JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

Virtual Target Screening: Validation Using Kinase Inhibitors

Daniel N. Santiago,[‡] Yuri Pevzner,[‡] Ashley A. Durand,[†] MinhPhuong Tran,[‡] Rachel R. Scheerer,[‡] Kenyon Daniel,[†] Shen-Shu Sung,[§] H. Lee Woodcock,[‡,⊥] Wayne C. Guida,*[,†,‡,⊥] and Wesley H. Brooks*[,†,‡]

[†]HTS & Chemistry Core, H. Lee Moffitt Cancer Institute & Research Institute, 12902 Magnolia Drive, Drug Discovery-SRB3, Tampa, Florida 33612, United States

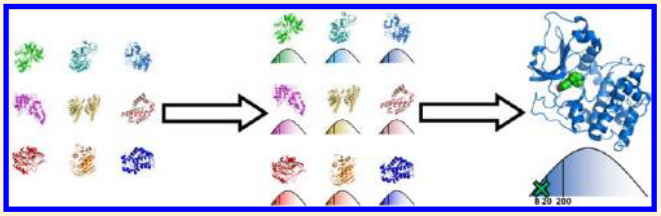
[‡]Department of Chemistry, University of South Florida, Tampa, Florida 33620, United States

[§]Department of Pharmacology, Milton S. Hershey Medical Cancer Institute, Pennsylvania State University, 500 University Drive, MC H072, Hershey, Pennsylvania 17033, United States

[⊥]Center for Molecular Diversity in Drug Design, Discovery and Delivery, University of South Florida, 4202 East Fowler Avenue, CHE 205, Tampa, Florida 33620, United States

Supporting Information

ABSTRACT: Computational methods involving virtual screening could potentially be employed to discover new biomolecular targets for an individual molecule of interest (MOI). However, existing scoring functions may not accurately differentiate proteins to which the MOI binds from a larger set of macromolecules in a protein structural database. An MOI will most likely have varying degrees of predicted binding affinities to many protein targets. However, correctly interpreting a docking score as a hit for the MOI docked to any individual protein can be problematic. In our method, which we term "Virtual Target Screening (VTS)", a set of small drug-like molecules are docked against each structure in the protein library to produce benchmark statistics. This calibration provides a reference for each protein so that hits can be identified for an MOI. VTS can then be used as tool for: drug repositioning (repurposing), specificity and toxicity testing, identifying potential metabolites, probing protein structures for allosteric sites, and testing focused libraries (collection of MOIs with similar chemotypes) for selectivity. To validate our VTS method, twenty kinase inhibitors were docked to a collection of calibrated protein structures. Here, we report our results where VTS predicted protein kinases as hits in preference to other proteins in our database. Concurrently, a graphical interface for VTS was developed.



INTRODUCTION

Drug discovery and development focuses initially on finding a lead candidate. The intent is to find a molecule capable of modifying (usually inhibiting) the activity of a protein to alter the course of a disease. Early in the process, it should be known whether or not the lead candidate, or a focused library built around a promising scaffold, is specific toward the target of interest and if there are potentially detrimental off-target effects. Alternatively, the scope of an existing drug can be expanded if new biomolecular targets can be identified, thus reducing the cost and time of developing new therapies. Fortunately, experimental methods are relatively well-developed to address drug specificity and promiscuity[1] (and online databases[2−4] and data mining techniques[5−7] can potentially aid in these efforts). On the other hand, testing a lead candidate or focused library against other proteins experimentally (referred to as specificity testing or counter-screening) is a tedious process and is often limited by cost, time, availability of proteins, and appropriate assays. In a method that has been referred to as virtual target screening (VTS), protein structures are screened by ranking a small molecule's docking scores to calibration docking statistics. These ranked dockings theoretically can reveal proteins that have significant interactions with a given small molecule. Improvements in virtual screening (VS) applications[8,9] and the availability of an increasing number (78 020 structures as of December 20, 2011) of solved protein structures from the Protein Data Bank[10] (PDB) help to make VTS a reality.

