H. J. N2-substituted O6-cyclohexylmethylguanine derivatives: Potent inhibitors of cyclin-dependent kinases 1 and 2. *J. Med. Chem.* **2004**, *47* (15), 3710−3722.
(39) Wang, S.; Meades, C.; Wood, G.; Osnowski, A.; Anderson, S.; Yuill, R.; Thomas, M.; Mezna, M.; Jackson, W.; Midgley, C.; Griffiths, G.; Fleming, I.; Green, S.; McNae, I.; Wu, S.-Y.; McInnes, C.; Zheleva, D.; Walkinshaw, M. D.; Fischer, P. M. 2-Anilino-4-(thiazol-5-yl)pyrimidine CDK inhibitors: Synthesis, SAR analysis, X-ray crystallography, and biological activity. *J. Med. Chem.* **2004**, *47* (7), 1662−1675.
(40) Fraley, M. E.; Steen, J. T.; Brnardic, E. J.; Arrington, K. L.; Spencer, K. L.; Hanney, B. A.; Kim, Y.; Hartman, G. D.; Stirdivant, S. M.; Drakas, B. A.; Rickert, K.; Walsh, E. S.; Hamilton, K.; Buser, C. A.; Hardwick, J.; Tao, W.; Beck, S. C.; Mao, X.; Lobell, R. B.; Sepp-Lorenzino, L.; Yan, Y.; Ikuta, M.; Munshi, S. K.; Kuo, L. C.; Kreatsoulas, C. 3-(Indol-2-yl)indazoles as Chek1 kinase inhibitors: Optimization of potency and selectivity via substitution at C6. *Bioorg. Med. Chem. Lett.* **2006**, *16* (23), 6049−6053.
(41) (a) Tong, Y.; Claiborne, A.; Pyzytulinska, M.; Tao, Z.-F.; Stewart, K. D.; Kovar, P.; Chen, Z.; Credo, R. B.; Guan, R.; Merta, P. J.; Zhang, H.; Bouska, J.; Everitt, E. A.; Murry, B. P.; Hickman, D.; Stratton, T. J.; Wu, J.; Rosenberg, S. H.; Sham, H. L.; Sowin, T. J.; Lin, N.-H. 1,4-Dihydroindeno[1,2-c]pyrazoles as potent checkpoint kinase 1 inhibitors: Extended exploration on phenyl ring substitutions and preliminary ADME/PK studies. *Bioorg. Med. Chem. Lett.* **2007**, *17* (13), 3618−3623. (b) Tao, Z.-F.; Li, G.; Tong, Y.; Stewart, K. D.; Chen, Z.; Bui, M.-H.; Merta, P.; Park, C.; Kovar, P.; Zhang, H.; Sham, H. L.; Rosenberg, S. H.; Sowin, T. J.; Lin, N.-H. Discovery of 4′-(1,4-dihydro-indeno[1,2-c]pyrazol-3-yl)-benzonitriles and 4′-(1,4-dihydro-indeno[1,2-c]pyrazol-3-yl)-pyridine-2′-carbonitriles as potent checkpoint kinase 1 (Chk1) inhibitors. *Bioorg. Med. Chem. Lett.* **2007**, *17* (21), 5944−5951.
(42) Wang, L.; Sullivan, G. M.; Hexamer, L. A.; Hasvold, L. A.; Thalji, R.; Przytulinska, M.; Tao, Z.-F.; Li, G.; Chen, Z.; Xiao, Z.; Gu, W.-Z.; Xue, J.; Bui, M.-H.; Merta, P.; Kovar, P.; Bouska, J. J.; Zhang, H.; Park, C.; Stewart, K. D.; Sham, H. L.; Sowin, T. J.; Rosenberg, S. H.; Lin, N.-H. Design, synthesis, and biological activity of 5,10-dihydro-dibenzo[b,e][1,4]diazepin-11-one-based potent and selective Chk-1 inhibitors. *J. Med. Chem.* **2007**, *50* (17), 4162−4176.
