Toward a Pharmacophore for Kinase Frequent Hitters

Alex M. Aronov* and Mark A. Murcko†

Vertex Pharmaceuticals Inc., Cambridge, Massachusetts 02139-4242

Received March 15, 2004

Abstract: Small molecule protein kinase inhibitors are widely employed as biological reagents and as leads in the design of drugs for a variety of diseases. One of the hardest challenges in kinase inhibitor design is achieving target selectivity. By utilizing X-ray structural information for four promiscuous inhibitors, we propose a five-point pharmacophore for kinase frequent hitters, demonstrate its ability to discriminate between frequent hitters and selective ligands, and suggest a strategy for selective inhibitor design.

Protein phosphorylation regulates most aspects of cell life, whereas abnormal phosphorylation is a cause or consequence of disease.[1] The importance of protein phosphorylation in eukaryotic signaling is reflected in the fact that protein kinase domains are found in ∼2% of eukaryotic genes.[2] Sequencing of the human genome has revealed at least 500 distinct kinases, which can be grouped into ∼20 known families on the basis of their structural relatedness.[3]

Protein kinases have now become the second most exploited group of drug targets after G-protein-coupled receptors (GPCR), accounting for 20−30% of drug discovery projects at many pharmaceutical companies.[1] Small molecule kinase inhibitors have generated much interest, both as potential therapeutics and as experimental tools for understanding the physiological roles of these enzymes. Various small molecule target-selective inhibitors of disease-relevant protein kinases are currently in different stages of clinical testing, and the first representatives of this class have already received FDA approval.[4,5]

The overwhelming majority of protein kinase inhibitors currently in clinical development are directed toward the adenosine triphosphate (ATP) binding site.[6] Despite the ATP binding site having acquired the status of a 'druggable' target, a number of concerns remain.[6] The two primary drawbacks of this approach to kinase inhibition are (i) the need for sufficient potency to compete with high intracellular ATP concentrations present in vivo, and (ii) the ubiquitous nature of the ATP binding site.[1,6]

Due to fold conservation and sequence similarity in the ATP site that many kinases share, as well as the commonality of the catalytic mechanism across the kinase families, one of the hardest challenges in kinase inhibitor design is achieving target selectivity.[1,6,7] Knowledge of the inhibitor's selectivity profile is a prerequisite for the accurate interpretation of its biological effects.[8] It seems likely that unpredictable pharmacology and unforseen patient side effects often encountered with these agents have arisen due to a lack of specificity.[9,10] Experimentally, inhibitor selectivity is typically assessed in parallel enzymatic assays for a set of recombinant protein kinases. Alternatively, affinity chromatography employing immobilized kinase inhibitors has been successfully used for proteome-wide assessment of inhibitor selectivity.[8,11,12] The structural information collected on kinases has gone a long way to explain the markedly different fingerprints of kinase inhibitory activity. However, crystal structures have been determined to date for some 30 protein kinases,[13] which represent less than 6% of the protein kinases in the human genome.[3] While the ability of homology models to pick up some of the crucial selectivity information has been demonstrated,[10] structural selectivity prediction still presents a major computational challenge. A wide range of flexibility, which includes side chain motion within the ATP binding site, loop flexibility, and domain motion between the N- and C-terminal kinase lobes, has led to the observation that the interkinase structural variation is not significantly greater than the intrakinase variation.[10] The corollary is that accurate prediction of kinase specificity for an ATP-competitive ligand may be difficult to achieve.

