

## Kinase Inhibitor Data Modeling and de Novo Inhibitor Design with Fragment Approaches

Michal Vieth,\* Jon Erickson, Jibo Wang, Yue Webster, Mary Mader, Richard Higgs, and Ian Watson

*Eli Lilly and Company Lilly Research Laboratories, Lilly Corporate Center, DC 1931, Indianapolis, Indiana 46285*

Received July 13, 2009

A reconstructive approach based on computational fragmentation of existing inhibitors and validated kinase potency models to recombine and create “de novo” kinase inhibitor small molecule libraries is described. The screening results from model selected molecules from the corporate database and seven computationally derived small molecule libraries were used to evaluate this approach. Specifically, 1895 model selected database molecules were screened at 20  $\mu$ M in six kinase assays and yielded an overall hit rate of 84%. These models were then used in the de novo design of seven chemical libraries consisting of 20–50 compounds each. Then 179 compounds from synthesized libraries were tested against these six kinases with an overall hit rate of 92%. Comparing predicted and observed selectivity profiles serves to highlight the strengths and limitations of the methodology, while analysis of functional group contributions from the libraries suggest general principles governing binding of ATP competitive compounds.

### Introduction

The kinase families' function in signal transduction pathways has given rise to multiple attractive targets for therapeutic intervention in many disease states such as cancer, diabetes, inflammation, and arthritis.<sup>1,2</sup> Recently, multiple drug launches and a variety of clinical studies with kinase inhibition as the primary mechanism of action has probed therapeutic space with inhibitors of this gene family. The success of drugs with kinase inhibition as the mode of action demonstrates the ability to deliver kinase inhibitor drugs with varying selectivity, potency, and pharmacokinetic properties.

Kinase-targeted compounds carry various degrees of inter-family selectivity as some marketed drugs inhibit several kinases (e.g., sunitinib), and others are more selective (e.g., gefitinib).<sup>3</sup> Thus, the understanding and practical application of design principles which lead to kinase inhibitors with appropriate selectivity are of high interest. Many clinical and marketed ATP competitive kinase inhibitors exhibit a similar pharmacophore.<sup>4</sup> This pharmacophore, exemplified by gefitinib among others (Figure 1a,b), contains a hinge binding element that is connected to both hydrophobic and solubilizing groups. A method that takes advantage of new technology for the efficient design of kinase inhibitors carrying this generalized pharmacophore has been developed. Here we describe an application of this method to design, synthesize, and assay kinase inhibitors to test the ability of our pharmacophore hypothesis to probe general kinase inhibition.

In order to understand kinase inhibitor potency and selectivity, we have developed informatics and computational approaches to characterize kinase target space from profiling data.<sup>5,6</sup> We have demonstrated that correlations exist between fragments or “privileged structures” within kinase inhibitors

and their binding affinity to particular kinases.<sup>7</sup> In addition, it was shown that similarity between kinase targets could be derived from the similarity of the fragments making up inhibitors of those particular kinases in a complementary fashion to relating targets by whole ligand similarity,<sup>8</sup> three-dimensional protein–ligand structure,<sup>9</sup> and full profiling approaches.<sup>10</sup> Also described was the use of small molecule inhibitor fragment frequencies for understanding kinase potency and selectivity.<sup>7</sup> In those efforts, Naïve Bayes models employing fragment frequencies provide highly interpretable and reliable means for predicting potency in individual kinases. Specifically, the models were utilized with moderate success (40% hit rate) in prospective screening for new kinase inhibitors.<sup>7</sup>

In this report, these concepts are taken a step further, beyond retrospective analysis to prospective generation of kinase inhibitor libraries. We introduce and prospectively validate improved computational models utilizing a support vector machine based method built with fingerprint descriptors (SVMFP<sup>a</sup>),<sup>11</sup> in the process of computer model driven selection of compounds for biological testing (also referred to as virtual screening<sup>12,13</sup>). The models are used as the driving force to select among new combinations of hinge-binding, hydrophobic and solubilizing fragments to design kinase inhibitors. Building blocks for these experiments were generated from fragmentation of existing inhibitors.<sup>7,14</sup> The resulting fragments were then recombined to create “de novo”

<sup>a</sup> Abbreviations: FBDD, fragment-based drug design; SBDD, structure-based drug design; QSAR, quantitative structure–activity relationship; SVM, support vector machine; SVMFP, support vector machine fingerprints; PPV, positive predicted value; PDB, Protein Databank; ATP, adenosine triphosphate; ABL1, c-abl oncogene 1, receptor tyrosine kinase; CHK2, CHK2 checkpoint kinase; FLT3, fms-related tyrosine kinase 3; MET, met proto-oncogene (hepatocyte growth factor receptor); P70S6K, RPS6KB1 ribosomal protein S6 kinase; ROCK2, Rho-associated, coiled-coil containing protein kinase 2; SP, single point; rms, root-mean-square; NN, nearest neighbor.

\*To whom correspondence should be addressed. Phone: (317) 277-3959. Fax: (317) 276-6545. E-mail: m.vieth@lilly.com.

chemical libraries and scored using qualified models trained on kinase potency. Analysis of the results from this experiment and data for a variety of universal hydrophobic and solubilizing groups for ABL1, CHK2, FLT3, MET, P70S6K, and ROCK2 kinases are presented.

## Methods

### Overview

- 1 SVMFP models used to drive library selection were validated in chronological testing (Supporting Information Figure 1) and prospective screening of 1895 compounds against six kinases (ABL1, CHK2, FLT3, MET, P70S6K, and ROCK2).
- 2 A general kinase pharmacophore, exemplified in the EGFR marketed drug gefitinib (Figure 1a), was used as a framework for libraries. SVMFP models were used to predict combinations of hypothetical hinge elements, hydrophobic and solubilizing groups from fragmented kinase inhibitors to design seven "de novo" kinase libraries with predicted activity in six kinases (ABL1, CHK2, FLT3, MET, P70S6K, and ROCK2) as outlined in Figure 1c,d
- 3 Fourteen candidate libraries with all compounds predicted active (predicted > 50% inhibition at 20  $\mu$ M, 60% of compounds had predicted activity > 70%) in at least one of six SVMFP kinase inhibition models, were initially selected and subsequently filtered to seven based on ease of synthesis. Compounds were then synthesized and assayed for single point (SP) % inhibition at 20  $\mu$ M in these six kinases. The results showed a 92% confirmation rate of model predictions.
- 4 Model performance was analyzed for each library focusing on group contributions of hydrophobic and solubilizing moieties for utility of different kinase inhibition.