In VTS (also referred to as inverse docking or virtual counter-screening), a molecule of interest (MOI) can be docked rapidly into each entry of a protein structure library. The key to an effective VTS system is correctly interpreting the relative importance of the individual ligand−protein docking scores to determine which proteins are of particular significance among all screened proteins. Identification of protein "hits" in previous inverse docking studies have included: direct ranking of scores, modification of scores (such as weighting against possible promiscuity and nonspecific binding), "fingerprinting" (comparisons between the molecule of interest and known inhibitors), and incorporation of larger binding data sets to improve scoring

Received: February 7, 2012
Published: July 2, 2012

functions for ligand docking.[11−19] Already, virtual counter-screening techniques have been developed to answer the need for drug positioning,[3,20−29] toxicology,[30−33] and selectivity of focused librares.[34]

Here, the development and validation of a new VTS system is described that employs a unique approach involving 20 known small molecule kinase inhibitors and more than 1400 protein structures and a structure-based counter-screening approach dependent on calibration with a diverse set of molecules. This benchmarking against our protein database yields a fast, yet robust, procedure for determining targets for a given molecule of interest (MOI). The individual protein structures have been calibrated against a drug-like set of compounds, the National Cancer Institute's (NCI) Diversity Set (more information below). An extensive case study of known kinase inhibitors is presented as validation of our methodology where protein hits identified by VTS within our calibrated protein library are compared to hundreds of published experimental data points. Additionally, a user-friendly interface has been developed to facilitate the VTS workflow.

MATERIALS AND METHODS

Hardware. Molecular modeling and VTS studies were performed using a Dell Precision 490 workstation running on Fedora 8 Linux with dual Xeon 3.06 GHz processors, 4 GB RAM, and a 250 GB hard drive.

Software. Schrödinger's Maestro 8.0[35] was used as the primary graphical user interface for molecule structure preparation. LigPrep 1.6[36] was used to convert the NCI Diversity Set I[37] from the provided 3D models in SDF file format to refined 3D models in Maestro file format. LigPrep 2.2[38] was later used for refinement of small MOIs that were to be screened in VTS. Maestro and MacroModel[39] 9.5 were used in preparation of enzyme coordinates for docking studies. Schrödinger's Grid-based Ligand Interaction Docking with Energetics (GLIDE) 5.0[40−42] was used for the generation of grid files and automated in silico docking (virtual screening). Perl scripts were created to automate protein structure calibration via command-line execution of Schrödinger applications. PyMol[43] from DeLano Scientific was used for graphical presentation of the results.

Small Molecule Calibration Structures. The National Cancer Institute (NCI) Diversity Set I, consisting of 1990 3D structures, was used as our small molecule calibration set. It is a representative subset of the entire NCI/NIH Development Therapeutics Program chemical collection of almost 140 000 compounds. As previously mentioned, ligand refinement was done using Schrödinger's LigPrep, which increased the number of structures to 2392. The original 1990 structures decreased, though, to 1875 due to lack of force field parameters (e.g., molecules containing arsenic were omitted). The additional structures represented different tautomers, ring conformations, and protonation states of the 1875 compounds. Energy minimization of the NCI Diversity Set I employing the MMFF force field[44−49] was also performed using LigPrep.

Protein Structures. Our collection of protein structures, currently at 1451 entries (Figures 1 and 2), was prepared from the PDB. We loosely applied a set of guidelines in selecting proteins to add to our library. Our main aim was to have a broad representation so that new target proteins could be probed. The guidelines used in selecting our initial set of proteins were the following: (1) human proteins preferred; (2) wild-type structures only; (3) full-length or near full length sequences;