(43) (a) Tao, Z.-F.; Wang, L.; Stewart, K. D.; Chen, Z.; Gu, W.; Bui, M.-H.; Merta, P.; Zhang, H.; Kovar, P.; Johnson, E.; Park, C.; Judge, R.; Rosenberg, S.; Sowin, T.; Lin, N.-H. Structure-based design, synthesis, and biological evaluation of potent and selective macrocyclic checkpoint kinase 1 inhibitors. *J. Med. Chem.* **2007**, *50* (7), 1514−1527. (b) Li, G.; Hasvold, L. A.; Tao, Z.-F.; Wang, G. T.; Gwaltney, S. L.; Patel, J.; Kovar, P.; Credo, R. B.; Chen, Z.; Zhang, H.; Park, C.; Sham, H. L.; Sowin, T.; Rosenberg, S. H.; Lin, N.-H. Synthesis and biological evaluation of 1-(2,4,5-trisubstituted phenyl)-3-(5-cyanopyrazin-2-yl)ureas as potent Chk1 kinase inhibitors. *Bioorg. Med. Chem. Lett.* **2006**, *16* (8), 2293−2298.
(44) (a) He, H.; Zatorska, D.; Kim, J.; Aguirre, J.; Llauger, L.; She, Y.; Wu, N.; Immormino, R. M.; Gewirth, D. T.; Chiosis, G. Identification of potent water soluble purine-scaffold inhibitors of the heat shock protein 90. *J. Med. Chem.* **2006**, *49* (1), 381−390. (b) Llauger, L.; He, H.; Kim, J.; Aguirre, J.; Rosen, N.; Peters, U.; Davies, P.; Chiosis, G. Evaluation of 8-arylsulfanyl, 8-arylsulfoxyl, and 8-arylsulfonyl adenine derivatives as inhibitors of the heat shock protein 90. *J. Med. Chem.* **2005**, *48* (8), 2892−2905. (c) Patel, P. D.; Yan, P.; Seidler, P. M.;

Patel, H. J.; Sun, W.; Yang, C.; Que, N. S.; Taldone, T.; Finotti, P.; Stephani, R. A.; Gewirth, D. T.; Chiosis, G. Paralog-selective Hsp90 inhibitors define tumor-specific regulation of HER2. *Nat. Chem. Biol.* **2013**, *9* (11), 677−684.

(45) Goldstein, D. M.; Soth, M.; Gabriel, T.; Dewdney, N.; Kuglstatter, A.; Arzeno, H.; Chen, J.; Bingenheimer, W.; Dalrymple, S. A.; Dunn, J.; Farrell, R.; Frauchiger, S.; La Fargue, J.; Ghate, M.; Graves, B.; Hill, R. J.; Li, F.; Litman, R.; Loe, B.; McIntosh, J.; McWeeney, D.; Papp, E.; Park, J.; Reese, H. F.; Roberts, R. T.; Rotstein, D.; San Pablo, B.; Sarma, K.; Stahl, M.; Sung, M.-L.; Suttman, R. T.; Sjogren, E. B.; Tan, Y.; Trejo, A.; Welch, M.; Weller, P.; Wong, B. R.; Zecic, H. Discovery of 6-(2,4-difluorophenoxy)-2-[3-hydroxy-1-(2-hydroxyethyl)propylamino]-8-methyl-8H-pyrido[2,3-d]pyrimidin-7-one (pamapimod) and 6-(2,4-difluorophenoxy)-8-methyl-2-(tetrahydro-2H-pyran-4-ylamino)pyrido[2,3-d]pyrimidin-7(8H)-one (R1487) as orally bioavailable and highly selective inhibitors of p38α mitogen-activated protein kinase. *J. Med. Chem.* **2011**, *54* (7), 2255−2265.

(46) (a) Wilson, D. P.; Wan, Z.-K.; Xu, W.-X.; Kirincich, S. J.; Follows, B. C.; Joseph-Mccarthy, D.; Foreman, K.; Moretto, A.; Wu, J.; Zhu, M.; Binnun, E.; Zhang, Y.-L.; Tam, M.; Erbe, D. V.; Tobin, J.; Xu, X.; Leung, L.; Shilling, A.; Tam, S. Y.; Mansour, T. S.; Lee, J. Structure-based optimization of protein tyrosine phosphatase 1B inhibitors: From the active site to the second phosphotyrosine binding site. *J. Med. Chem.* **2007**, *50* (19), 4681−4698. (b) Wan, Z.-K.; Lee, J.; Hotchandani, R.; Moretto, A.; Binnun, E.; Wilson, D. P.; Kirincich, S. J.; Follows, B. C.; Ipek, M.; Xu, W.; Joseph-Mccarthy, D.; Zhang, Y.-L.; Tam, M.; Erbe, D. V.; Tobin, J. F.; Li, W.; Tam, S. Y.; Mansour, T. S.; Wu, J. Structure-based optimization of protein tyrosine phosphatase-1 B inhibitors: Capturing interactions with arginine 24. *ChemMedChem* **2008**, *3* (10), 1525−1529.