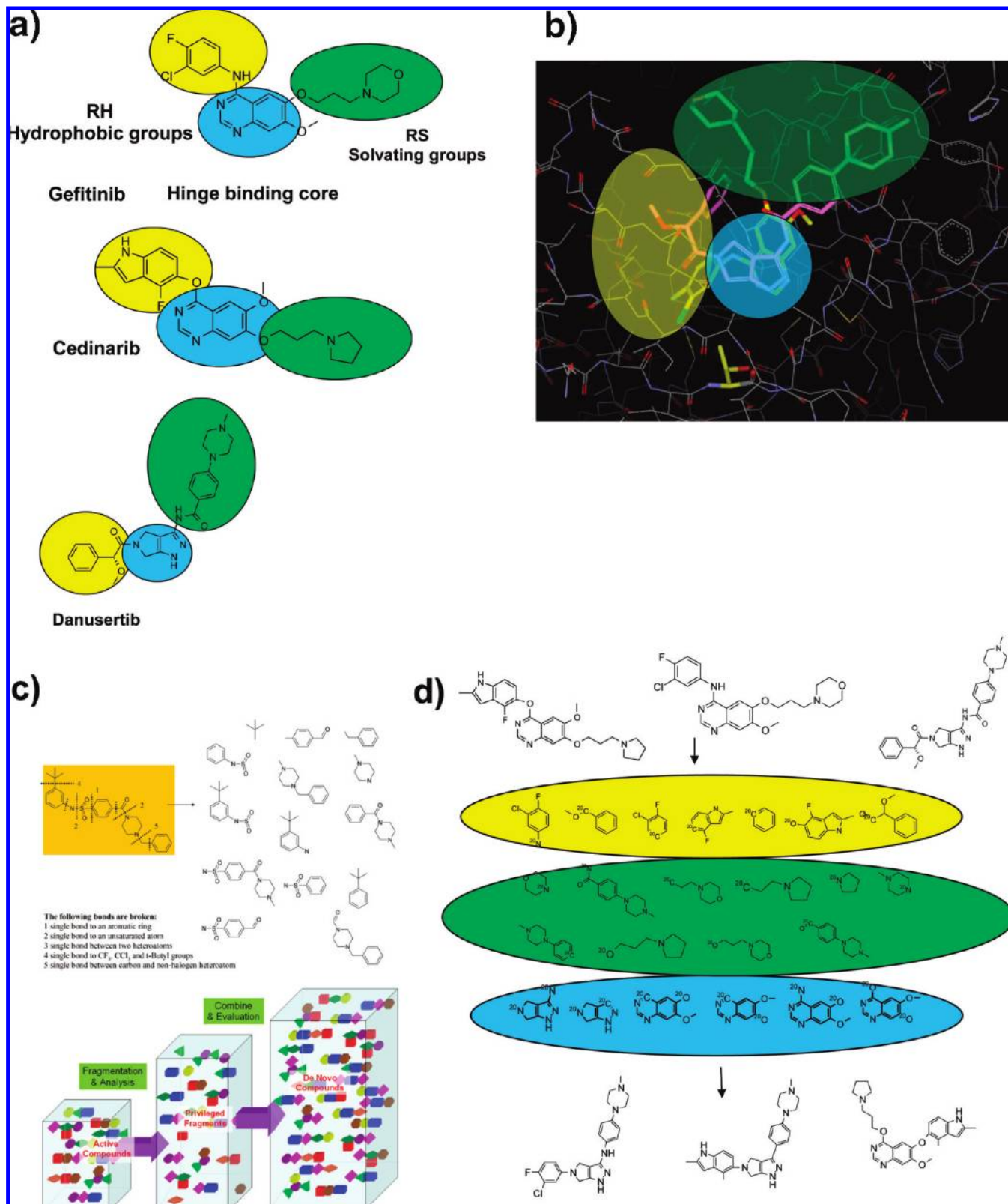
**SVMFP Model Development and Validation.** Previously described models used for kinase inhibition prediction<sup>7</sup> utilized fragment frequency based naïve Bayes models. In this work, support vector machine (SVM)<sup>15</sup> models using Merck atom pair<sup>16</sup> derived fingerprints were employed. These models provided superior results with respect to prediction accuracy as compared to the naïve Bayes models. Support vector machines (SVMs) are supervised learning methods used for classification and regression.<sup>17,18</sup> They function by simultaneously minimizing the empirical classification error and maximizing the geometric margin. Others have also reported on the utility of SVM models in QSAR-based virtual screening.<sup>11,17</sup> As with most models, the quality and relevance of model predictions and metrics of model performance depend on the number and structural diversity of compounds in the training set as well as the structural diversity of compounds to be predicted, i.e., things close to the training set generally get more reliable predictions than those further away but specific results might vary. Best practices indicate that one must split data into a test and training set to get reliable unbiased performance estimates.<sup>19,20</sup> Validation studies of the SVMFP models used here are based on multiple test/train splits and statistical measures, including positive predicted value (PPV) and active enrichments representing the average of the results as reported in earlier work.<sup>7</sup> To further validate the models, chronological test/train splits were used to validate these activity models using earlier collected training data to predict test data

collected at a later date. This was done to provide a more realistic estimate of model performance in a prospective study with existing compounds than the standard random selection of train/test splits, which can be overly optimistic given the sequential structural explorations inherent to an ongoing SAR studies. The less biased error estimates from chronological validation originate from the observation that new chemical diversity tends to be tested over time, better revealing a model's "extrapolation" prediction error on new chemical diversity. This chronological model validation scheme provides additional assurance that the models are suitable for use in the design of new compounds. In this work a variety of parameters were considered to qualify the models including, number of observations in the training and test sets, the number of actives, PPV, negative predictive value (NPV), sensitivity, and specificity.

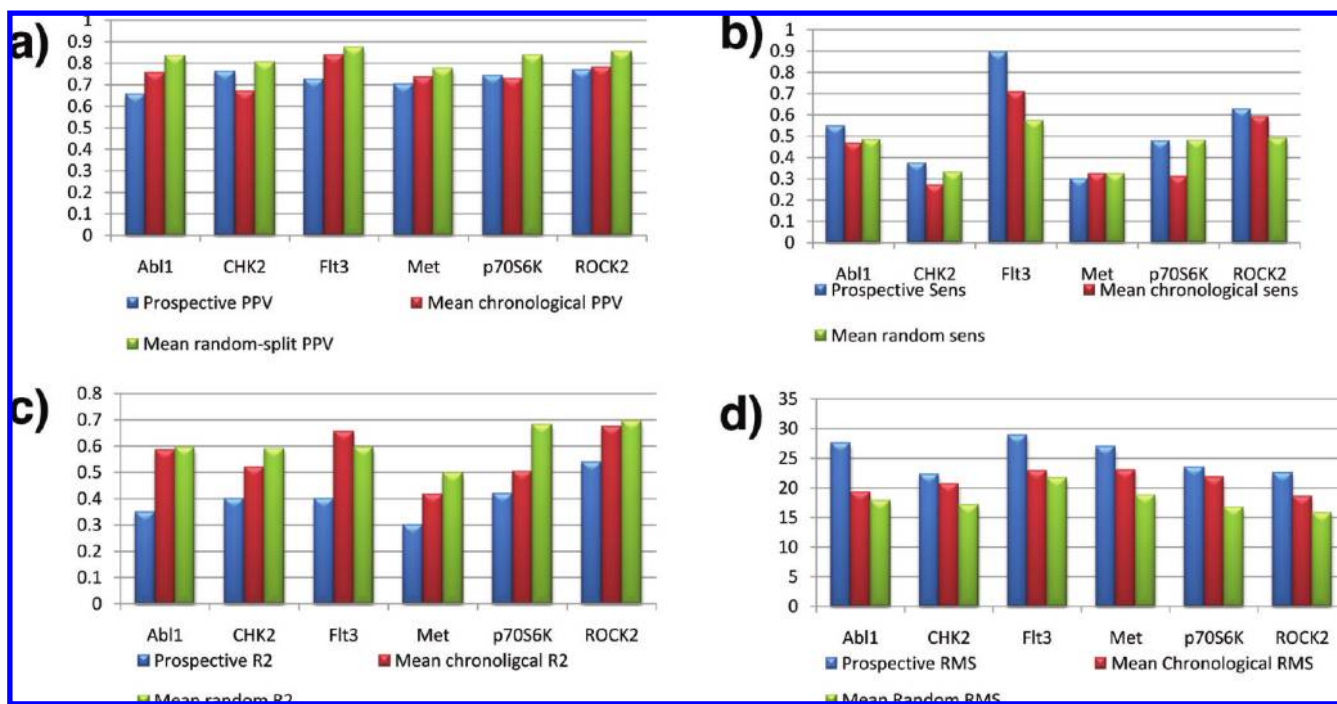
A set of six previously described kinase assays was assembled from our internal research efforts for this study;<sup>7</sup> we substituted MKK6 with MET. To increase statistical significance, at least 1400 actives ( $\geq 70\%$  inhibition at 20  $\mu$ M test substance concentration) and 16000 compounds assayed (17K average) for each of the six kinase assays was required. Details of the data used in modeling are given in Supporting Information Table 1. Data for each target was sorted in a chronological fashion, and the first 80% of compounds was used to build models, while the next 10% of compounds was used to test the model performance. The models were also built with the first 90% of the data, for training, and the last 10% to test model performance. This allowed for two chronological evaluations of model performance. In addition, three standard random 50–50 splits were performed using the first 80% of the data and the first 90% of the data, providing an additional six random test sets. The final model for prospective testing was then built using all data for each target. A graphical representation of the validation protocol and performance metrics is presented in Figure 1 of Supporting Information.