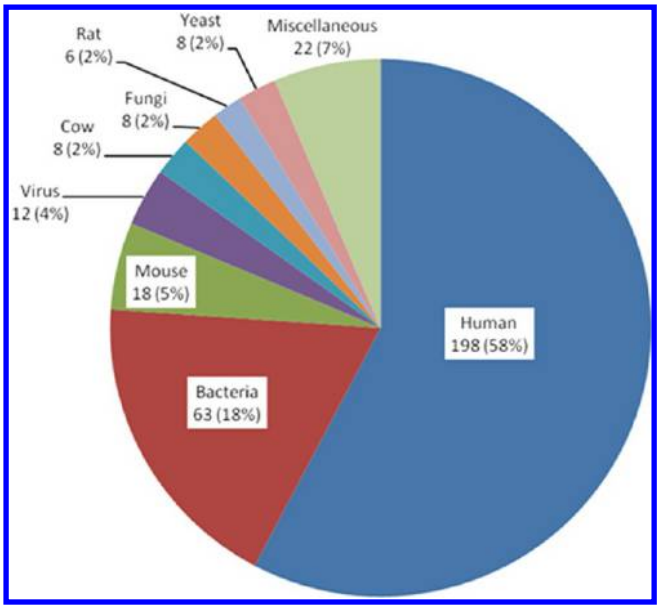


Figure 1. Organism composition shown for VTS protein structure library of 1451 total entries. "Miscellaneous" consists of organism types each totaling less than 2% of the library: dog, ray, chicken, rabbit, plant, protozoa, snake, and pig.

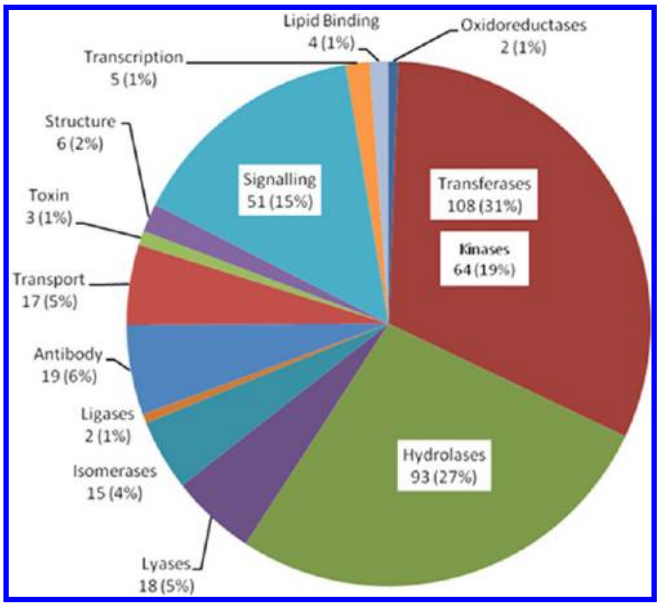


Figure 2. Protein type composition shown for VTS protein structure library of 1451 total entries. Kinases are about 2/3 of transferases, 19% overall.

(4) X-ray structures rather than NMR or homology (theoretical) models; (5) resolution better than 3 Å; and (6) a ligand noncovalently bound in the protein's active site or binding interface. Proteins with more than one molecule in the asymmetric unit (i.e., more than one copy of the protein in the crystal structure) were inspected, and those with lower B-factors were chosen while the other subunits were deleted. Of the 1451 protein structures, 343 (24%) are unique, so the remaining structures are redundant in sequence. Human protein structures account for 59% of the VTS library.

Ligand Structures. Fabian et al. collected K_d information for 119 kinase and kinase-like targets against 20 kinase inhibitors[50] (Figure 3). We used these inhibitors as our test set for

