Kinase-Targeted Library Design through the Application of the PharmPrint Methodology

Felix Deanda,* Eugene L. Stewart, Michael J. Reno, and David H. Drewry

GlaxoSmithKline, Five Moore Drive, Research Triangle Park, North Carolina 27709

Received August 7, 2008

The PharmPrint methodology, as modified and implemented by Deanda and Stewart, was prospectively evaluated for use as a virtual high-throughput screening tool by applying it to the design of target-focused arrays. To this end, PharmPrint quantitative structure−activity relationship (QSAR) models for the prediction of AKT1, Aurora-A, and ROCK1 inhibition were constructed and used to virtually screen two large combinatorial libraries. Based on predicted activities, an Aurora-A targeted array and a ROCK1 targeted array were designed and synthesized. One control group per designed array was also synthesized to assess the enrichment levels achieved by the QSAR models. For the Aurora-A targeted array, the hit rate, against the intended target, was 42.9%, whereas that of the control group was 0%. Thus, the enrichment level achieved by the Aurora-A QSAR model was incalculable. For the ROCK1 targeted array, the hit rate against the intended target was 30.6%, whereas that of the control group was 5.10%, making the enrichment level achieved by the ROCK1 QSAR model 6-fold above control. Clearly, these results support the use of the PharmPrint methodology as a virtual screening tool for the design of kinase-targeted arrays.

INTRODUCTION

With continuing advances in combinatorial chemistry and an ever-growing list of commercially available reagents, medicinal chemists have, at their disposal, the means to synthesize vast numbers of compounds. Complementarily, advances in high-throughput screening (HTS) make it possible to test tens to hundreds of thousands of compounds for biological activity against targets of pharmaceutical interest in a single day.[1−7] Early work associated with these two technologies emphasized the synthesis and screening of large, diverse libraries for lead exploration. The widely held belief was that if one synthesized as many diverse compounds as possible, a lead compound would be found in the library. Moreover, subsequent optimization of the "lead" into a drug candidate would be fairly straightforward, given the expectation that sufficient structure−activity relationship (SAR) data would be elucidated from the library itself.[1−7] However, practical experience soon revealed that lead identification from screening large, diverse collections of compounds was not so simple. In addition to lower-than-expected hit rates, the few hits that were identified had a tendency to have less-than-desirable physicochemical properties, making lead optimization difficult. Consequently, the focus recently has shifted from the synthesis of large, diverse libraries to smaller, target-focused arrays that incorporate as much available knowledge about the target of interest as possible, in an effort to make lead discovery more productive, efficient, and cost-effective.[1−7]

In designing target-focused arrays, medicinal chemists can apply one of several different computer-assisted drug design (CADD) methods, from ligand-based techniques to structure-based techniques. For numerous protein kinases, perhaps the most obvious and logical choice is structure-based library design, given the wealth of 3D structural data that has accumulated over the last two decades from NMR and X-ray crystallography.[2,3,8−12] With a protein kinase structure in hand, a receptor−ligand docking and scoring method engineered for use with combinatorial libraries can be applied to virtually screen a library and calculate predicted binding modes and affinities to rank-order compounds. Next, by utilizing the docking scores, a library design program can be used to design a kinase-targeted array, whereby one design objective would involve maximizing the predicted "activities" for the select subset of compounds to be synthesized.[3−8,12−17]

Although docking and scoring have intuitive appeal, the virtual screening of a large combinatorial library by such techniques remains a computationally expensive process, requiring substantial amounts of computing resource and CPU time.[3−5,7−10,15,17−20] Further complicating matters is that seldom do we want to virtually screen a single library against a single kinase target. In our experience, there are often multiple combinatorial libraries in hand that we desire to screen across multiple kinases, which is clearly a daunting task with structure-based library design methods. Thus, a less computationally intensive, yet reliable and efficient, approach is needed as a virtual HTS tool to reduce large combinatorial libraries to smaller, more-practical subsets for synthesis and subsequent screening against targets of interest.

In addition to 3D kinase structures, the pharmaceutical and biotechnology industries have accumulated large volumes of SAR data. Consequently, quantitative structure−activity relationship (QSAR) methods represent an attractive solution to our problem, given that many such applications require less computer resources and CPU time, compared to structure-based methods.[2,15,18,21−25] To this end, we evaluated the

* Author to whom correspondence should be addressed. Telephone: (919) 483-9482. Fax: (919) 483-6053. E-mail: felix.g.deanda@gsk.com.

10.1021/ci800276t CCC: $40.75
Published on Web 11/21/2008

Table 1. Summary of Statistics from the PharmPrint/SAS-PLS Analyses for the Kinase Panel

protein kinase	number of compounds	number of 3-point P'phores[a]	number of PLS factors	minimum root mean PRESS	non-CV r^2 value[b]	standard error of prediction[c]
AKT1	1816	7169	9	0.770	0.661	0.51
Aurora-A	3220	8143	14	0.616	0.810	0.54
ROCK1	2478	6935	13	0.687	0.742	0.56

[a] Number of three-point pharmacophores in the enriched PharmPrint fingerprint. [b] Noncross-validated r^2 value. [c] Standard error of prediction.

PharmPrint methodology[25−27] as a virtual HTS tool for use in lead discovery. McGregor and Muskal[26,27] have published details on PharmPrint, which is a three-point pharmacophore fingerprint, and its use as a set of molecular descriptors in generating 3D QSAR models by partial least-squares (PLS) regression. In their protein kinase-related work, Deanda and Stewart[25] modified this CADD tool by implementing several simple, yet significant, procedural and algorithmic changes to improve the predictive performance of their PharmPrint QSAR models. In this study, we evaluated this modified PharmPrint methodology as a virtual HTS tool by prospectively applying it to the design of kinase-targeted arrays.

PharmPrint QSAR models for the prediction of AKT1, Aurora-A, and ROCK1 inhibition were constructed and used to virtually screen two large combinatorial libraries. The predicted $p\mathrm{IC}_{50}$ values were next used to profile each library, to determine which kinase to target with which library. One library scored well against Aurora-A, whereas the other scored well against ROCK1. Therefore, an Aurora-A targeted array and a ROCK1 targeted array were designed, synthesized, and subsequently screened against the kinase panel. Two additional arrays were also synthesized—one from each combinatorial library—to serve as control groups. The latter arrays were synthesized from kinase-biased, hand-selected sets of reagents. In this report, the biological results for the designed arrays are summarized and the predictive abilities of the PharmPrint QSAR models discussed. The enrichment levels achieved by the QSAR models versus control groups in the design of kinase-targeted arrays are also assessed and discussed.