The final models were used to select a set of 1895 previously untested compounds from our compound collection. Any compound predicted to be more than 70% active in any of six models was considered for screening. As much as possible we tried to vary predictive selectivity of compounds so that we could test performance of each model in this experiment.

**De Novo Design with Kinase Pharmacophore.** To demonstrate proof of concept for the utility of the SVMFP models in driving the design of kinase inhibitors, the pharmacophore shared by many ATP competitive kinase clinical and marketed compounds, exemplified in Figure 1, were used. The elements of the pharmacophore were supplied by fragmentation of existing kinase actives (defined as < 1  $\mu$ M IC<sub>50</sub> in any kinase) with our previously described dicer approach,<sup>7</sup> followed by binding mode predictions of hypothetical hinge binding elements.<sup>21,22</sup> Inputs for the virtual library enumeration arose from 300 hinge binders, hydrophobic (less than or equal to two heteroatoms) and solubilizing group (presence of primary or secondary amine) coming from diced fragments to more than 300 hinge binders. SVMFP models were used to narrow down selection of hinge binders. A positional scanning<sup>23</sup>like strategy was used to avoid impractical enumeration of all possible structures. In this strategy, substituents for each candidate group (i.e., hydrophobic and solubilizing groups) at each of two positions on the hinge-binding group were scored relative to a random selection of



**Figure 1.** (a) Kinase pharmacophore exemplified in marketed and clinical compounds. Solubilizing groups are highlighted with green, hydrophobic with yellow, and hinge binders with cyan. (b) Three-dimensional picture of pharmacophore drawn over X-ray structures of gefitinib (yellow carbons, 2ity.pdb) in EGFR and danusertib in Aurora A (magenta carbons, 2j50.pdb). EGFR kinase from 2ity.pdb is shown. Transparent ovals in yellow, cyan, and green are drawn over hydrophobic, hinge binding, and solubilizing groups, respectively. Note that the positions of both solubilizing and hydrophobic groups occupy slightly different region in both structures in contrast to the hinge binding elements, which occupy similar space. Thr790 gate keeper in EGFR is shown in yellow stick representation. (c) Schematic representation of de novo design process starting from actives, fragmentation, and reassembly. Reproduced with ACS permission from ref 7. (d) Hypothetical application to three kinase drugs from Figure 1. Solubilizing groups are highlighted in green, hydrophobic in yellow, and hinge binders in cyan. Fragment connection atoms contain isotopic labels.



**Figure 2.** Comparison of Chronological, random validation with prospective metrics. (a) PPVs. (b) Sensitivities. (c) Root mean square prediction error. (d)  $R^2$  values.

five substituents for the other groups. Independence of the two positions/groups was assumed for this strategy. This assumption is likely to be working better for groups, which are spatially separated and not as well for groups which are in close proximity. It was hoped that the use of 5 substituents at the other position instead of just one would allow for some correction to this assumption. The enumeration resulted in 7500 compounds ( $5 \times 5 \times 300$ ), which were then selected by ranking the top 50 groups in each position by the highest predicted activities for each target. A final  $50 \times 50 \times 50$  matrix was then enumerated for each target, and the final best scoring 20 hinge binders and their best scoring compounds were considered for synthetic feasibility. Once the chemistry schemes were prepared, the libraries were re-enumerated using reasonably priced reagents and scored with the model for each target to select 20–60 compounds predicted to have > 50% activity in any of the six models. Initially, we were hoping to use 70% threshold, but for many libraries that proved rather difficult as models were too conservative in venturing into new chemical space. This difficulty was particularly evident for library 6, used as a positive control in this experiment. The compounds were described in the literature as MET inhibitors,<sup>24</sup> however, only 3 of 41 were predicted to have more than 70% inhibition in one of the kinases. A total of 265 compounds were targeted for the seven libraries ranging from 19 to 60 compounds each. The final 179 synthesized compounds were evaluated in SP assays in six kinases at 20  $\mu\text{M}$ , with SP hits defined as > 70% inhibition.