dx.doi.org/10.1021/ci300073m | J. Chem. Inf. Model. 2012, 52, 2192−2203

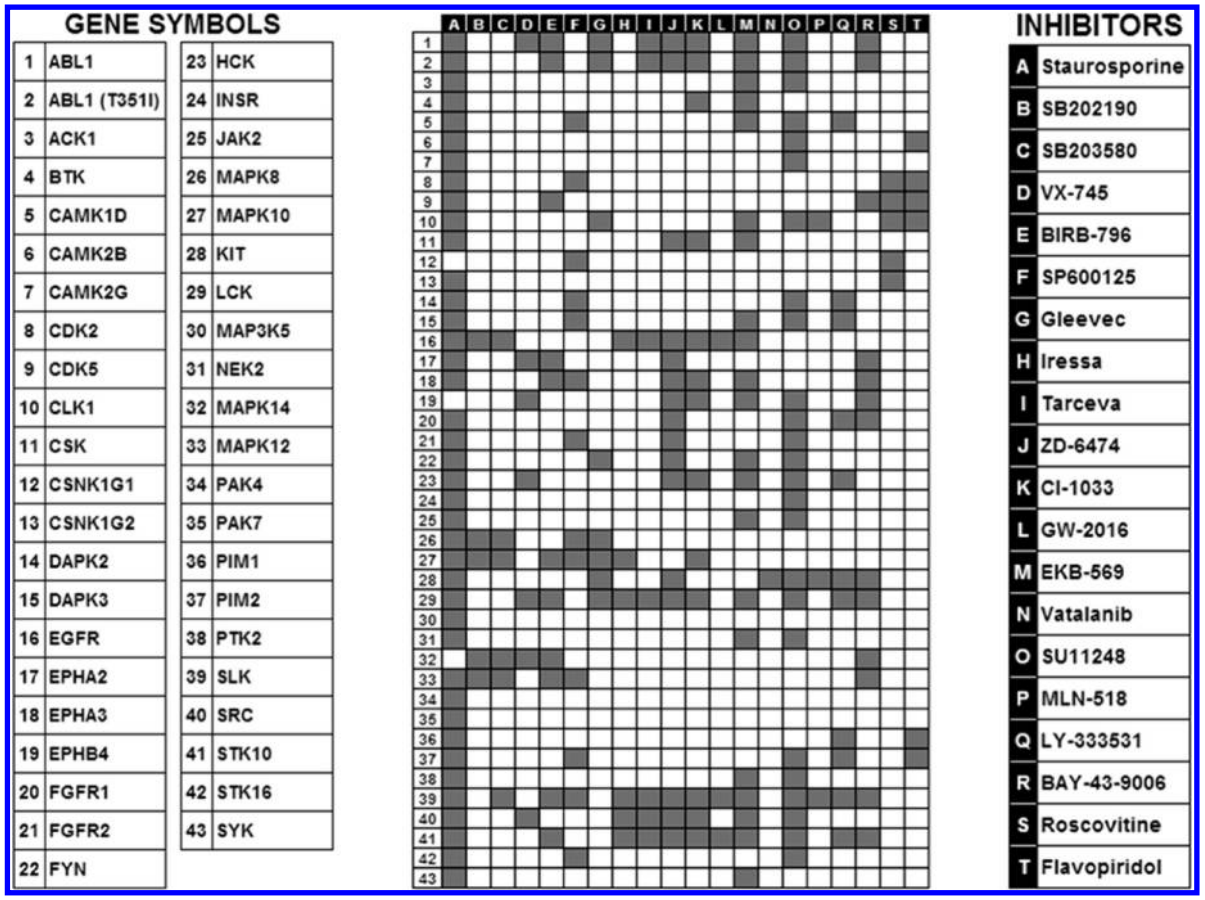


Figure 3. Matrix of reported activities on kinase inhibitor assays at 10 μM. Of the 119 kinases reported by Fabian et al., 43 were available from the PDB to include in the VTS library. 220 (26%, gray squares) of the 860 inhibitor−kinase combinations have target binding affinities (K_d < 10 μM). This cutoff is used as our metric to compare VTS hits with reported data.

the VTS system. Maestro was used to create each ligand structure, and LigPrep was used to create tautomers and produce different ionization states within a pH range of 7 ± 2. Energy minimization of ligand structures was effected using LigPrep.