EXPERIMENTAL METHODS

Protein Kinase Datasets. Experimental $p\mathrm{IC}_{50}$ data for AKT1, Aurora-A, and ROCK1 inhibitors were collected from the corporate database for the purpose of constructing PharmPrint QSAR models. AKT1, Aurora-A, and ROCK1 were selected as representatives of the AGC subfamily of Ser/Thr protein kinases due primarily to the large volume of available biological and chemical data as well as interest in inhibitors of these enzymes as potential therapeutic agents. The AKT1 data set was comprised of 1816 inhibitors with $p\mathrm{IC}_{50}$ values ranging from 3.67 to 7.87. Altogether, 85 different substructural classes of compounds were represented in the set, as determined by an in-house chemical classification scheme. The Aurora-A data set included 3220 inhibitors, each representing one of 92 substructural classes. The range of $p\mathrm{IC}_{50}$ values for this set was 4.61—9.03. Lastly, the ROCK1 dataset was comprised of 2478 inhibitors, representing 95 substructural classes of compounds with $p\mathrm{IC}_{50}$ values ranging from 3.42 to 9.09.

The experimental error associated with each training set was estimated to be less than 0.50 log units. For all three datasets, the number of experimental $p\mathrm{IC}_{50}$ values associated with each compound was greater than 2. In addition, there were no modifiers (i.e., < or >) associated with any of these values. Lastly, the standard deviation for the $p\mathrm{IC}_{50}$ values of any given compound was less than 0.50 log units.

SAS/PLS-Derived, PharmPrint QSAR Models. The PharmPrint methodology, as modified and implemented by Deanda and Stewart,[25] was used to generate 3D QSAR models for the prediction of AKT1, Aurora-A, and ROCK1 inhibition. The PharmPrint/PLS algorithm provided by McGregor and Muskal was replaced with the SAS/PLS procedure from the SAS System.[28] The SAS/Singular Value Decomposition (SVD) algorithm was selected for latent factor extraction and split cross-validation was performed to identify the optimal number of PLS components to include in each regression model. The statistic used in identifying the optimal number of latent factors was the root mean PRESS (predicted residual sum of squares).[25]

In constructing the QSAR models, the PharmPrint fingerprints for all compounds in each of the three training sets were first computed. Next, the fingerprints associated with a given dataset were enriched by eliminating three-point pharmacophores present in or absent from all compounds in that set, as these descriptors could be safely eliminated without affecting the regression model.[25] There were a total of 10549 three-point pharmacophores encoded into the original PharmPrint fingerprint; Table 1 lists the numbers of three-point pharmacophores that remained in the fingerprints for each dataset after the enrichment was completed. For the set of AKT1 inhibitors, the enriched fingerprints included 7169 three-point pharmacophores. For the sets of Aurora-A and ROCK1 inhibitors, the enriched fingerprints were comprised of 8143 and 6935 three-point pharmacophores, respectively.

A SAS/PLS-derived QSAR model was constructed from each training set by fitting the experimental $p\mathrm{IC}_{50}$ values against the corresponding enriched PharmPrint fingerprints. Table 1 provides a brief summary of statistics from the PLS regression analyses. Note that the PharmPrint QSAR model for AKT1 inhibition was comprised of nine PLS components and had a minimum root mean PRESS of 0.77. The non-cross-validated r^2 value was 0.661, and the standard error of prediction was 0.51 log units. A plot of predicted versus experimental $p\mathrm{IC}_{50}$ values for the AKT1 training set is illustrated in Figure 1. Figure 2 shows a plot of predicted versus experimental $p\mathrm{IC}_{50}$ values for the Aurora-A training set. The PharmPrint QSAR model for Aurora-A inhibition included 14 PLS components with a minimum root mean PRESS of 0.616. The non-cross-validated r^2 value was 0.810, with a standard error of prediction of 0.54 log units. A plot of predicted versus experimental $p\mathrm{IC}_{50}$ values for the ROCK1 training set is illustrated in Figure 3. From Table 1, note

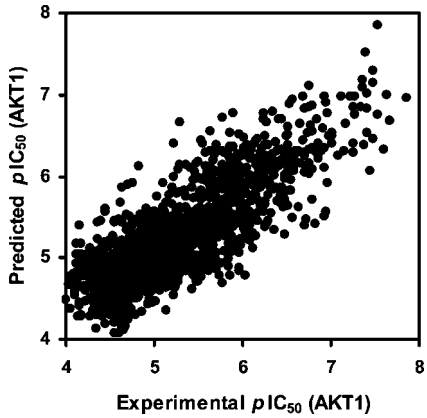


Figure 1. Predicted versus experimental pIC_{50} values for the training set of AKT1 inhibitors.

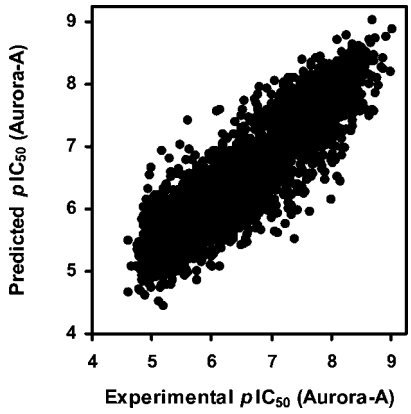


Figure 2. Predicted versus experimental pIC_{50} values for the training set of Aurora-A inhibitors.

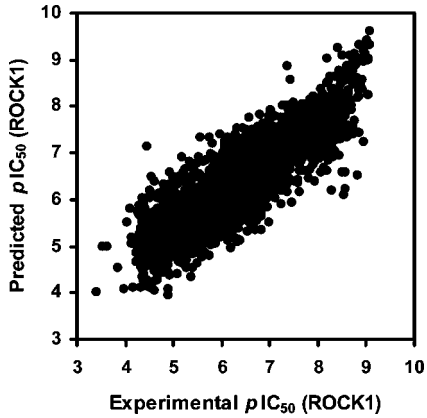
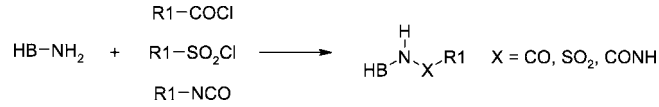


Figure 3. Predicted versus experimental pIC_{50} values for the training set of ROCK1 inhibitors.

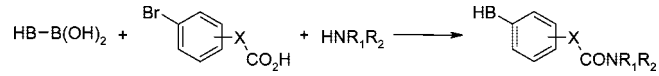
that the PharmPrint QSAR model for ROCK1 inhibition consisted of 13 PLS components and had a minimum root mean PRESS of 0.687. The non-cross-validated r^2 value was 0.742, with a standard error of prediction of 0.56 log units.