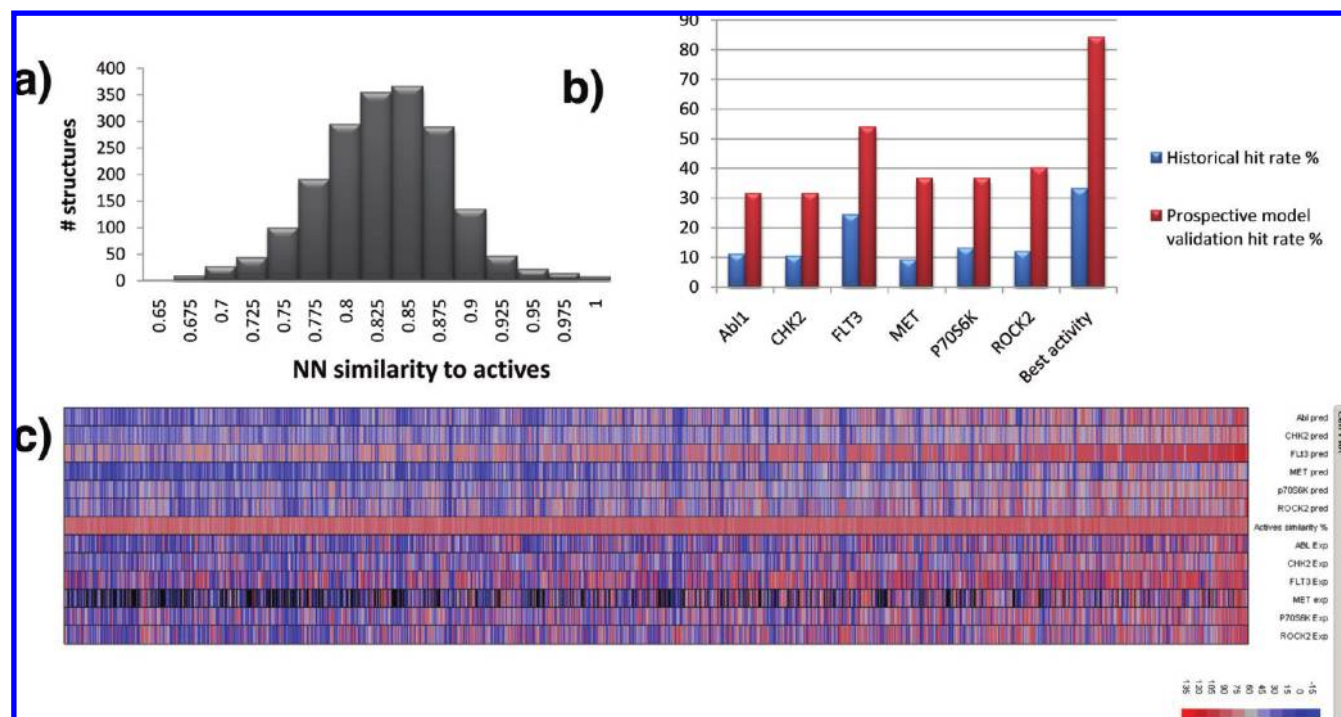
## Results and Discussion

It should be noted that the hypotheses discussed here are based on interpretation of actives as defined by single point (SP) percent inhibition data measured at a 20  $\mu\text{M}$  concentration. Although this measure is a good indication of a binding event, any selectivity or potency comparison is limited relative

to analysis of dose response data. The percent inhibition data was used for model comparison and analysis of group activities for our libraries because this response was consistent with values used to train the models.  $\text{IC}_{50}$  data for one library is presented to understand the correlation to dose response data. Overall, the translation is good, but the final dose response data tends to vary in magnitude for compounds showing similar, high percent inhibition.

**SVMFP Model Performance and Validation.** An element critical to these retrospective studies was a sufficient number of training set compounds and actives to properly estimate the model performance in blind prospective validation tests. In addition, model qualification for “de novo” library design employed a stringent chronological validation studies to give a more realistic estimate of model performance metrics. Parts a–d of Figure 2 show a comparison of performance metrics means derived from the prospective study along with two chronological validation studies and six traditional, randomized validation studies. The positive predictive value (PPV), which is the most critical performance metric for utilization of models in “de novo” design studies, is above 0.6 for all studied targets in the prospective study results and appears to be reasonably estimated by chronological and random split validation studies. For MET, P70S6K, and ROCK2, prospective study PPV appears to be very well estimated by chronological validation, while random split validation appears to provide a slight overestimate. For the ABL1, FLT3, the prospective study PPV is overestimated by both chronological and random split validation with the former, giving more realistic estimates. For CHK2, chronological validation gives a lower estimate, with random split validation giving a slight overestimate of prospective PPV.

Prospective sensitivity is important in understanding the false negative ratio and high sensitivity values are essential for prediction of selectivity. Figure 2b shows that prediction



**Figure 3.** Results from prospective testing. Distribution of nearest neighbor similarity of 1895 compounds with respect to 5870 historical actives. The fingerprints used for this study are a combination of molecular properties (atom count, size of largest ring, number of rings, number of ring atoms, number of aromatic atoms, number of fused ring atoms, number of heteroatoms), MACCS keys,<sup>25</sup> and extended connectivity fingerprints with size 3 shell. Tanimoto similarity in this fingerprint space between kinase drugs from Figure 1 was the following: gefinitib—cedinarib 0.728, danusertib—cedinarib 0.592, danusertib—gefinitib 0.585. (a) Distribution of nearest neighbor similarity. (b) Comparison of historical hit rates with prospective validation. (c) Bioprint of predicted (top) and observed (bottom) activities at 20 $\mu$ M for 1895 compounds in 6 kinases. MET had 1392 measurements.

sensitivity varies from 0.3, for the historically lowest hit rate target MET, to 0.9 for FLT3. In almost all cases, prospective sensitivity is underestimated by both random and chronological validations, with random validation estimates giving significantly better estimates for P70S6K and CHK2 and chronological validation being closer for ROCK2 and FLT3.

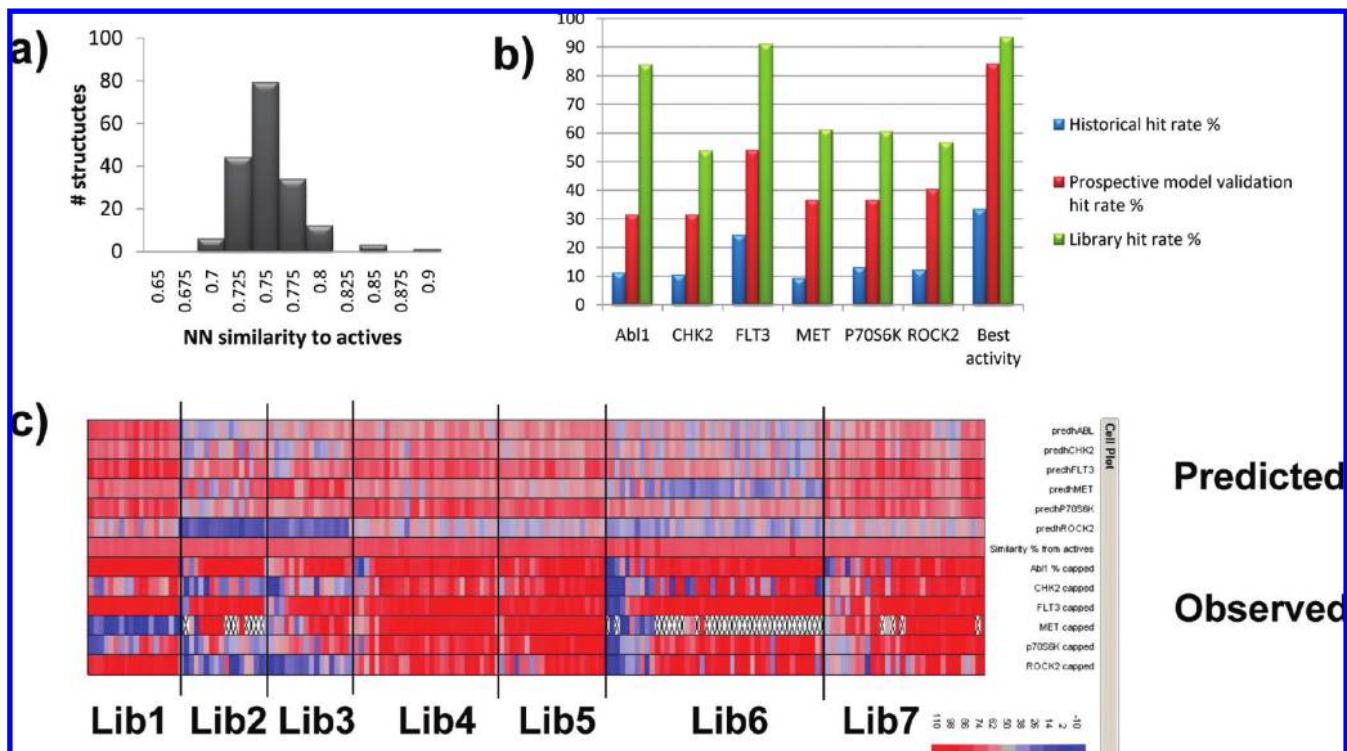
The rms error (see Figure 2c) is another metric important for selectivity predictions. The prospective study rms results are significantly higher in chronological and random split validations for two targets (ABL1, FLT3) and only slightly higher than chronological validation error for CHK2 and P70S6K. For the two other targets, medium range differences from chronological results (MET, ROCK2) are observed. In all cases, random split validation is underestimating the prospective errors more than the chronological method.