In Silico Docking. For compatibility with the Optimized Potentials for Liquid Simulations (OPLS) all-atom force field[51,52] used by GLIDE, selenium atoms were changed to sulfur atoms. (Selenium parameters have since been added to the OPLS force field; however, the selenium-to-sulfur substitution was used for those proteins subjected to X-ray crystallography studies using the method of multiwavelength anomalous diffraction.[53]) All explicit water molecules were removed as well as repeated structures with high B-factors. The GLIDE protein preparation module, available from Schrodinger,[54] was employed to add explicit hydrogen atoms, define appropriate charges from the OPLS force field for protein atoms, and perform restrained minimization of the protein structures. Small-molecule structures were docked with GLIDE at standard precision (SP) into the protein target to approximate absolute binding free energies using the GLIDE Score (Gscore) scoring function.[40,55] Default SP settings include 400 poses kept for energy minimization of the initial 5000 poses within a scoring window of 100.00 kcal/mol, energy minimization performed with a distance-dependent dielectric of 2.0 with a maximum number of 100 steps, and a scaling factor of 0.8 for van der Waals radii of ligand atoms with an absolute value of partial charge less than 0.15. These Gscores are calculated and used to rank the structures relative to each other in a docking run. Default grid files were centered on the available ligand in each protein structure. Due to software limitations, GLIDE is unable to dock ligands (such as polypeptides) with more than 300 atoms and/or 50 rotatable bonds. If there was a small molecule originally in the protein structure, it was docked (referred to as "self-docking") and the resulting Gscore was recorded.

Calibrating. Docking the NCI Diversity Set I against the VTS library provided a reference for each protein structure. Statistics on the calibration dockings are eventually compared to the MOI docking score in the VTS. How well an MOI ranks against a protein's calibration reference is employed to determine if that protein is a hit. For each protein structure, a Perl script (written in-house) was employed to execute Schrödinger command-line utilities to calculate top-200, top-20, and Boltzmann weighted averages of the Gscores from the calibration dockings. These averages served as criteria to determine whether a particular protein is a hit. The formula for the Boltzmann weighted average, B_j, is defined with $\Delta G_{i,j}$ as the approximate binding free energy (Glide docking score or Gscore, kcal/mol) for the ith structure in the NCI Diversity Set I docked into the jth protein structure in the VTS library.

$$B_j = \sum_i \left(\Delta G_{i,j} \frac{e^{-\Delta G_{i,j}/kT}}{\sum_i e^{-\Delta G_{i,j}/kT}} \right)$$

The number of successful dockings (out of 1875 unique structures) is also logged. These statistics, along with any self-dock Gscore, are tabulated for each protein in the VTS library.

dx.doi.org/10.1021/ci300073m | J. Chem. Inf. Model. 2012, 52, 2192−2203

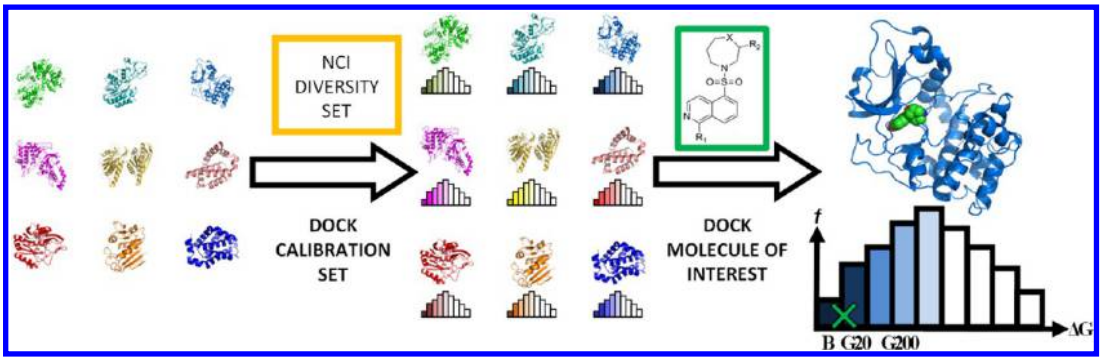


Figure 4. Scheme outlining the basic steps of the VTS system. Each protein structure is first calibrated (left arrow) by docking the NCI Diversity Set I and calculating statistics. The Boltzmann average, top-20 average, and top-200 average criteria (represented on right with "B", "G20", and "G200", respectively, and shown on a hypothetical plot of score frequency vs ΔG). These "average" scores determine protein hits for a molecule of interest (MOI). The "×" represents a hypothetical MOI Gscore that scored better than the top-200 and top-20 averages but not the Boltzmann average for the protein entry.