Virtual Combinatorial Library 1. Virtual Library 1 was a two-component combinatorial library (see Scheme 1) that was enumerated from a set of 51 kinase hinge binders (that is, reagents with scaffolds known to bind to the kinase hinge region) and a set of 966 acyl chlorides, sulfonyl chlorides, and isocyanates. The set of hinge binders was hand-picked to include scaffolds under-represented in the corporate collection and scaffolds for which further SAR exploration was desired. Each hinge binder possessed a primary or secondary amine to which an acyl, sulfonyl, or isocyanate

Scheme 1. Virtual Library 1 Was Enumerated from a Set of Hinge Binders (HB) and a Set of Acyl Chlorides, Sulfonyl Chlorides, and Isocyanates



Scheme 2. Virtual Library 2 Was Enumerated from a Set of Hinge Binders (HB), a Set of Bromophenyl Carboxylic Acids, and a Set of Primary Alkyl- and Aryl-Amines and Secondary Alkyl-Amines



could be coupled to yield an amide, sulfonamide, or urea derivative, respectively. A set of ~7250 acyl chlorides, sulfonyl chlorides, and isocyanates was initially compiled from a series of substructure searches of the corporate database and the Available Chemicals Directory (ACD).[29] The monomer set then was filtered, to eliminate those reagents exceeding a molecular weight of 300, a hydrogen-bond donor count of three, a hydrogen-bond acceptor count of five, and a rotatable bond count of six. These criteria were applied to the reagents' substructures that were to become part of the final products. As a final filter, the remaining reagents were visually inspected, to remove additional compounds that had undesirable functional groups (for example, groups that would interfere with synthesis). In imposing our molecular property criteria on the set of acyl chlorides, sulfonyl chlorides, and isocyanates, our goal was to ultimately compile a subset of reagents such that the vast majority of the combinatorial products would have molecular weights of ≤550, ≤5 hydrogen-bond donors, ≤10 hydrogen-bond acceptors, and ≤8 rotatable bonds. Roughly 99% of the compounds in Virtual Library 1 satisfied all four molecular properties.

Virtual Combinatorial Library 2. Virtual Library 2 was a three-component combinatorial library (see Scheme 2) that was enumerated from a set of 7 kinase hinge binders, 11 bromophenyl carboxylic acids, and 1255 primary alkyl- and aryl-amines and secondary alkyl-amines. The set of hinge binders was again hand-picked to include scaffolds under-represented in the corporate collection and scaffolds for which further SAR exploration was desired. The set of carboxylic acids was also hand-selected, with the intention of exploring the effects of flexibility in the central linker. In contrast, the set of amines was compiled from a series of substructure searches of the corporate database and ACD. The initial set was comprised of more than 10000 reagents, which were subsequently filtered to eliminate those exceeding a molecular weight of 270, a hydrogen-bond donor count of three, a hydrogen-bond acceptor count of five, and a rotatable bond count of six. Again, these criteria were applied to the reagents' substructures that were to be incorporated into the final products. The remaining amines were next visually inspected to remove additional compounds with undesirable functional groups. As was the goal with Virtual Library 1, the final subset of amines was assembled such that the vast majority of the combinatorial products would have molecular weights of ≤550, ≤5 hydrogen-bond donors, ≤10 hydrogen-bond acceptors, and ≤8 rotatable bonds. For Virtual Library 2, ~99% of the compounds satisfied all four molecular properties.

Library Enumeration and Virtual Screening. Virtual Libraries 1 and 2 were enumerated using CombiLibMaker,[30] which generated libraries with the expected theoretical number of virtual products (49266 and 96635 compounds, respectively). The output generated by CombiLibMaker was two sets of SMILES strings—one for each virtual library. An in-house software tool was written to take a SMILES file as input, generate 3D structures via Concord,[31] and call upon the PharmPrint tools kit to calculate fingerprints. Also incorporated into this tool was PipeComm,[32] which allowed for the distributed processing of all computational work across multiple CPUs. As a result of distributed processing, the time required to calculate the fingerprints for Virtual Library 1 was 10.3 CPU min on 32 SGI R12000 processors and 24.8 CPU min for Virtual Library 2. A second in-house software tool was also written to enrich the PharmPrint fingerprints of the library compounds in identical fashion to those of the three training sets and then compute predicted $p\mathrm{IC}_{50}$ values against each of the three protein kinases by applying the PharmPrint QSAR models for AKT1, Aurora-A, and ROCK1 inhibition. The time required to enrich the fingerprints and compute the predicted activities for Virtual Library 1 was 2.40 CPU min on a single SGI R12000 processor and 4.80 CPU min for Virtual Library 2.

Aurora-A and ROCK1 Targeted Arrays. Library design is often treated as a multiobjective optimization problem, where numerous parameters such as predicted activities and druglike properties are simultaneously optimized.[1,8] Because our primary goal was to prospectively assess the utility of the PharmPrint methodology as a virtual screening tool, had we included design parameters other than predicted activities, the PharmPrint evaluation would have taken on a greater level of complexity. Nevertheless, a sensible lead should possess suitable physicochemical properties to serve as a proper starting point for further drug development. Therefore, we imposed thresholds on molecular weight, hydrogen-bond donors, hydrogen-bond acceptors, and rotatable bonds, to eliminate "druglike" as an optimization parameter and, thus, reduce the complexity of the library design process and, in turn, simplify the PharmPrint evaluation. The Reactant-Biased, Product-Based (RBPB) library design algorithm within DiverseSolutions (DVS)[33] was used for array design, and by imposing our molecular property criteria, we were able to focus the design objective on simply maximizing the average predicted $p\mathrm{IC}_{50}$ values against the kinase targets of interest.

For Virtual Library 1, the distributions of predicted $p\mathrm{IC}_{50}$ values for the kinase panel were approximately normal, as illustrated in Figure 4. Note that this library did not score well against AKT1, as evidenced by its distribution of predicted activities with a mean of 5.01 and a small standard deviation of 0.20 log units. In fact, there were only five combinatorial products in the entire library with predicted $p\mathrm{IC}_{50}$ values of greater than 6.0, and the maximum predicted activity was only 6.08. In contrast, the distributions of predicted $p\mathrm{IC}_{50}$ values for Aurora-A and ROCK1 were shifted further right, compared to that of AKT1. The Aurora-A and ROCK1 distributions were fairly similar, with means of 5.95 and 5.88 and standard deviations of 0.42 and 0.48 log units, respectively. While it would have been appropriate to design an array to target both of these protein kinases or selectively target one over the other, the decision

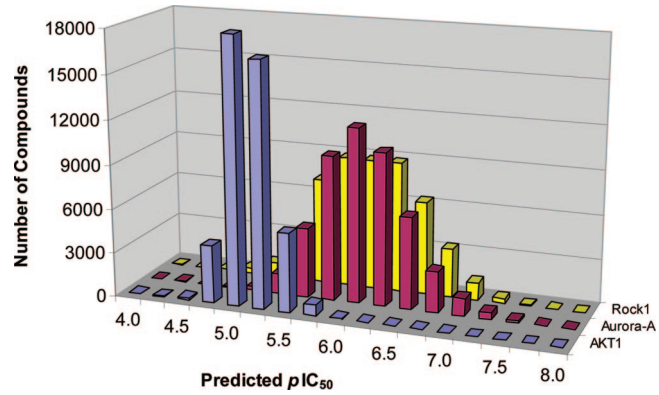


Figure 4. Distributions of predicted $p\mathrm{IC}_{50}$ values for AKT1, Aurora-A, and ROCK1 for the fully enumerated Virtual Library 1.