The concept that particular molecules have a high propensity for binding to protein targets has been an active research area for many years. According to the original definition by Evans,[14] a "privileged substructure" is a "single molecular framework able to provide ligands for diverse receptors". Privileged substructures have been identified for known drugs,[15] protein binding,[16] and GPCRs.[17,18] As part of a larger paradigm shift in the industry from target 'cherry picking' to broader chemogenomics approaches to gene family drug discovery, a recent trend has seen a move away from large, generalized libraries of diverse compounds toward smaller, focused libraries often designed and directed against target families.[18] Structural requirements for activity of a compound against a target family of interest can be derived from structural information on the target gene family or from the analysis of known ligands. The consistent ability to identify these substructures is critical to the process of library focusing, but it has proven to be elusive thus far. In related work by Mason[17] and co-workers, pharmacophore fingerprints have been used to capture the range of 3D properties seen in target family active compounds, regardless of their specific target. However, to our knowledge, this Letter constitutes the first report of a single pharmacophore that captures the combination of features common to promiscuous inhibitors of the target family. Herein, we will demonstrate a different approach to selectivity prediction in the kinase gene family format. Instead of attempting to predict specificity of ligands for select targets, we chose to profile those molecular features that tend to coincide with a propensity for nonselective inhibition of multiple kinases. Borrowing from the literature, we termed these ligands 'kinase frequent hitters'. An important distinction between this type of kinase inhibitors and nonspecific kinase inhibi-

* To whom correspondence should be addressed: Vertex Pharmaceuticals Inc., 130 Waverly St., Cambridge, MA 02139-4242. Telephone: (617) 444-6804. Fax: (617) 444-6566. E-mail: alex_aronov@vrtx.com.

† E-mail: mark_murcko@vrtx.com.

10.1021/jm049793g CCC: $27.50
Published on Web 10/12/2004

tors as described by Shoichet and co-workers[7,19] is that we investigated ligands exhibiting ATP-competitive kinase inhibition at nanomolar concentrations, as opposed to the promiscuous aggregators acting at micromolar or higher concentrations that have been reported previously. By utilizing X-ray structural information for four promiscuous inhibitors representing two distinct chemical classes, we propose a pharmacophore for kinase frequent hitters and demonstrate its ability to discriminate between frequent hitters and selective ligands. We also relate the chemical space information about kinase frequent hitters to the biological space, i.e., the structural information about contacts with specific conserved residues within the kinase ATP binding site that act to predispose a kinase ligand to promiscuous behavior.

We assembled the frequent hitter dataset and the selective datasets from the kinase screening data compiled from our corporate collection. The following five kinases were chosen to represent the kinase space: protein kinase A (PKA), Src, Cdk2, Erk2, and Gsk3. Of these, PKA is a member of the AGC family of serine-threonine kinases, Src belongs to the tyrosine kinase superfamily, and the remaining three kinases represent various evolutionary branches of the CMGC family of serine-threonine kinases.[20] The highest amino acid pairwise identity for this kinase set is 33% between Gsk3 and Cdk2. While one could envision a different set of targets, our choice was primarily aimed at the construction of a sizable matrix of data representing diverse ligands and kinases, and the five aforementioned kinases represent some of the most common therapeutic and counterscreening targets from the kinase superfamily, as signified by an average of ca. 7500 references on Medline[21] for each kinase. Activity based on the determined K_I was partitioned into three bins: green (K_I < 50 nM), yellow (50 nM < K_I < 2 μM), and red (K_I > 2 μM). Each compound can then be represented by an activity fingerprint containing five elements, one per kinase. The set of 43 frequent hitter compounds was selected which contained at least two green and no red elements, i.e., the frequent hitters are <2 μM against all five targets. The selective inhibitor set (Selective-50, 209 compounds) was selected which contained only one green element. Two subsets of the Selective-50 set were generated by setting the activity threshold higher for the green bin, to 10 nM and 2 nM, respectively, for Selective-10 (130 compounds) and Selective-2 (61 compounds) subsets. In other words, Selective-10 and Selective-2 contain molecules that are >5-fold and >25-fold selective for a single kinase, respectively. All of the datasets appeared evenly diverse, as shown by the internal pairwise Tanimoto similarity calculated using Daylight[22] fingerprints (Table 2). The activity fingerprints for the frequent hitter set and Selective-50 are shown in Figure 1.