Virtual Target Screening. LigPrep was used to prepare all MOIs. A Perl script uses the statistics from the calibration dockings as a reference after the MOI is docked into each structure in the protein library. Thus, when an MOI is docked, its Gscore is compared to the calibration averages. If its Gscore is better (i.e., more negative than the top-200 average, top-20 average, or Boltzmann average), the protein to which it was docked is determined to be a hit. (See Figure 4.)

VTS Web-Based Interface. To facilitate the application of the VTS protocol, we have developed a web-based interface that provides a user-friendly, quick, and automated tool for docking MOIs into collections of user-defined proteins. The framework for the online VTS interface is based on the open source CHARMM interface and graphics[56] (CHARMMing) package, which was originally designed to provide an easy to use interface to the Chemistry at HARvard Macromolecular Mechanics[57] (CHARMM) modeling package and force field.[58] The underlying framework upon which both the VTS interface and CHARMMing are built is based on the Django engine—a high-level Python Web framework.

Derived from CHARMMing, the graphical virtual target screening (gVTS) system includes tools necessary to set up and initiate VTS experiments. Functionality currently implemented includes the following:

- Maintain a library of protein grids for docking of small molecules.
  - User-prepared grids, based on the proteins of interest can be uploaded and stored in the internal database.
  - Ability to create custom grid sets that can represent structures specific to a given VTS experiment.
- Maintain database of MOIs.
  - User can submit MOIs either by uploading the Cartesian coordinates or by drawing a molecule via a 2D chemical drawing interface, JChemPaint (jchempaint.sourceforge.net), which is included in gVTS (see Figure 5).
  - All submitted MOIs are atom-typed and then energy minimized with MacroModel.
- Initiate VTS runs and analyze results
  - VTS jobs can be set up with any number of MOIs against either the entire library of proteins or a custom created subset.

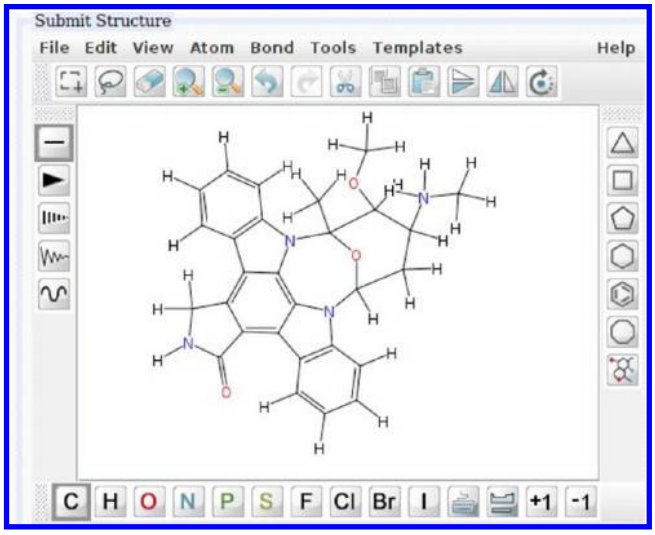


Figure 5. Illustration of the JChemPaint applet for the input of staurosporine into gVTS.

  - Job runtime estimation algorithm predicts an approximate execution time based on the number of MOIs, rotatable bonds per MOI, and number of screened proteins.
  - Job scheduling and queuing system provided by the CHARMMing interface allows for submission of multiple jobs and can be interfaced to popular queuing systems such as Torque, PBS, and sun grid engine (with slight modifications).
  - Complete, up to the second, information on any job currently running or run in the past is available and includes but is not limited to the status of the job, information about any resulting hits, structures being screened/hit, log, and output files. In addition, the user is able to visualize the docking pose of any MOI in a protein hit (see Figure 6).
- User authentication system
  - To ensure privacy of the data, the Django/CHARMMing based user authentication system in combination with database identifiers protects each user's information such as MOIs, protein structures, jobs, etc. and allows access only by authorized persons.