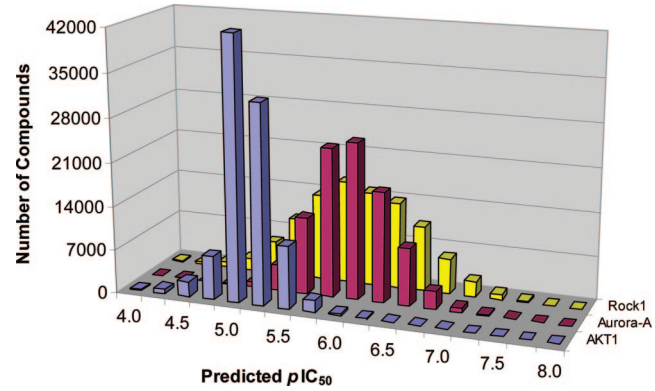


Figure 5. Distributions of predicted $p\mathrm{IC}_{50}$ values for AKT1, Aurora-A, and ROCK1 for the fully enumerated Virtual Library 2.

was made to target just one and not involve any others, to keep our PharmPrint evaluation simple. As such, we chose to design a small 48-member array targeting Aurora-A given that its $p\mathrm{IC}_{50}$ distribution had a 1.5-fold larger population of compounds at predicted activities of greater than or equal to 7.0 compared to that of ROCK1. The RBPB library design algorithm was used to design the array with the objective of maximizing the average predicted $p\mathrm{IC}_{50}$ value against Aurora-A. The array configuration was specified upfront as four kinase hinge binders and 12 reagents from the set of acyl chlorides, sulfonyl chlorides and isocyanates. The top-scoring candidate array reported by the RBPB algorithm was selected for synthesis.

Illustrated in Figure 5 are the distributions of predicted $p\mathrm{IC}_{50}$ values for AKT1, Aurora-A, and ROCK1 for Virtual Library 2. Note that this library also scored poorly against AKT1; the average predicted $p\mathrm{IC}_{50}$ value against this kinase was 5.03, with a standard deviation of 0.24 log units, and the highest scoring compound had a predicted activity of only 5.71. In contrast, the $p\mathrm{IC}_{50}$ distributions for Aurora-A and ROCK1 were centered further to the right at almost identical means. The average predicted $p\mathrm{IC}_{50}$ value against Aurora-A was 5.84, whereas for ROCK1, it was 5.77. The standard deviation of the ROCK1 activities, however, was modestly larger, compared to that of Aurora-A, with values of 0.57 and 0.37 log units, respectively. Note that the right tail-end of the ROCK1 distribution at $p\mathrm{IC}_{50}$ values of greater than or equal to 7.0 is heavier compared to that of Aurora-A. In fact, the compound population for this portion of the ROCK1 distribution was 3.2-fold larger than for that of Aurora-A. As a result, we chose to design a 600-member

Table 2. Summary of Kinase Inhibitory Data for the Aurora-A Targeted Array and Control Group

		active compounds			
protein kinase	array design	number	average pIC_{50} (SD)	pIC_{50} range	inactive compounds
AKT1	PharmPrint	0	NA	NA	43
	control	0	NA	NA	68
Aurora-A	PharmPrint	18	6.77 (0.63)	6.03 − 8.48	25
	control	0	NA	NA	68
ROCK1	PharmPrint	4	6.41 (0.40)	6.03 − 6.96	39
	control	8	6.44 (0.40)	6.03 − 7.07	60

Table 3. Summary of Compounds Correctly and Incorrectly Classified as Active and Inactive by the PharmPrint QSAR Models for the Aurora-A Targeted Array

	compounds predicted as active			compounds predicted as inactive		
protein kinase	total	active	inactive	total	inactive	active
AKT1	0	0 (NA)	0 (NA)	43	43 (100%)	0 (0%)
Aurora-A	42	18 (42.9%)	24 (57.1%)	1	1 (100%)	0 (0%)
ROCK1	34	3 (8.80%)	31 (91.2%)	9	8 (88.9%)	1 (11.1%)

array targeting ROCK1. Here again, the RBPB library design algorithm of DVS was used to design the target-focused array with the objective of maximizing the average predicted pIC_{50} value against ROCK1. The array configuration was specified upfront as 4 hinge binders, 6 bromophenyl carboxylic acids, and 25 amines. The top-scoring candidate array reported by the RBPB algorithm was selected for synthesis.

Aurora-A and ROCK1 Control Groups. Rather than compare the hit rates of our designed arrays against those from a random collection of compounds, we chose instead to synthesize control arrays from monomers that were hand-picked on the basis of experience with kinase X-ray structures, known kinase inhibitor SAR, and an intuitive assessment of chemical diversity. One array was designed in this manner from each combinatorial library. From Virtual Library 1, four kinase hinge binders and 25 reagents from the set of acyl chlorides, sulfonyl chlorides, and isocyanates were hand-selected that were distinctly different from the PharmPrint-based selections to produce the Aurora-A control group. From Virtual Library 2, 26 amines were hand-picked that were distinctly different from the PharmPrint-based selections, in addition to 6 of the 7 kinase hinge binders and 10 of the 11 bromophenyl carboxylic acids to produce the ROCK1 control group. Given that almost all reagents for the latter two monomer sets were selected, it was obviously not surprising to learn that all PharmPrint-based selections were part of the hand-picked set.

RESULTS AND DISCUSSION

The Aurora-A and ROCK1 targeted arrays were designed in efforts to evaluate the utility of the PharmPrint methodology as a virtual HTS tool. To this end, PharmPrint QSAR models for the prediction of AKT1, Aurora-A, and ROCK1 inhibition were constructed and used in the design of our target-focused arrays. In this section, the biological results obtained for the designed arrays and their corresponding control groups are summarized. To allow comparison of results between the designed and control arrays, we established at the outset of the study definitions for active and inactive compounds. An active compound was defined as one with an experimental pIC_{50} value of greater than or equal to 6.0 against a given kinase target, whereas a compound with a pIC_{50} value of less than 6.0 was defined as inactive. Based on these definitions, we assessed the predictive abilities of the PharmPrint QSAR models by evaluating the numbers of compounds correctly and incorrectly classified as active and inactive. The performance of computational design versus control was also assessed by evaluating the enrichment levels achieved by the QSAR models. The enrichment levels were calculated simply from the ratios of hit rates between the designed and control arrays.