$R^2$  values, i.e., the variance in real data explained by the models, may also be important for ranking the compounds and determining model performance (Figure 2d). For all targets,  $R^2$  prospective validation values are significantly overestimated by both chronological and random split validations, with the chronological validation method giving lower overestimates for all targets except FLT3—note that all overestimates are large, especially for ABL1 and FLT3.

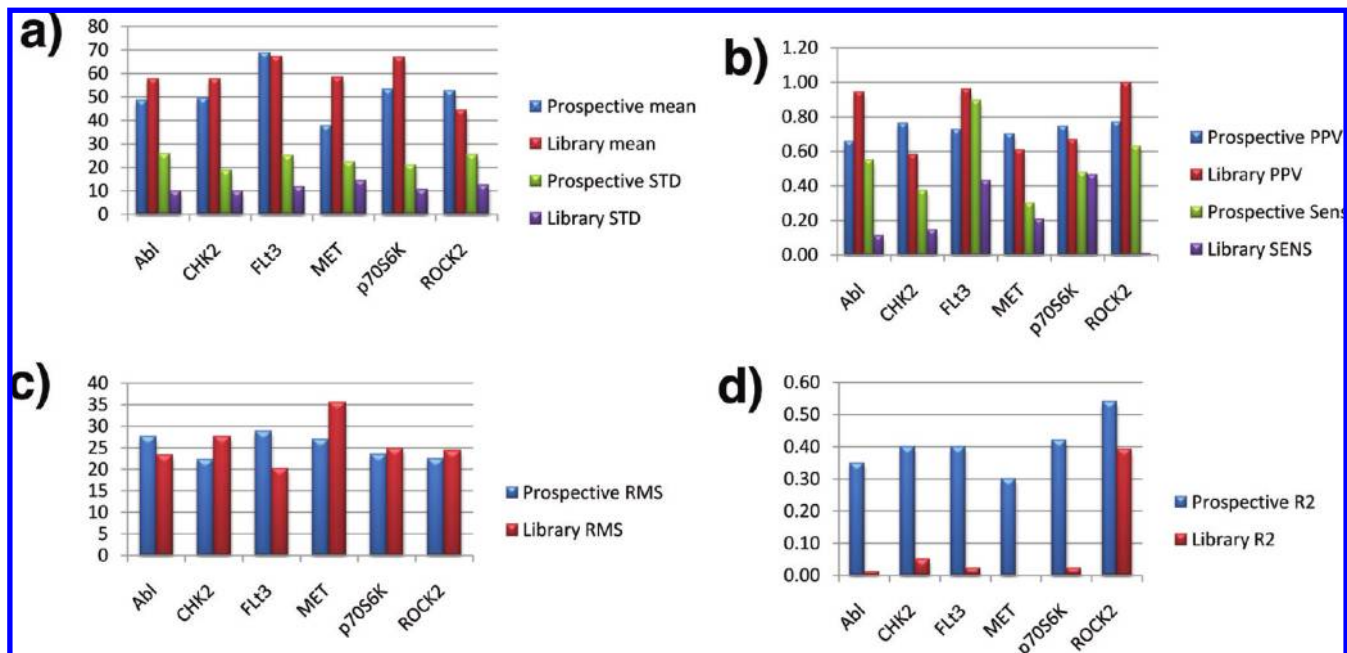
Figure 3 (and Table 1 in Supporting Information) shows the general statistics for SVMFP model prospective validation studies of 1895 compounds in six kinase assays. Per target hit rate, described as % of compounds > 70% at 20  $\mu$ M, are 2–4 fold higher than observed historically. Overall kinase hit rate for the prospective validation compounds increased 2.5 fold to 84%. The hit rate increased most (4 fold)

for MET, which historically was the lowest hit rate target at 9%, while the increase was the lowest at 2.2 fold (to 54%) for FLT3—historically the most promiscuous of the six kinase targets with a hit rate of 24%. Figure 3a shows the distribution of nearest neighbors (NN) using the fingerprint similarity (a Tanimoto similarity index using molecular properties, MACCS keys, and extended connectivity fingerprints<sup>7,25</sup>) to the actives used to derive the models of the compounds selected for the prospective experiment. The test set was designed to eliminate as many closely related compounds to the training set as possible, but a small number of closely related structures were also included (with similarity > 0.9, where a value of 1 is identical) to test the similarity dependence of active predictions. Very few compounds from our collection with similarity < 0.8 to known actives were predicted active in the models. The mean similarity of the selected set to existing actives was 0.82. We also examined the dependence of PPV as a function of similarity (NN similarity bins of > 0.9, < 0.8, and > 0.8) and unexpectedly found that the PPVs for all targets were mostly independent of similarity. We note, however, that the test set contained very few compounds that are very similar to training set by design. Figure 3c shows a bioprint visual comparison (i.e., a heat map depiction of actives/inactives<sup>7</sup>) of predicted activities with experimentally determined values for all prospective compounds. The comparison looks favorable for all targets and qualitatively confirms the values shown for the performance metrics.

**Library Design Results.** Armed with prospectively validated models for six targets (overall 84% hit rate, good PPV estimates, and sensitivities) high hit rates, although probably lower than obtained for the validation, was anticipated for



**Figure 4.** Results from library testing. (a) Distribution of nearest neighbor similarity of 179 compounds with respect to 6422 historical actives. (b) Comparison of historical hit rates with prospective validation and library hit rate. (c) Bioprint of predicted (top) and observed (bottom) activities at 20  $\mu$ M for 179 compounds in six kinases. MET had 116 measurements..



**Figure 5.** Comparison of model performance metrics for prospective validation and library compounds. (a) Mean model predictions and standard deviations for six targets. All differences in means are statistically significant except for FLT3. (b) PPVs and sensitivities. (c) Prediction rms. (d)  $R^2$  values: the low values for library compounds might be influenced by higher means for predicted values and lower standard deviations..

the “de novo” libraries. This was the first time the models were applied to previously unexplored chemical space where a majority of compounds are less than 0.8, similar to the training set actives of the six kinase targets. As shown in Figure 4a, the NN similarity distribution of library compounds averaged 0.74, much lower than the prospective validation set. On the other hand, all library compounds

shared a common pharmacophore and hinge binding elements that were present in actives, although with different connections and R groups. Results for all libraries significantly exceeded expectations as shown in Figure 4b. In fact the overall hit rate of 92% exceeded even that of the prospective test set. Hit and enrichment rates for each target were also higher, with values ranging from 50% (CHK2) to

90% (FLT3). The largest increase was observed for ABL1, with >80% hit rate compared to 10% for historical compounds and 30% for prospective test set. Not surprisingly, a qualitative correspondence shown in the bioprint format is low, as the activity of many compounds were significantly under predicted in cases where compounds showed activity across all targets (Figure 4c). While it is rewarding that the models led to active compounds, it appeared that the strategy and pharmacophore led to compounds with multiple kinase activities, i.e., promiscuous inhibitors. To learn from this experiment, the model performance metrics were contrasted for the library compounds and those from the prospective validation studies. The results are graphically displayed in Figure 5a–d (and in Table 2 in Supporting Information). First, the mean prediction values are significantly higher for library compounds in four targets, with FLT3 and ROCK2 minimally lower, than predictions for the validation set. Standard deviations of these values are significantly lower for the libraries than the validation set, suggesting that the predictions for library compounds were higher and within a narrower range, i.e., most library compounds were predicted to be more active across the board. Model PPV values were significantly higher for ABL1, FLT3, and ROCK2, comparable for MET and P70S6K and lower than those found in the prospective set for CHK2. Sensitivities were notably lower for library compounds, with the largest differences for ROCK2 (0.63 vs 0.01 for libraries) and ABL1 (0.55 vs 0.11). The only target with comparable sensitivity between the library and validation sets was P70S6K (0.48 vs 0.46). Low sensitivities point to significant underpredictions of activity for all targets, representing a significant challenge in model utilization for selectivity design. As a consequence of narrower, higher predictions, the  $R^2$  values are no better than random indicators for model quality for all targets except ROCK2 (target with marginal sensitivity).