The interface has been developed using the Python 2.6 programming language and the Django 0.96 object framework;

dx.doi.org/10.1021/ci300073m | J. Chem. Inf. Model. 2012, 52, 2192−2203

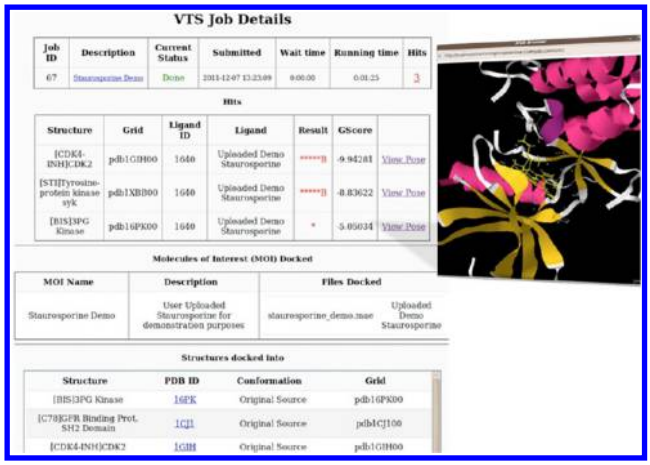


Figure 6. Illustration of the results panel and 3D docking pose of staurosporine in a kinase.

however, future versions of CHARMMing (and gVTS) will fully support the latest language/framework. A MySQL 5.1.37 database is used to maintain system information and user generated data. Perl scripts provide the interface to the Schrödinger software suite. The JChemPaint Java applet is used to draw molecular structures.

RESULTS AND DISCUSSION

Kinase Data Set and Protein Hit Determination. Twenty small molecule kinase inhibitors were chosen as a test set for our VTS procedure. Fabian et al. had previously collected data for these 20 molecules and 119 protein kinases[46] included in a complete table (their Supplementary Table 4) of binding constants (K_d values if K_d < 10 μM). For this study, our metric compared this table of binding data with the VTS protein hit data on the 20 kinase inhibitors. We gauged VTS accuracy by its ability to predict protein hits by reference to protein−inhibitor pairs that have K_d < 10 μM. Of the 119 kinases, 43 structures were available for download from the PDB allowing for 860 data points to be compared (Figure 7) of which 220 (26%) protein−inhibitor combinations had K_d < 10 μM.

Among the protein hit criteria for determining a VTS hit, the top-20 average was optimal in this study for predicting specific kinase hits as reported, slightly better statistically than the Boltzmann average criterion. "Optimal" is defined as having the maximum number of data points matching between VTS and experimental data. In the tables at the top of Figure 7, the top-20 average afforded the best correlation between VTS hits and experimental K_d < 10 μM and VTS nonhits vs experimental $K_d \geq 10$ μM. The top-200 average yielded the highest hit rate (49%) for simply determining protein kinase−inhibitor interactions with K_d < 10 μM, but it also yielded the largest number of false positives and false negatives (37% vs 28% or 30%).

It should be noted that the activity data is based on a primary kinase screening assay run at 10 μM.[50] VTS sensitivity (ability to predict a protein−inhibitor combination to be a hit considering only reported K_d < 10 μM) is relatively low: 22−49%. (Here sensitivity is defined, from the VTS hits, as the ratio of the upper-left value to the 220 values with K_d < 10 μM.)