Aurora-A Targeted Array and Control. There were a total of 48 combinatorial products possible for the Aurora-A targeted array, of which 43 were successfully synthesized. For the control group, 100 combinatorial products were theoretically possible, but only 68 were synthesized. After all compounds had undergone quality assurance, they were registered and screened against the kinase panel.

Although it is instructive to compare predicted versus experimental pIC_{50} values for the target-focused array, it was not reasonable to do so given the narrow ranges of predicted activities associated with these compounds against all three kinases. For example, against the intended target, Aurora-A, ~79% of the compounds had predicted pIC_{50} values that were within 0.50 log units of the mean predicted activity of the array, which was not surprising, considering our library design objective. However, to properly assess the predictive ability of the Aurora-A QSAR model, a set of compounds with a broader and more populous range of predicted activities would be needed. Almost identical observations were also made and conclusions drawn for AKT1 and ROCK1. Despite these limitations, the numbers of compounds correctly and incorrectly classified as active and inactive could still be assessed and the enrichment levels achieved by the PharmPrint methodology over the control set could be evaluated.

A summary of biological results for the Aurora-A targeted array and the control group is listed in Table 2 for all three kinases. Note that of the 43 compounds in the designed array, 18 were active against the intended target with an average experimental pIC_{50} value of 6.77 and a standard deviation of 0.63 log units. The range of measured activities for the actives set was 6.03 to 8.48. Not all 43 compounds, however, were predicted to be active against Aurora-A by the

corresponding QSAR model. Table 3 lists the numbers of compounds correctly and incorrectly classified as active and inactive by the PharmPrint QSAR models across the kinase panel. Note that, for Aurora-A, 42 compounds were actually predicted to be active. Thus, the hit rate for the Aurora-A QSAR model was 42.9%. For the control group, all compounds were inactive against Aurora-A with $p\mathrm{IC}_{50}$ values of less than 5.80. Given that the hit rate for the control group was 0%, the enrichment level achieved by the PharmPrint QSAR model was incalculable.

There were also 24 inactive compounds predicted to be active against Aurora-A (see Table 3). In efforts to understand why the corresponding QSAR model had incorrectly predicted their biological outcomes, their chemical structures, PharmPrint fingerprints, and predicted $p\mathrm{IC}_{50}$ values were carefully scrutinized against known SAR. From this analysis, the predicted activities were rationalized as the compounds possessed the pharmacophoric features needed for Aurora-A inhibition. However, there were also structural and chemical features associated with these compounds that were not observed in those of our training set. Thus, we concluded that these features were, at least in part, the cause of the discrepancies encountered between the predicted and experimental values. While we recognize that no CADD-related approach is perfect, we realize that the scope of a QSAR model's predictive ability is only as broad as the SAR data present in a training set.

Finally, note from Table 3 that the one compound predicted to be inactive by the Aurora-A QSAR model was inactive; its experimental $p\mathrm{IC}_{50}$ value was less than 5.0.

Four compounds from the Aurora-A targeted array were also found to be active against ROCK1, with an average $p\mathrm{IC}_{50}$ value of 6.41 and a standard deviation of 0.40 log units. Their range of measured activities was 6.03− 6.96. Although this array was not designed to target ROCK1, 34 compounds were predicted to be active by the corresponding QSAR model. Still, their distribution of predicted activities against this kinase was centered at a $p\mathrm{IC}_{50}$ value of only 6.23 (standard deviation of 0.40 log units), which lay near the dividing line between active and inactive. In contrast, the distribution of predicted activities against Aurora-A was centered at a much higher value, a $p\mathrm{IC}_{50}$ of 6.95 (standard deviation of 0.45 log units). Based on these observations and given the magnitude of the standard error of prediction of the ROCK1 QSAR model, a higher false positive rate was expected, compared to that observed with Aurora-A, which was indeed the case. Only 3 of the 34 compounds predicted to be active were actually active, for a hit rate of 8.80%. For the control group, 8 of its 68 compounds were active against ROCK1 with an average $p\mathrm{IC}_{50}$ value of 6.44 and a standard deviation of 0.40 log units. The range of activities here was 6.03−7.07. Although the hit rate for the control group was higher (11.8%), it was merely 1.3-fold above that of the designed array.

Of the nine compounds predicted to be inactive against ROCK1 for the Aurora-A targeted array, eight were actually inactive. Thus, the percentage of compounds correctly classified as inactive was 88.9%. The experimental $p\mathrm{IC}_{50}$ value for the one false negative was 6.20, which was not significantly different from its predicted $p\mathrm{IC}_{50}$ value of 5.80 and certainly was well within the standard error of prediction.

Recall that Virtual Library 1 had scored poorly against AKT1. Not surprisingly, for the designed array, the average predicted $p\mathrm{IC}_{50}$ value against this kinase was 5.24 with a small standard deviation of 0.19 log units. Also, not one compound was predicted to have an activity of greater than 5.70. The PharmPrint-based prediction that all 43 compounds would be inactive against AKT1 proved correct; all compounds had experimental $p\mathrm{IC}_{50}$ values of less than 5.10. For the control group, all 68 compounds were also inactive, with measured $p\mathrm{IC}_{50}$ values of less than 5.40 against AKT1.

ROCK1 Targeted Array and Control. Of the 600 possible combinatorial products for the ROCK1 targeted array, a total of 297 compounds were successfully synthesized. The loss of such a large fraction of compounds might lead one to speculate as to whether an uneven number of predicted actives or inactives were not synthesized. This was not the case as the loss of compounds was fairly even across the distribution of predicted activities. Essentially the same success rate of synthesis was observed for the control group. Of the 1560 theoretically possible products, only 749 were synthesized. Once all compounds had undergone quality assurance, they were registered and screened against the kinase panel.

As we saw with the Aurora-A targeted array, there were also narrow ranges of predicted activities associated with the compounds from the ROCK1 targeted array against the kinase panel. For example, against ROCK1, ∼80% of the compounds had predicted $p\mathrm{IC}_{50}$ values that were within 0.50 log units of the mean predicted activity of the array. However, to properly assess the predictive ability of the ROCK1 QSAR model, we knew that a set of compounds with a broader and more populous range of predicted activities was needed. Similar observations were also made and conclusions drawn for AKT1 and Aurora-A. As observed previously, despite these limitations, the numbers of compounds correctly and incorrectly classified as active and inactive could still be assessed and the enrichment levels achieved by the PharmPrint methodology over the control set could be evaluated.