Three-dimensional coordinates for the molecules in the datasets were generated with CORINA.[23] Multiconformer databases were then prepared for pharmacophore query searching using OMEGA.[24] Up to 1000 conformers/molecule were allowed, the energy window was set to 20 kcal/mol, and root-mean-square deviation (RMSD) cutoff of 1 Å was applied to ensure adequate and diverse conformational representation.

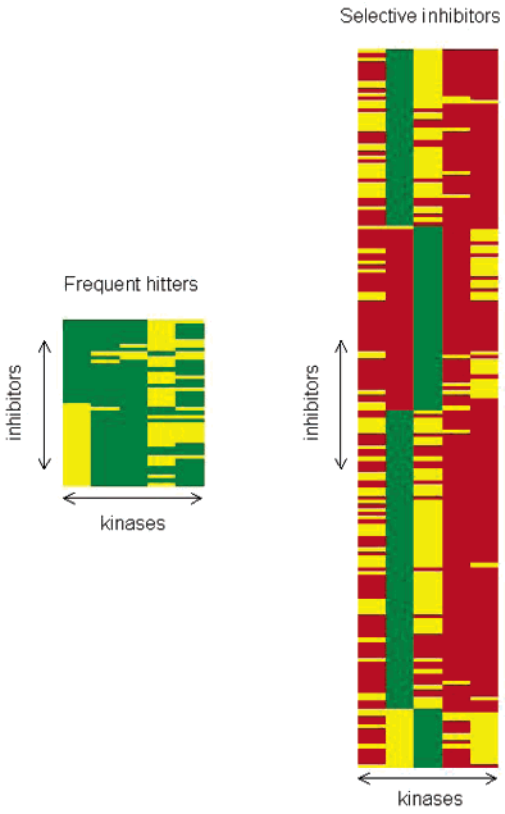


Figure 1. Kinase selectivity profiles for two datasets. Each row represents a compound, five columns represent five kinases (Src, PKA, Gsk3, Erk2, Cdk2). The color-coding scheme is as follows: K_I < 0.05 μM (green), 0.05 μM < K_I < 2 μM (yellow), and K_I > 2 μM (red). Selective inhibitor profile contains a single green element, while a frequent hitter turns on multiple green elements.

Table 1. Summary of the Pharmacophore Geometric Parameters[a]

distances (Å)		distances (Å)	
F1F2[1]	1.5−2.5	F3F4	2.5−3.1
F1F3	3.9−4.9	F1F5	5.3−8.1
F2F3	1.9−2.9	F2F5	5.8−9.2
F1F4	6.5−7.5	F3F5	7.2−10.0
F2F4	4.9−5.5	F4F5	8.2−11.0

[a] Pharmacophore features are as follows: F1 donor, F2 acceptor, F3 donor, F4 aromatic, and F5 acceptor (see Figure 2B).

We chose to utilize the information from our corporate collection of ligand-bound kinase structures to facilitate the search for a kinase frequent hitter pharmacophore. Structures for four of the 43 promiscuous inhibitors were available. They represented two different scaffolds. After aligning the structures in the common frame of reference and pharmacophore annotation using the polar-charge-hydrophobe (PCH) scheme as implemented in MOE,[25] we selected five apparent feature clusters as shown in Figure 2a. Positions of these clusters were used to define the resulting five-point pharmacophore for kinase frequent hitters. It contains a hydrogen bond acceptor F2 flanked by two hydrogen bond donors (F1 and F3), as well as an aromatic moiety F4 and acceptor F5 (Figure 2b). Intrafeature distances are shown in Table 1. Two examples of promiscuous kinase inhibitors are shown in Figure 2c. In the first ligand, F1−F3 hydrogen bonding features are provided by the aminopyrazole moiety, the pyrimidine spacer gives the