Another way to look into the results for the prospective set and library compounds is to compute cross-target correlations for both experimental and predicted values. The results are presented in Table 1 for all target pairs. Correlations between pairs of targets for the experimental values for 1895 prospective compounds is low, except for P70S6K and CHK2 ( $R^2 = 0.35$ ,  $R = 0.6$ ). On the other hand, the 179 library compounds give higher correlations for P70S6K and MET ( $R^2 = 0.6$ ), FLT3 and ABL1 ( $R^2 = 0.5$ ), ROCK2 and MET ( $R^2 = 0.44$ ), ROCK2 and P70S6K ( $R^2 = 0.37$ ), and similar for P70S6K and CHK2 ( $R^2 = 0.36$ ). Thus, it appears that the pharmacophore used in the designed libraries increases the likelihood of getting similar potency for multiple pairs of targets. Interestingly, the models predict the highest correlations for FLT3 and CHK2 for library compounds—this is not observed in the experimental data. For library compounds, the models also predict high correlations for P70S6K against four other targets, with three of these confirmed by the experimental data. Most notably, the models underestimate the highest experimental correlations between P70S6K and MET. These results (Table 1) show the difficulty of accurately estimating target relationships (and thus selectivity) for new data sets either from previous relationships or predictive models. In summary of the model performance, we noted that utilization of model combination strategy (selecting molecules predicted active in any of the models) worked well in designing active compounds, while we still remain significantly challenged in utilization of models to design selective compounds. It is likely that a

**Table 1.** Target Cross-Correlation  $R^2$  Tables for Experimental and Predicted % Inhibition Values<sup>a</sup>

	ROCK2	P70S6K	MET	FLT3	CHK2	ABL1
	AGC (M)	AGC (L)	TK (M)	TK (F)	CAMK (L)	TK (T)
ROCK2		0.24 (0.37)	0.06 (0.44)	0.00* (0.30)	0.05 (0.00*)	0.00 (0.16)
P70S6K	0.26 (0.43)	-	0.12 (0.62)	0.03 (0.30)	0.35 (0.36)	0.04 (0.16)
MET	0.01 (0.00*)	0.05 (0.12)		0.11 (0.24)	0.19 (0.17)	0.16 (0.10)
FLT3	0.03 (0.22)	0.01 (0.55)	0.34 (0.06)	-	0.17 (0.07)	0.23 (0.50)
CHK2	0.03 (0.22)	0.20 (0.50)	0.17 (0.02)	0.27 (0.72)	-	0.15 (0.05)
ABL1	0.03 (0.22)	0.01 (0.41)	0.42 (0.21)	0.48 (0.43)	0.19 (0.40)	-



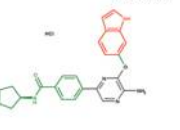
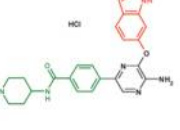
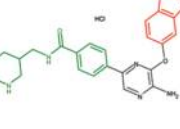


<sup>a</sup>  $R^2$  values are reported. Upper diagonal: experimental values, prospective  $N = 1895$ , except for MET  $N = 1392$ , library in parentheses  $N = 179$ . (MET  $N = 116$ ). Lower diagonal: predicted values,  $N = 179$ , except for MET  $N = 116$ , library compound correlations in parentheses. Kinase group and gate keeper residues (in parentheses) are included in table headings. Sequence identities between six kinases are available in Supporting Information. \* $P > 0.05$ .

different, more target specific pharmacophore, more similar chemotypes to existing actives, or a different set of targets could result in a more specific design, but this experiment highlights successes and challenges of good model driven de novo design.

**Selected Data for One Standard Library.** As a part of our proof of concept, one library (library 6) was used as positive control. For this library, we used model driven substituent selection to make up compounds previously described as MET kinase inhibitors.<sup>24</sup> Interestingly, most compounds from this library showed rather low predicted activity as only 3 of 44 synthesized compounds were predicted to be higher than 70% percent in any of six kinases (all compounds were predicted > 50%). This compares to 76% predicted actives in the other six libraries. As anticipated from the literature<sup>24</sup> all compounds from this library showed activity (> 70% at 20  $\mu$ M) in at least one kinase. A sample of data, including dose response, is shown in Table 2. To summarize, the data for these compounds highlights good translation of % inhibition to dose response for CHK2, underestimation of activity for this library, and little differentiation between % inhibition and the  $IC_{50}$  values (necessitating determination of  $IC_{50}$ s for compounds of specific interest). Compounds from this library exemplify the multikinase inhibition nature of other libraries designed in this study.

**Analysis of Hydrophobic and Solubilizing Groups.** To gain structural insights into the results of solubilizing and hydrophobic groups in our libraries, the data for compounds in which common solubilizing and hydrophobic groups were present were analyzed. Groups appearing in more than one library were identified to ensure the results were not dependent on a particular choice of a hinge binder. Parts a and b of Table 3 show mean % inhibition values for all compounds containing a particular group for the 12 solubilizing groups (Table 3a) and nine hydrophobic groups (Table 3b) examined. In addition, we examined the mean best target (the best activity for all targets) % inhibitions, mean number of atoms in compounds with that group, together with mean polar surface area (PSA)<sup>26</sup> and CLOGD<sup>27</sup> values. The groups are sorted based on their mean best target inhibition.

**Table 2.** Structures and Data for Example of Compounds from Library 6<sup>a</sup>

Structure	MW	CHK2	hMET	ABL1 %	MET %	FLT3 %	CHK2 %	P70 %	ROCK2 %
		FB IC50 uM	FB IC50 uM						
	537	1.13	1.21	11	53	61	60	45	42
predicted				57	54	42	51	54	45
	465	0.58	3.55	87	75	93	98	89	98
predicted				59	57	61	36	54	45
 AND Enantiomer	451	0.61	13.68	84	65	86	96	100	37
predicted				63	59	58	35	55	48
	465	0.80	5.52	85	73	102	98	98	99
predicted				53	57	58	38	49	42
	479	0.89	7.09	81	65	99	98	101	98
predicted				57	55	54	41	54	44
	465	0.96	4.27	87	77	100	98	86	97
predicted				52	52	62	33	46	39
	479	1.12	3.33	94	96	94	97	92	98
predicted				57	59	61	38	52	42

<sup>a</sup>Hydrophobic groups are drawn in red, solubilizing in green. Predicted values are shown in bottom row for each structure. Heat map color coding is used to denote the degree of predicted and observed inhibition. Dose response data for CHK2 and MET shown.

Some solubilizing groups, like the phenyl sulfonamide piperazine, are highly potent across all targets. Others like meta- and ortho-substituted benzyl piperidines are more active in Abl1, FLT3 (above in 90%) but have low activity in MET. Meta-substituted phenyl piperazines appear to give the lowest activity likely due to steric hindrance. The smallest solubilizing group used, *para*-benzyl-*N*-Me-amine, shows mean activity >80.5% in MET and FLT3, with lowest

activity in ROCK2. All solubilizing groups appear on average most potent in FLT3 as compared to the other five kinases.