A summary of biological results for the ROCK1 targeted array and the control group is listed in Table 4 for all three kinases. Note that 58 of the 297 compounds from the designed array were active against the intended target. The average experimental $p\mathrm{IC}_{50}$ value for the actives set was 6.55, with a standard deviation of 0.44 log units and the range of measured activities was 6.0−7.86. Here again, not all 58 compounds were predicted to be active against ROCK1 by the corresponding QSAR model. Table 5 lists the numbers of compounds correctly and incorrectly classified as active and inactive by the PharmPrint QSAR models across the kinase panel. Note that, of the 144 compounds predicted to be active against ROCK1, 44 were actually active, for a hit rate of 30.6%. In the control group, 38 of the 749 compounds were active against the kinase, with an average $p\mathrm{IC}_{50}$ value of 6.46 and a standard deviation of 0.43 log units (see Table 4). The range of activities here was 6.01−7.77. Given that the hit rate for the control group was 5.10%, the enrichment level achieved by the PharmPrint QSAR model against the intended target was 6-fold times greater than that of the control.

As a point of interest, only three compounds from the designed array were predicted to have $p\mathrm{IC}_{50}$ values of greater

Table 4. Summary of Kinase Inhibitory Data for the ROCK1 Targeted Array and Control Group

		active compounds			
protein kinase	array design	number	average $p\mathrm{IC}_{50}$ (SD)	$p\mathrm{IC}_{50}$ range	inactive compounds
AKT1	PharmPrint	5	6.51 (0.47)	6.02−7.24	292
	control	5	6.49 (0.32)	6.26−7.00	744
Aurora-A	PharmPrint	11	6.25 (0.26)	6.02−6.91	286
	control	6	6.19 (0.16)	6.04−6.45	743
ROCK1	PharmPrint	58	6.55 (0.44)	6.00−7.86	239
	control	38	6.46 (0.43)	6.01−7.77	711

Table 5. Summary of Compounds Correctly and Incorrectly Classified as Active and Inactive by the PharmPrint QSAR Models for the ROCK1 Targeted Array

	compounds predicted as active			compounds predicted as inactive		
protein kinase	total	active	inactive	total	inactive	active
AKT1	0	0 (NA)	0 (NA)	297	292 (98.3%)	5 (1.70%)
Aurora-A	210	7 (3.30%)	203 (96.7%)	87	83 (95.4%)	4 (4.60%)
ROCK1	144	44 (30.6%)	100 (69.4%)	153	139 (90.8%)	14 (9.20%)

than 7.0 against ROCK1. In part, this was due to the fact that roughly half of the compounds were not synthesized; however, the large size of the designed array was also significant. The average predicted $p\mathrm{IC}_{50}$ value against ROCK1 for the 600 virtual compounds was 6.29, with a standard deviation of 0.47 log units. If a parallel array the size of the Aurora-A targeted array had been designed, the virtual library would have been better sampled at the high-end of the $p\mathrm{IC}_{50}$ distribution, leading to an array with a significantly higher average predicted activity. To explore this, we undertook the exercise where a 48-member array was designed; its average predicted $p\mathrm{IC}_{50}$ value was 7.30 with a standard deviation of 0.32 log units. At this point, we were left to speculate that had the smaller designed array been synthesized, its hit rate may have been more comparable to that of the Aurora-A targeted array, given its higher average predicted activity.

Note from Table 5 that 100 inactive compounds were also predicted to be active against ROCK1. As previously noted, to gain an understanding of why the corresponding QSAR model had incorrectly predicted their biological outcomes, their chemical structures, PharmPrint fingerprints, and predicted $p\mathrm{IC}_{50}$ values were carefully scrutinized against known SAR. From this analysis, the predicted activities were rationalized, as the compounds possessed the pharmacophoric features needed for ROCK1 inhibition. However, here again, there were structural and chemical features associated with these compounds that were not observed in those of our training set. Given these observations, we were led to the same conclusion as that of the inactive compounds predicted active against Aurora-A.

Of the 153 designed products predicted to be inactive against ROCK1, 139 had experimental $p\mathrm{IC}_{50}$ values of less than 6.0 (see Table 5). Thus, the percentage of compounds correctly classified as inactive was 90.8%. Fourteen compounds, however, proved to be false negatives, four of which had their biological activities underestimated by more than 1.0 log unit. The worst case was for an inhibitor that had a predicted $p\mathrm{IC}_{50}$ value of 5.47 but an experimental value of 7.40. Seven of the ten remaining false negatives had their experimental $p\mathrm{IC}_{50}$ values underestimated by 0.50−1.0 log units, while the last three had their activities underestimated by less than 0.50 log units. The chemical structures and PharmPrint fingerprints of the top 10 false negatives were carefully scrutinized along with their predicted $p\mathrm{IC}_{50}$ values which were underestimated by more than 0.60 log units. From this analysis, we observed that these compounds all possessed multiple, identical chemical substructures, which, in turn, resulted in multiple, identical three-point pharmacophores, most of which made positive contributions to the PharmPrint score. Because the contribution from each of these pharmacophores was added only once to the predicted activities, we reasoned that, at least in part, this had led to the poor estimates of biological activity. As a final point, compounds with multiple, identical chemical substructures were not found within the training set.

Eleven compounds from the ROCK1 targeted array were also determined to be active against Aurora-A, with an average $p\mathrm{IC}_{50}$ value of 6.25 and a standard deviation of 0.26 log units. The range of measured activities was 6.02−6.91. Although this target-focused array was not designed to target Aurora-A, 210 of the 297 compounds were predicted to be active by the corresponding PharmPrint QSAR model. However, here again, we observed that the average predicted $p\mathrm{IC}_{50}$ value against this kinase was 6.15 (standard deviation of 0.29 log units), which centered the distribution of predicted activities very close to the dividing line between active and inactive. Based on this observation and given the magnitude of the standard error of prediction of the Aurora-A QSAR model, we expected a high false positive rate for the designed array, which was indeed the case. Seven of the 210 compounds predicted to be active were actually active, for a hit rate of 3.30%. In the control group, only 6 of the 749 compounds were active against Aurora-A, with an average $p\mathrm{IC}_{50}$ value of 6.19 and a standard deviation of 0.16 log units. The range of activities here was 6.04−6.45. Because the hit rate for the control group was merely 0.80%, the PharmPrint QSAR model did manage to achieve a 4.2-fold enrichment over the control set, despite its own low hit rate.

In the designed array, 87 compound were predicted to be inactive against Aurora-A, 83 of which were confirmed to be inactive. Thus, the percentage of compounds correctly classified as inactive was 95.4%. Obviously, the remaining four compounds were false negatives. However, it is important to note that two of these compounds had predicted $p\mathrm{IC}_{50}$ values that were within 0.20 log units of their

experimental values. The remaining two compounds had predicted activities that were within 0.60 log units of their experimental $p\mathrm{IC}_{50}$ values.