Table 2. Datasets and Classification Results Using the Kinase Frequent Hitter Pharmacophore

data set	no. in set	frameworks[a]	similarity/ ±std dev[b]	no. matching/ not matching[c]	% matching/ not matching
frequent hitters	43	23	0.63/±0.12	38/5	88/12
selective-50	209	64	0.69/±0.11	10/199	5/95
selective-10	130	39	0.71/±0.11	2/128	2/98
selective-2	61	17	0.73/±0.09	0/61	0/100
all	252	85	0.65/±0.12		

[a] Molecular frameworks were derived as described by Bemis and Murcko.[15] [b] Mean and standard deviation for internal pairwise Tanimoto similarity using Daylight fingerprints.[22] [c] Pharmacophore match represents a full match to the five-point pharmacophore query (Table 1).

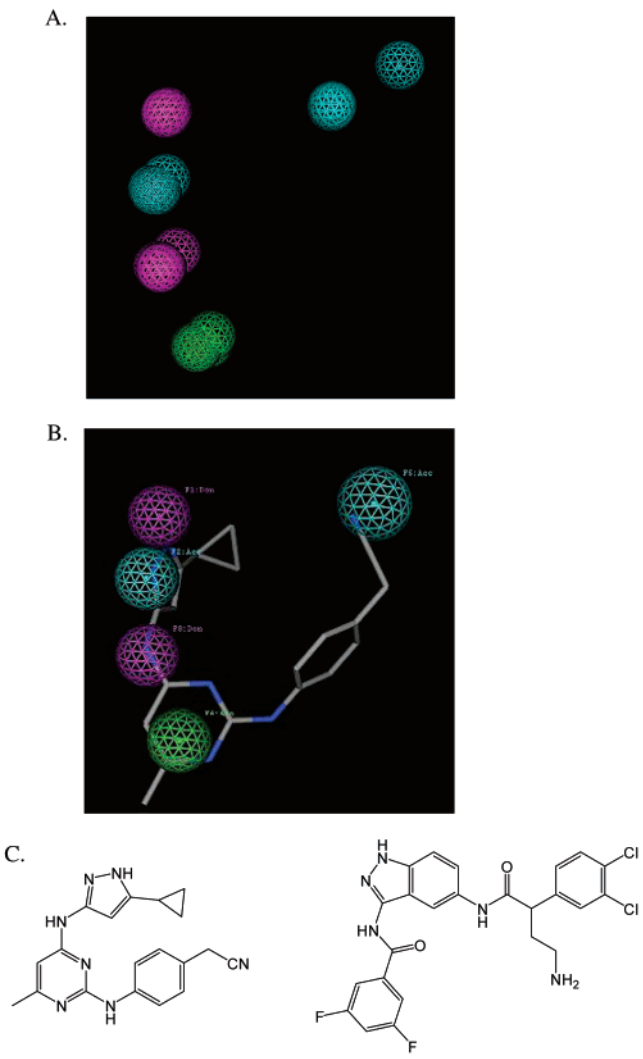


Figure 2. Derivation of a pharmacophore for kinase frequent hitters. A. Clusters of features derived from annotations of four X-ray structures of promiscuous kinase inhibitors. Colors are magenta (donor), cyan (acceptor), and green (aromatic). B. Resulting five-point pharmacophore used as a query for kinase frequent hitters. Pharmacophore features are as shown: F1 donor, F2 acceptor, F3 donor, F4 aromatic, and F5 acceptor. Sphere radii used for pharmacophore searching in MOE[25] were 1 Å for F1−F4, and 1.2 Å for F5. C. Examples of kinase frequent hitters from the Vertex corporate database.