(47) (a) Fevig, J. M.; Wexler, R. R. Chapter 9: Anticoagulants: Thrombin and factor Xa inhibitors. *Annu. Rev. Med. Chem.* **1999**, 1−12. (b) Burgey, C. S.; Robinson, K. A.; Lyle, T. A.; Nantermet, P. G.; Selnick, H. G.; Isaacs, R. C. A.; Lewis, S. D.; Lucas, B. J.; Krueger, J. A.; Singh, R.; Miller-Stein, C.; White, R. B.; Wong, B.; Lyle, E. A.; Stranieri, M. T.; Cook, J. J.; McMasters, D. R.; Pellicore, J. M.; Pal, S.; Wallace, A. A.; Clayton, F. C.; Bohn, D.; Welsh, D. C.; Lynch, J. J.; Yan, Y.; Chen, Z.; Kuo, L.; Gardell, S. J.; Shafer, J. A.; Vacca, J. P. Pharmacokinetic optimization of 3-amino-6-chloropyrazinone acetamide thrombin inhibitors. Implementation of P3 pyridine N-oxides to deliver an orally bioavailable series containing P1 N-benzylamides. *Bioorg. Med. Chem. Lett.* **2003**, *13* (7), 1353−1357. (c) Burgey, C. S.; Robinson, K. A.; Lyle, T. A.; Sanderson, P. E. J.; Lewis, S. D.; Lucas, B. J.; Krueger, J. A.; Singh, R.; Miller-Stein, C.; White, R. B.; Wong, B.; Lyle, E. A.; Williams, P. D.; Coburn, C. A.; Dorsey, B. D.; Barrow, J. C.; Stranieri, M. T.; Holahan, M. A.; Sitko, G. R.; Cook, J. J.; McMasters, D. R.; McDonough, C. M.; Sanders, W. M.; Wallace, A. A.; Clayton, F. C.; Bohn, D.; Leonard, Y. M.; Detwiler, T. J.; Lynch, J. J.; Yan, Y.; Chen, Z.; Kuo, L.; Gardell, S. J.; Shafer, J. A.; Vacca, J. P. Metabolism-directed optimization of 3-aminopyrazinone acetamide thrombin inhibitors. Development of an orally bioavailable series containing P1 and P3 pyridines. *J. Med. Chem.* **2003**, *46* (4), 461−473. (d) Young, M. B.; Barrow, J. C.; Glass, K. L.; Lundell, G. F.; Newton, C. L.; Pellicore, J. M.; Rittle, K. E.; Selnick, H. G.; Stauffer, K. J.; Vacca, J. P.; Williams, P. D.; Bohn, D.; Clayton, F. C.; Cook, J. J.; Krueger, J. A.; Kuo, L. C.; Lewis, S. D.; Lucas, B. J.; McMasters, D. R.; Miller-Stein, C.; Pietrak, B. L.; Wallace, A. A.; White, R. B.; Wong, B.; Yan, Y.; Nantermet, P. G. Discovery and evaluation of potent P1 aryl heterocycle-based thrombin inhibitors. *J. Med. Chem.* **2004**, *47* (12), 2995−3008. (e) Rittle, K. E.; Barrow, J. C.; Cutrona, K. J.; Glass, K. L.; Krueger, J. A.; Kuo, L. C.; Lewis, S. D.; Lucas, B. J.; McMasters, D. R.; Morrissette, M. M.; Nantermet, P. G.; Newton, C. L.; Sanders, W. M.; Yan, Y.; Vacca, J. P.; Selnick, H. G. Unexpected enhancement of thrombin inhibitor potency with o-aminoalkylbenzylamides in the P1 position. *Bioorg. Med. Chem. Lett.* **2003**, *13* (20), 3477−3482. (f) Nantermet, P. G.; Barrow, J. C.; Newton, C. L.; Pellicore, J. M.; Young, M.; Lewis, S. D.; Lucas, B. J.; Krueger, J. A.; McMasters, D. R.; Yan, Y.; Kuo, L. C.; Vacca, J. P.; Selnick, H. G. Design and synthesis of potent and selective macrocyclic thrombin inhibitors. *Bioorg. Med. Chem. Lett.* **2003**, *13* (16), 2781−2784.