Recall that Virtual Library 2 had also scored poorly against AKT1. As such, it was not surprising that the average predicted $p\mathrm{IC}_{50}$ value against AKT1 for the designed array was only 5.13 with a small standard deviation of 0.26 log units. In addition, not one compound had a predicted $p\mathrm{IC}_{50}$ value of greater than 5.75. The PharmPrint-based prediction that all 297 compounds would be inactive against AKT1 proved to be mostly correct. Two-hundred ninety two compounds were inactive against AKT1. Thus, 98.3% of the compounds were correctly classified as inactive. The five remaining compounds were clearly false negatives, with an average $p\mathrm{IC}_{50}$ value of 6.51 and a standard deviation of 0.47 log units. Three of these compounds had their biological activities underestimated by more than 1.0 log unit. The worst case was for an inhibitor that had a predicted $p\mathrm{IC}_{50}$ value of 5.53, but an experimental value of 7.24. The two remaining false negatives had their biological activities underestimated by roughly 0.90 log units. Because these five compounds possessed multiple, identical substructures, we assumed that their biological activities had been underestimated for the same reasons as those cited for the ROCK1 false negatives. In the control group, five of the 749 compounds were active against AKT1, for a hit rate of 0.67%. The average $p\mathrm{IC}_{50}$ value for the actives was 6.49, with a standard deviation of 0.32 log units and a range of activities of 6.26−7.0.

CONCLUSIONS

In this prospective study, our in-house version of the PharmPrint methodology was evaluated as a virtual high-throughput screening (HTS) tool by applying it to the design of two kinase-targeted arrays. First, PharmPrint quantative structure−activity relationship (QSAR) models for the prediction of AKT1, Aurora-A, and ROCK1 inhibition were constructed by fitting experimental $p\mathrm{IC}_{50}$ values against enriched PharmPrint fingerprints. A 9-component QSAR model was constructed from the set of AKT1 inhibitors, which accounted for 66.1% of the variance in the experimental data, with a standard error of prediction of 0.51 log units. The PharmPrint QSAR model for Aurora-A inhibition was comprised of 14 PLS components and was able to account for 81% of the variance in the biological data, with a standard error of prediction of 0.54 log units. Lastly, a 13-component QSAR model was constructed from the set of ROCK1 inhibitors, which accounted for 74.2% of the variance in the experimental data, with a standard error of prediction of 0.56 log units.

The PharmPrint QSAR models were used to virtually screen two large combinatorial libraries, which were designated Virtual Libraries 1 and 2, by computing the predicted $p\mathrm{IC}_{50}$ values for all compounds. The predicted activities then were used to profile each library, to identify which kinase to target with which library. Given that Virtual Library 1 had scored well against Aurora-A, a 48-member array was designed to target this kinase, of which 43 compounds were successfully synthesized. Since Virtual Library 2 had scored well against ROCK1, a 600-member array was also designed to target this kinase, of which only 297 compounds were synthesized. One control group was synthesized for each designed array, which would then allow for the calculation of enrichment levels that might be produced by the PharmPrint methodology. The reagent selections for the control groups were kinase-biased, where the choices made were based on experience with kinase X-ray structures, known kinase inhibitor structure-activity relationships (SAR), and an intuitive assessment of chemical diversity.

All four sets of synthesized compounds were screened against the kinase panel. The biological results for the kinase-targeted arrays against the intended targets revealed promising hit rates and demonstrated that the PharmPrint QSAR models for Aurora-A and ROCK1 inhibition had produced significant enrichments over the control groups. For the Aurora-A targeted array, the hit rate against its intended target was 42.9%. Given that the corresponding control group yielded no hits against Aurora-A, the enrichment level achieved by the Aurora-A QSAR model was incalculable. For the ROCK1 targeted array, the hit rate against its intended target was 30.6%, whereas that of the control group was 5.10%. Thus, the enrichment level achieved by the ROCK1 QSAR model was 6-fold greater than the control. When the Aurora-A and ROCK1 targeted arrays were screened against ROCK1 and Aurora-A, respectively, the observed hit rates were significantly lower, but not unexpected, because the predicted $p\mathrm{IC}_{50}$ values were clustered near the dividing line between active and inactive, and, therefore, we had anticipated high false positive rates. The hit rate for the Aurora-A targeted array against ROCK1 was 8.80%, which was modestly lower than that of the control group at 11.8%. While the hit rate for the ROCK1 targeted array against Aurora-A was only 3.30%, the hit rate for the corresponding control group was significantly lower at merely 0.80%.

In addition to predicted actives, numerous compounds from both kinase-targeted arrays were predicted to be inactive by the PharmPrint QSAR models. One compound from the Aurora-A targeted array was predicted to be inactive against Aurora-A, and that compound was inactive. Nine Aurora-A targeted compounds were also predicted to be inactive against ROCK1, of which eight (88.9%) were inactive. The PharmPrint-based prediction that the Aurora-A targeted array would be inactive against AKT1 was proven entirely correct with all 43 compounds inactive against the kinase. For the ROCK1 targeted array, 153 compounds were predicted to be inactive against ROCK1 of which 139 (90.8%) were inactive. Of the 87 ROCK1 targeted compounds predicted to be inactive against Aurora-A, 83 (95.4%) were inactive. Lastly, the PharmPrint-based prediction that the ROCK1 targeted array would be inactive against AKT1 was proven mostly correct with 292 (98.3%) of the 297 compounds inactive against the kinase.

Given the high hit rates achieved against the intended kinase targets, our results support the use of the PharmPrint methodology as a virtual HTS tool for the design of kinase-targeted arrays. There were also high percentages of compounds correctly classified as inactive across the kinase panel for both designed arrays, which indicate that the PharmPrint methodology could be used to design compounds selective for a given kinase. While designing in kinase selectivity was not part of our objectives, we recognize that this is a critical issue facing medicinal chemists who are working with kinase inhibitors. As such, we intend to investigate this matter

further by designing arrays that maximize PharmPrint predicted activities against a given kinase while also minimizing the activities against others. This will be the topic of a future communication.

REFERENCES AND NOTES

(1) Gillet, V. J. Designing Combinatorial Libraries Optimized on Multiple Objectives. In *Chemoinformatics: Concepts, Methods, and Tools for Drug Discovery*; Bajorath, J., Ed.; Humana Press: Totowa, NJ, 2004; Vol. 275, pp 335−354.

(2) Tommasi, R.; Cornella, I. Focused Libraries: The Evolution in Strategy from Large-Diversity Libraries to the Focused Library Approach. In *Exploiting Chemical Diversity for Drug Discovery*; Bartlett, P. A., Entzeroth, M., Eds.; Royal Society of Chemistry: Cambridge, U.K., 2006; Vol. 163, pp 163−183.