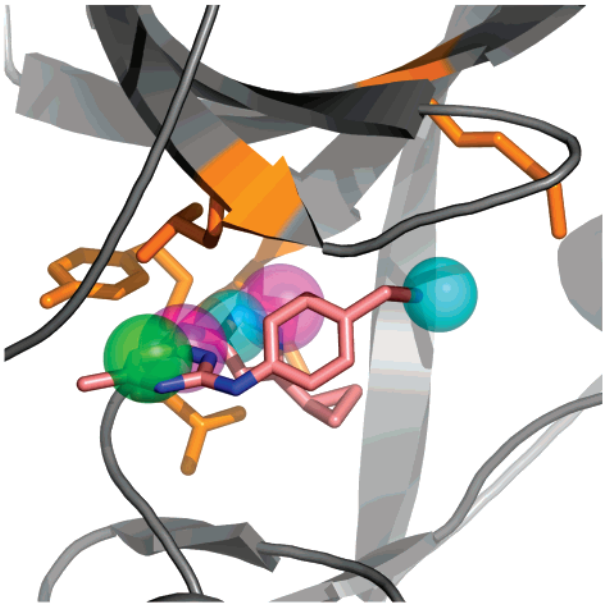
match with F4, and phenylacetonitrile makes the F5 acceptor interaction (Figure 2b). In the case of the second molecule, the aminoindazole, the benzamide, and the amide carbonyl of the substituted phenylacetamide serve as the frequent hitter interaction points.

The pharmacophore query search was performed using the pharmacophore matching routine in MOE. Search results for the datasets described above using the proposed frequent hitter pharmacophore (Figure 2b) are shown in Table 2. Of the 43 molecules in the frequent hitter set, 38 (88%) molecules produced a query match. When subjected to the same query search, the set of selective ligands (Selective-50) recorded only 10 matches (5%). To estimate the robustness of the approach, we chose subsets of the original selective ligand set for further validation. Selective-10 represents the more selective subset of Selective-50, as described above. Only two matches to the frequent hitter query were observed for this ligand subset (2%). Finally, Selective-2 subset containing the most selective ligands in our database was profiled, and revealed no matches to the pharmacophore query (Table 2). These results suggest the utility of the proposed five-point kinase frequent hitter pharmacophore query as a rapid filter for promiscuous kinase inhibitors.

Knowledge of the link between the annotated features and the structural information about contacts with specific conserved residues within the kinase ATP binding site enables one to pinpoint the specific interactions that act to predispose a kinase ligand to promiscuous binding behavior. The five identified pharmacophore features were placed in the ATP binding site of protein kinase A (PKA) (Figure 3a). PKA is one of the kinases with the most public structural information available, with over 20 structures deposited in the Protein Data Bank.[26] Key structural elements within the ATP site of PKA responsible for interaction with the features of the frequent hitter pharmacophore are: Leu 49, Lys 72, Glu 121, Tyr 122, and Val 123. The three residues, Glu 121, Tyr 122, and Val 123 form the central part of the hinge region between the two kinase domains, and the lack of conservation (Figure 3b) observed for them may be attributed to the fact that their side chains do not participate in the key interactions in the ATP site. Acceptor F2 forms a hydrogen bond to the backbone NH of Val 123, and the flanking donors F1 and F3 form hydrogen bonds to the backbone carbonyls of Glu 121 and Val 123, respectively. The aromatic feature F4 interacts with the hydrophobic Leu 49 side chain, which points down into the active site from the glycine-rich loop. Little variation is observed at that position for the five kinases (Figure 3b). Finally, acceptor functionality F5 forms a hydrogen bond to the basic nitrogen of the conserved catalytic Lys 72. While different combinations of these interactions can be seen in a variety of ligand-bound kinase structures, it appears that a five-point combination is needed for promiscuous inhibition across different kinase families.

In conclusion, we have identified a five-point pharmacophore for kinase frequent hitters and demonstrated its ability to discriminate between frequent hitters and

A.



B.