In slight contrast to the solubilizing groups, the nine common hydrophobic groups show, on average, less promiscuity and more target differentiation. This may be understandable, as they target the gate keeper region which is often hypothesized to be responsible for selectivity.<sup>4,28</sup> However



**Table 3.** Structures and Experimentally Observed, Averaged % Inhibitions for Library Compounds: (a) Solubilizing Groups, (b) Hydrophobic Groups<sup>a</sup>

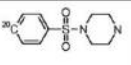
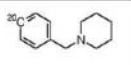
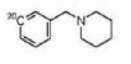
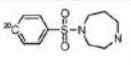
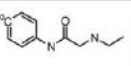
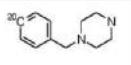
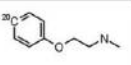
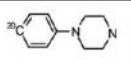
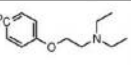
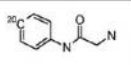
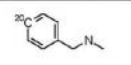
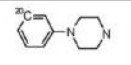









a) solubilizing groups											
R1-solubilizing group	#	Mean ABL1 %	Mean CHK2 %	Mean FLT3 %	Mean MET %	Mean P70 %	Mean ROCK 2 %	Mean BEST target %	H A	PSA	Mean ClogD PH=7.4
	11	92	92	93	97	94	92	100	34	121	2.1
	6	99	67	96	4	62	77	99	31	63	3.0
	4	98	61	97	25	65	77	99	31	61	3.4
	7	95	89	94	91	91	90	99	35	119	1.4
	8	90	86	93	97	90	77	98	31	106	2.9
	6	70	78	90	80	78	77	97	33	79	2.9
	4	87	83	96	96	78	74	97	30	85	2.6
	8	89	90	91	95	87	87	97	32	80	3.2
	6	82	60	90	67	64	63	93	34	71	3.5
	6	65	67	85	87	77	68	93	31	120	3.2
	9	75	67	85	81	68	55	90	30	76	2.2
	8	65	68	88	76	75	71	89	32	79	3.4
b) hydrophobic groups											
R2-hydrophobic groups	#	Mean ABL1 %	Mean CHK2 %	Mean FLT3 %	Mean MET %	Mean P70 %	Mean ROCK2 %	Mean BEST target %	HA	PSA	Mean ClogD PH=7.4
	16	97	41	95	75	69	68	100	33	126	0.4
	3	93	76	96	67	83	95	99	31	102	2.0
	6	93	74	94	61	74	87	97	29	89	2.3
	18	89	70	89	74	73	58	96	34	86	3.2

Table 3. Continued

b) hydrophobic groups (continued)											
R2 - hydrophobic groups	#	Mean ABL1 %	Mean CHK2 %	Mean FLT3 %	Mean MET %	Mean P70 %	Mean ROCK2 %	Mean BEST target %	HA	PSA	Mean ClogD PH=7.4
	10	74	81	89	92	81	83	95	31	88	2.3
	18	86	87	91	90	88	84	95	33	125	0.3
	19	86	45	87	57	55	50	94	35	89	2.8
	10	78	78	89	80	72	67	93	28	88	2.0
	7	70	78	86	64	82	79	90	34	103	2.6

<sup>a</sup> Each depicted fragment had to be present in minimum of two libraries. Mean values are given for all targets and properties. Red to blue heat-map color coding with distinct colors for >90%, >70%, >50%, and <50%. Mean values for atom counts, CLOGD, and PSA are shown to give connection to physical properties. Connection atoms contain isotopic labels.

gate keeper identity is not likely to be the only determinant of hydrophobic group differentiation. While similarity between activity of hydrophobic groups exemplified in Table 3b for two Leu gate keeper targets (P70 and CHK2) is high, there is also high correlation of averaged activity of these groups between Abl (Thr gate keeper) and Flt3 (Phe gate keeper) as well as between ROCK2 (Met gate keeper) and P70S6K (Leu gate keeper).

Hydrophobic groups in Table 3b are sorted with respect to their mean maximum % inhibition. 4-Substituted-indole carries high (>90%) mean activity in Abl1 and FLT3, with lower values in MET, while 5-substituted indole is more ubiquitous across the board, with >90% mean inhibition in FLT3 and MET. Phenyl is quite similar for all targets with <70% mean inhibition only in ROCK2, but benzyl gives the highest mean % inhibition in MET across all targets and groups. CHK2 appears to have preference for 5-substituted indole from all examined groups. Thus the CHK2 activity of compounds from library 6 exemplified in Table 2 likely comes from the presence of the 5-substituted indole. The presence of halogens in the multiple hydrophobic groups exemplified here may play a role in potential halogen bonding<sup>29</sup> phenomenon in addition to providing steric and hydrophobic effect.

Mean values for atom counts, CLOGD, and polar surface area (PSA) are shown in Table 3 to give connection to physical properties. Most structures are in favorable physical property space,<sup>5,30,31</sup> with computed CLOGD values less than 3, a reasonable size with less than 35 heavy atoms and less than 120 PSA values.

### Conclusions

In summary, we have shown how model driven kinase inhibitor design can be used to probe kinase space with new inhibitors. It was demonstrated that highly validated models can be used in virtual screening and to design active, multi-

kinase inhibitors. Our prospective results show more balanced predictability for all targets and better potential to probe selectivity for compounds closely related to training set compounds. The library results, which make use of a generic framework, showed that while this pharmacophore approach is good at uncovering kinase actives, it is challenged in terms of designing selective compounds. For instance, the inability to achieve good model sensitivity suggests that more work is needed to utilize this methodology to create selective tools for kinase target validation. Nevertheless, the group analysis described here sheds some light into the future design of more targeted inhibitors. In general, the ability to validate and use predictive modeling in blind/prospective experiments and de novo design of new compounds from existing information was demonstrated. To our knowledge, the overall virtual screening results show unprecedented success as compared to previous prospective literature reports.<sup>12</sup> Our de novo driven library results are even more encouraging when compared to reported computer driven designs, which utilized minor modifications of existing inhibitors with 30–40% success rate.<sup>32</sup> The implications of this research span fragment based drug discovery,<sup>33</sup> selectivity considerations, and predictive modeling applications/expectations. Understanding the proper utilization of predictive models demonstrated by our work should raise the standards of new compound design not only for protein kinases but for other gene families for which sufficient high quality data is available.

**Acknowledgment.** We thank Drs. Sutherland, Robertson, Hemmerle, Toth, and Bloem, and the Lilly Library Sciences Group for supporting this research.