(3) Burello, E.; Bologa, C.; Frecer, V.; Miertus, S. Application of Computer Assisted Combinatorial Chemistry in Antiviral, Antimalarial and Anticancer Agents Design. *Mol. Phys.* **2002**, *100*, 3187–3198.

(4) Deng, Z.; Chuaqui, C.; Singh, J. Knowledge-Based Design of Target-Focused Libraries Using Protein−Ligand Interaction Constraints. *J. Med. Chem.* **2006**, *49*, 490–500.

(5) Sun, Y.; Ewing, T. J. A.; Skillman, A. G.; Kuntz, I. D. CombiDOCK: Structure-Based Combinatorial Docking and Library Design. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 597–604.

(6) Frecer, V.; Burello, E.; Miertus, S. Combinatorial Design of Non-symmetrical Cyclic Urea Inhibitors of Aspartic Protease of HIV-1. *Bioorg. Med. Chem.* **2005**, *13*, 5492–5501.

(7) Chen, G.; Zheng, S.; Luo, X.; Shen, J.; Zhu, W.; Liu, H.; Gui, C.; Zhang, J.; Zheng, M.; Puah, C. M.; Chen, K.; Jiang, H. Focused Combinatorial Library Design Based on Structural Diversity, Drug-likeness and Binding Affinity Score. *J. Comb. Chem.* **2005**, *7*, 398–406.

(8) Pickett S. D. Library Design: Reactant and Product-Based Approaches. In *Comprehensive Medicinal Chemistry II*; Taylor, J. B., Triggle, D. J., Eds.; Elsevier: Oxford, U.K., 2006; Vol. 4, pp 337−378.

(9) Thomas, M. P.; McInnes, C.; Fischer, P. M. Protein Structures in Virtual Screening: A Case Study with CDK2. *J. Med. Chem.* **2006**, *49*, 92–104.

(10) Todorov, N. P.; Buenemann, C. L.; Alberts, L. Combinatorial Ligand Design Targeted at Protein Families. *J. Chem. Inf. Model.* **2005**, *45*, 314–320.

(11) Polgar, T.; Baki, A.; Szendrei, G. I.; Keseru, G. M. Comparative Virtual and Experimental High-Throughput Screening for Glycogen Synthase Kinase-3β Inhibitors. *J. Med. Chem.* **2005**, *48*, 7946–7959.

(12) Forino, M.; Jung, D.; Easton, J. B.; Houghton, P. J.; Pellecchia, M. Virtual Docking Approaches to Protein Kinase B Inhibition. *J. Med. Chem.* **2005**, *48*, 2278–2281.

(13) Krier, M.; Araujo-Junior, J. X.; Schmitt, M.; Duranton, J.; Justiano-Basaran, H.; Lugnier, C.; Bourguignon, J.-J.; Rognan, D. Design of Small-Sized Libraries by Combinatorial Assembly of Linkers and Functional Groups to a Given Scaffold: Application to the Structure-Based Optimization of a Phosphodiesterase 4 Inhibitor. *J. Med. Chem.* **2005**, *48*, 3816–3822.

(14) Zhang, Y. J.; Wang, Z.; Sprous, D.; Nabioullin, R. Silico Design and Synthesis of Piperazine-1-Pyrrololidine-2,5-dione Scaffold-Based Novel Malic Enzyme Inhibitors. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 525–528.

(15) Yoon, S.; Smellie, A.; Hartsough, D.; Filikov, A. Surrogate Docking: Structure-Based Virtual Screening at High Throughput Speed. *J. Comput.-Aided Mol. Des.* **2005**, *19*, 483–497.

(16) Joseph-McCarthy, D.; Tsang, S. K.; Filman, D. J.; Hogle, J. M.; Karplus, M. Use of MCSS to Design Small Targeted Libraries: Application to Picornavirus Ligands. *J. Am. Chem. Soc.* **2001**, *123*, 12758–12769.

(17) Wyss, P. C.; Gerber, P.; Hartman, P. G.; Hubschwerlen, C.; Locher, H.; Marty, H.-P.; Stahl, M. Novel Dihydrofolate Reductase Inhibitors. Structure-Based versus Diversity-Based Library Design and High-Throughput Synthesis and Screening. *J. Med. Chem.* **2003**, *46*, 2304–2312.

(18) Waszkowycz, B. Structure-Based Approaches to Drug Design and Virtual Screening. *Curr. Opin. Drug Discovery Dev.* **2002**, *5*, 407–413.

(19) Fernandes, M. X.; Kairys, V.; Gilson, M. K. Comparing Ligand Interactions with Multiple Receptors via Serial Docking. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1961–1970.

(20) Jacobsson, M.; Karlen, A. Ligand Bias of Scoring Functions in Structure-Based Virtual Screening. *J. Chem. Inf. Model* **2006**, *46*, 1334–1343.

(21) Tong, W.; Lowis, D. R.; Perkins, R.; Chen, Y.; Welsh, W. J.; Goddette, D. W.; Heritage, T. W.; Sheehan, D. M. Evaluation of Quantitative Structure−Activity Relationship Methods for Large-Scale Prediction of Chemicals Binding to the Estrogen Receptor. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 669–677.

(22) Martin, Y. C. 3D QSAR: Current State, Scope, and Limitations. In *3D QSAR in Drug Design: Recent Advances*; Kubinyi, H., Folkers, G., Martin, Y. C., Eds.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1998; Vol. 12/13/14, pp 3−23.

(23) Golbraikh, A.; Shen, M.; Xiao, Z.; Xiao, Y.; Lee, K.; Tropsha, A. Rational selection of training and test sets for the development of validated QSAR models. *J. Comput.-Aided Mol. Des.* **2003**, *17*, 241–253.

(24) Stanton, D. T. On the Physical Interpretation of QSAR Models. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1423–1433.

(25) Deanda, F.; Stewart, E. L. Application of the PharmPrint Methodology to Two Protein Kinases. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1803–1809.

(26) McGregor, M. J.; Muskal, S. M. Pharmacophore Fingerprinting. 1. Application to QSAR and Focused Library Design. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 569–574.

(27) McGregor, M. J.; Muskal, S. M. Pharmacophore Fingerprinting. 2. Application to Primary Library Design. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 117–125.

(28) *The SAS System for Windows, Version 8.01*; SAS Institute: Cary, NC, 2000.

(29) *Available Chemicals Directory (ACD)*, MDL Information Systems: San Leandro, CA, 2003.

(30) CombiLibMaker, Version 4.3.2; Tripos, Inc.: St. Louis, MO, 2005.

(31) Concord, Version 4.0.7; Tripos, Inc.: St. Louis, MO, 2002.

(32) PipeComm, Version 2.1; The University of Texas: Austin, TX, 1998.

(33) DiverseSolutions, Version 6.2; Tripos, Inc.: St. Louis, MO, 2005.

CI800276T