PKAα num.	49	57	70	72	104	121	122	123	124	173	183
PKAα	*L*	V	A	*K*	V	*E*	*Y*	*V*	P	L	T
Cdk2	*I*	V	A	*K*	V	*E*	*E*	*L*	H	L	A
Erk2	*I*	V	A	*K*	I	*D*	*L*	*M*	E	L	C
Gsk3β	*I*	V	A	*K*	V	*D*	*Y*	*V*	P	L	C
Src	*L*	V	A	*K*	V	*E*	*Y*	*M*	S	L	A

Figure 3. Kinase frequent hitter pharmacophore. A. The pharmacophore features were placed in the ATP binding site of protein kinase A (PKA). The crystal structure of PKA (1STC),[27] a serine/threonine kinase, is shown in ribbon representation. The ATP binding site was aligned with other kinase structures using heavy atoms of residues 116−123 in the hinge region. A portion of the C-terminus was deleted for illustrative clarity. Key structural elements within the kinase ATP binding site responsible for interaction with the features of the frequent hitter pharmacophore are colored orange: Leu 49, Lys 72, Glu 121, Tyr 122, and Val 123. Color representation of pharmacophore features is from Figure 2: magenta (hydrogen bond donor), cyan (hydrogen bond acceptor), and green (aromatic). A kinase frequent hitter is included as a reference. B. Sequence alignment of the residues lining the ATP binding site for five kinases. Residue numbering is based on PKA sequence. Conserved residues are shown in green. Residues interacting with the frequent hitter pharmacophore are italicized.

selective ligands. This pharmacophore can be used for rapid virtual screening of compound libraries for molecules with a potential for nonselective inhibition of kinases. We linked the feature information about kinase frequent hitters to the structural information about conserved residues within the kinase ATP binding site that appear to increase a ligand's propensity for promiscuous behavior. We expect that the ability to predict in silico the ligand's propensity for being a frequent hitter will enable medicinal chemists to make more informed decisions in the context of cross-kinase selectivity. It may indeed be possible to introduce sufficient additional components into a molecule that could render it selective despite the presence of features F1−5. A different approach involves excision of a moiety responsible for one of the interactions described by the promiscuous pharmacophore in lead compounds. In our hands, this approach has proven to be a successful strategy in a quest for selective inhibitors of individual kinases, a critical point in the realization of the chemogenomics concept.