**Supporting Information Available:** General statistics for models, historical hit rates, and prospective validation result; model performance metrics for prospective validation (1895 compounds) and libraries (179 compounds); target correlation

table with sequence identity included; graphical demonstration of chronological validation and PPV concepts; biological assay description. Biological profiling data and compound data in fingerprint format are available from authors upon request. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## References

- Grant, S. K. Therapeutic Protein Kinase Inhibitors. *Cell. Mol. Life Sci.* **2009**, *66*, 1163–1177.
- Fabbro, D.; Garcia-Echeverria, C. G. Targeting protein kinases in cancer therapy. *Curr. Opin. Drug Discovery Dev.* **2002**, *5*, 701–712.
- Fabian, M. A.; Biggs, W. H., III; Treiber, D. K.; Atteridge, C. E.; Azimioara, M. D.; Benedetti, M. G.; Carter, T. A.; Ciceri, P.; Edeen, P. T.; Floyd, M.; Ford, J. M.; Galvin, M.; Gerlach, J. L.; Grotzfeld, R. M.; Herrgard, S.; Insko, D. E.; Insko, M. A.; Lai, A. G.; Lelias, J. M.; Mehta, S. A.; Milanov, Z. V.; Velasco, A. M.; Wodicka, L. M.; Patel, H. K.; Zarrinkar, P. P.; Lockhart, D. J. A small molecule–kinase interaction map for clinical kinase inhibitors. *Nat. Biotechnol.* **2005**, *23*, 329–336.
- Gill, A. L.; Verdonk, M.; Boyle, R. G.; Taylor, R. A Comparison of Physicochemical Property Profiles of Marketed Oral Drugs and Orally Bioavailable Anti-Cancer Protein Kinase Inhibitors in Clinical Development. *Curr. Top. Med. Chem.* **2007**, *7*, 1408–1422.
- Vieth, M.; Siegel, M. G.; Higgs, R. E.; Watson, I. A.; Robertson, D. H.; Savin, K. A.; Durst, G. L.; Hipskind, P. A. Characteristic physical properties and structural fragments of marketed oral drugs. *J. Med. Chem.* **2004**, *47*, 224–232.
- Vieth, M.; Sutherland, J. J.; Robertson, D. H.; Campbell, R. M. Kinomics: characterizing the therapeutically validated kinase space. *Drug Discovery Today* **2005**, *10*, 839–46.
- Sutherland, J. J.; Higgs, R. E.; Watson, I.; Vieth, M. Chemical fragments as foundations for understanding target space and activity prediction. *J. Med. Chem.* **2008**, *51*, 2689–2700.
- Keiser, M. J.; Roth, B. L.; Armbruster, B. N.; Ernsberger, P.; Irwin, J. J.; Shoichet, B. K. Relating protein pharmacology by ligand chemistry. *Nat. Biotechnol.* **2007**, *25*, 197–206.
- Fernandez, A.; Maddipati, S. A priori inference of cross reactivity for drug-targeted kinases. *J. Med. Chem.* **2006**, *49*, 3092–3100.
- Paolini, G. V.; Shapland, R. H.; van Hoorn, W. P.; Mason, J. S.; Hopkins, A. L. Global mapping of pharmacological space. *Nat. Biotechnol.* **2006**, *24*, 805–815.
- Liew, C. Y. M.; Xiao, H.; Liu, X.; Yap, C. W. SVM Model for Virtual Screening of Lck Inhibitors. *J. Chem. Inf. Model.* **2009**, *49*, 877–885.
- Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. Docking and Scoring in Virtual Screening for Drug Discovery: Methods and Applications. *Nat. Rev. Drug Discovery* **2004**, *3*, 935–949.
- Shoichet, B. K. Virtual screening of chemical libraries. *Nature* **2004**, *432*, 862–865.
- Vieth, M.; Siegel, M. Structural fragments in marketed oral drugs. In *Fragment Based Drug Discovery*; Erlanson, J., Ed.; Wiley-VCH: Weinheim, 2006; pp 123–135.
- Joachims, T. Making Large-Scale SVM Learning Practical. In *Advances in Kernel Methods—Support Vector Learning*; Scholkopf, B.; Burges, C.; Smola, A., Eds.; MIT Press: Cambridge, 1999.
- Carhart, R. E.; Smith, D. H.; Venkataraghavan, R. Atom pairs as molecular features in structure–activity studies: definition and applications. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 64–73.
- Doucet, J.-P. B.; Xia, F.; Panaye, H.; Fan, A.; Nonlinear, B. SVM Approaches to QSPR/QSAR Studies and Drug Design. *Curr. Comput.-Aided Drug Des.* **2007**, *3*, 263–289.
- Han, L. Y.; Ma, X. H.; Lin, H. H.; Jia, J.; Zhu, F.; Xue, Y.; Li, Z. R.; Cao, Z. W.; Ji, Z. L.; Chen, Y. Z. A support vector machines approach for virtual screening of active compounds of single and multiple mechanisms from large libraries at an improved hit-rate and enrichment factor. *J. Mol. Graph. Model.* **2008**, *26*, 1276–1286.
- Gedeck, P.; Lewis, R. A. Exploiting QSAR models in lead optimization. *Curr. Opin. Drug Discovery Dev.* **2008**, *11*, 569–575.
- Tropsha, A. Variable selection QSAR modeling, model validation, and virtual screening. *Annu. Rep. Comput. Chem.* **2006**, *2*, 113–126.
- Sutherland, J. J.; Nandigam, R. K.; Erickson, J. A.; Vieth, M. Lessons in Molecular Recognition. 2. Assessing and Improving Cross-Docking Accuracy. *J. Chem. Inf. Model.* **2007**, *47*, 2293–2302.
- Nandigam, R. K.; Evans, D. A.; Erickson, J. A.; Kim, S.; Sutherland, J. J. Predicting the accuracy of ligand overlay methods with Random Forest models. *J. Chem. Inf. Model.* **2008**, *48*, 2386–2394.
- Dooley, C.; Houghten, R. The use of positional scanning synthetic peptide combinatorial libraries for the rapid determination of opioid receptor ligands. *Life Sci.* **1993**, *52*, 1509–1517.
- Cui, J.; Funk, L. A.; Jia, L.; Kung, P.-P.; Meng, J. J.; Nambu, M. D.; Pairish, M. A.; Sheng, H.; Tran-Dube, M. B. Aminoheteroaryl compounds as protein tyrosine kinase inhibitors. PCT Gazette, WO 2006/021886 A1, **2006**.
- Givehchi, A.; Bender, A.; Glen, R. C. Analysis of activity space by fragment fingerprints, 2D descriptors, and multitarget dependent transformation of 2D descriptors. *J. Chem. Inf. Model.* **2006**, *46*, 1078–1083.
- Ertl, P.; Rohde, B.; Selzer, P. Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties. *J. Med. Chem.* **2000**, *43*, 3714–3727.
- Chemaxon LogD*; ChemAxon Kft: Budapest, 1037 Hungary, 2008; [www.chemaxon.com](http://www.chemaxon.com).
- Vieth, M.; Higgs, R. E.; Robertson, D. H.; Shapiro, M.; Gragg, E. A.; Hemmerle, H. Kinomics—structural biology and chemogenomics of kinase inhibitors and targets. *Biochim. Biophys. Acta* **2004**, *1697*, 243–257.
- Voth, R. A.; Ho, S. P. The Role of Halogen Bonding in Inhibitor Recognition and Binding by Protein Kinases. *Curr. Top. Med. Chem.* **2007**, *7*, 1336–1348.
- Muegge, I. Selection criteria for drug-like compounds. *Med. Res. Rev.* **2003**, *23*, 302–321.
- Lipinski, C.; Hopkins, A. Navigating chemical space for biology and medicine. *Nature* **2004**, *432*, 855–861.
- Deanda, F.; Stewart, E. L.; Reno, M. J.; Drewry, D. H. Kinase-Targeted Library Design through the Application of the PharmPrint Methodology. *J. Chem. Inf. Model.* **2008**, *48*, 2395–2403.
- Pellecchia, M. Fragment-based drug discovery takes a virtual turn. *Nat. Chem. Biol.* **2009**, *5*, 274–275.