References

(1) Cohen, P. Protein kinases—the major drug targets of the twenty-first century? *Nat. Rev. Drug Discovery* **2002**, *1*, 309−315.
(2) Rubin, G. M.; Yandell, M. D.; Wortman, J. R.; Gabor Miklos, G. L.; Nelson, C. R.; Hariharan, I. K.; Fortini, M. E.; Li, P. W.; Apweiler, R.; Fleischmann, W., et al. Comparative genomics of the eukaryotes. *Science* **2000**, *287*, 2204−2215.
(3) Manning, G.; Whyte, D. B.; Martinez, R.; Hunter, T.; Sudarsanam, S. The protein kinase complement of the human genome. *Science* **2002**, *298*, 1912−1934.
(4) Capdeville, R.; Buchdunger, E.; Zimmermann, J.; Matter, A. Glivec (STI571, imatinib), a rationally developed, targeted anticancer drug. *Nat. Rev. Drug Discovery* **2002**, *1*, 493−502.
(5) Muhsin, M.; Graham, J.; Kirkpatrick, P. Gefitinib. *Nat. Rev. Drug Discovery* **2003**, *2*, 515−516.
(6) Dancey, J.; Sausville, E. A. Issues and progress with protein kinase inhibitors for cancer treatment. *Nat. Rev. Drug Discovery* **2003**, *2*, 296−313.
(7) McGovern, S. L.; Shoichet, B. K. Kinase inhibitors: not just for kinases anymore. *J. Med. Chem.* **2003**, *46*, 1478−1483.
(8) Godl, K.; Wissing, J.; Kurtenbach, A.; Habenberger, P.; Blencke, S.; Gutbrod, H.; Salassidis, K.; Stein-Gerlach, M.; Missio, A.; Cotten, M.; Daub, H. An efficient proteomics method to identify the cellular targets of protein kinase inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 15434−15439.
(9) Fischer, P. M. CDK versus GSK-3 inhibition: a purple haze no longer? *Chem. Biol.* **2003**, *10*, 1144−1146.
(10) Diller, D. J.; Li, R. Kinases, homology models, and high throughput docking. *J. Med. Chem.* **2003**, *46*, 4638−4647.
(11) Knockaert, M.; Gray, N.; Damiens, E.; Chang, Y. T.; Grellier, P.; Grant, K.; Fergusson, D.; Mottram, J.; Soete, M.; Dubremetz, J. F.; Le Roch, K.; Doerig, C.; Schultz, P.; Meijer, L. Intracellular targets of cyclin-dependent kinase inhibitors: identification by affinity chromatography using immobilised inhibitors. *Chem. Biol.* **2000**, *7*, 411−422.
(12) Knockaert, M.; Wieking, K.; Schmitt, S.; Leost, M.; Grant, K. M.; Mottram, J. C.; Kunick, C.; Meijer, L. Intracellular Targets of Paullones. Identification following affinity purification on immobilized inhibitor. *J. Biol. Chem.* **2002**, *277*, 25493−25501.
(13) Breitenlechner, C.; Gassel, M.; Hidaka, H.; Kinzel, V.; Huber, R.; Engh, R. A.; Bossemeyer, D. Protein kinase A in complex with Rho-kinase inhibitors Y-27632, Fasudil, and H-1152P: structural basis of selectivity. *Structure (Camb)* **2003**, *11*, 1595−1607.
(14) Evans, B. E.; Rittle, K. E.; Bock, M. G.; DiPardo, R. M.; Freidinger, R. M.; Whitter, W. L.; Lundell, G. F.; Veber, D. F.; Anderson, P. S.; Chang, R. S.; et al. Methods for drug discovery: development of potent, selective, orally effective cholecystokinin antagonists. *J. Med. Chem.* **1988**, *31*, 2235−2246.
(15) Bemis, G. W.; Murcko, M. A. The properties of known drugs. 1. Molecular frameworks. *J. Med. Chem.* **1996**, *39*, 2887−2893.
(16) Hajduk, P. J.; Bures, M.; Praestgaard, J.; Fesik, S. W. Privileged molecules for protein binding identified from NMR-based screening. *J. Med. Chem.* **2000**, *43*, 3443−3447.
(17) Mason, J. S.; Morize, I.; Menard, P. R.; Cheney, D. L.; Hulme, C.; Labaudiniere, R. F. New 4-point pharmacophore method for molecular similarity and diversity applications: overview of the method and applications, including a novel approach to the design of combinatorial libraries containing privileged substructures. *J. Med. Chem.* **1999**, *42*, 3251−3264.
(18) Klabunde, T.; Hessler, G. Drug design strategies for targeting G-protein-coupled receptors. *Chembiochem* **2002**, *3*, 928−944.
(19) McGovern, S. L.; Caselli, E.; Grigorieff, N.; Shoichet, B. K. A common mechanism underlying promiscuous inhibitors from virtual and high-throughput screening. *J. Med. Chem.* **2002**, *45*, 1712−1722.
(20) http://www.kinase.com.
(21) http://www.ncbi.nlm.nih.gov/PubMed/PubMed.
(22) Daylight Chemical Information Systems, Inc., 2004.
(23) Corina; Molecular Networks GmbH: Erlangen, Germany, 2001.
(24) Omega; OpenEye Scientific Software, Inc.: Santa Fe, NM, 2001.
(25) Molecular Operating Environment; Chemical Computing Group Inc.: Montreal, Canada, 2002.
(26) Protein Data Bank http://www.rcsb.org/pdb.
(27) Prade, L.; Engh, R. A.; Girod, A.; Kinzel, V.; Huber, R.; Bossemeyer, D. Staurosporine-induced conformational changes of cAMP-dependent protein kinase catalytic subunit explain inhibitory potential. *Structure* **1997**, *5*, 1627−1637.

JM049793G