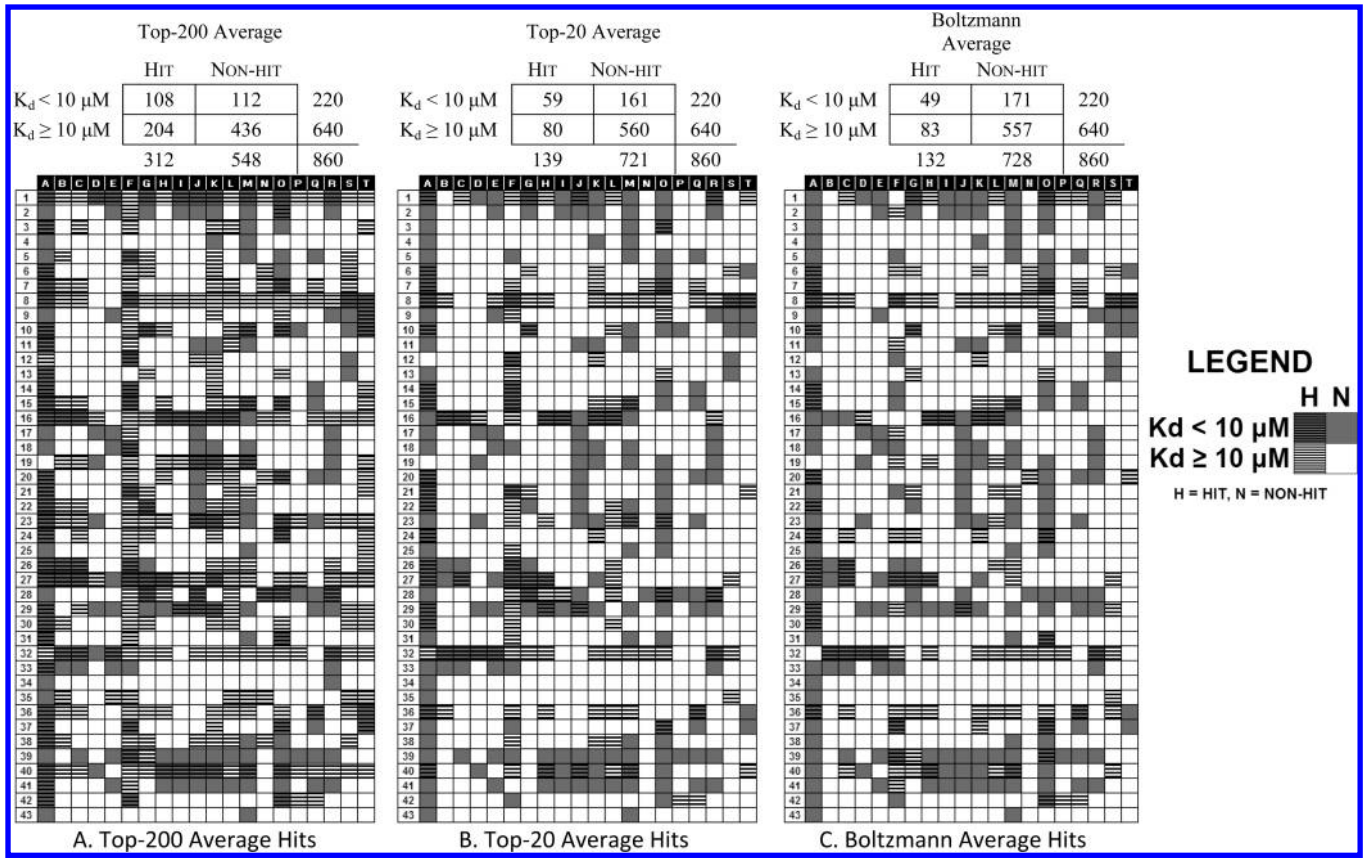


Figure 7. Tallies (top) and matrices (bottom) for VTS protein hit data and K_d values reported by Fabian et al.[50] for 860 combinations of 43 available protein kinase structures and 20 kinase inhibitors. Tallies include totals in the far right and bottom values. Numbers and letters correspond with protein kinases and kinase inhibitors, respectively, as in Figure 3.

dx.doi.org/10.1021/ci300073m | J. Chem. Inf. Model. 2012, 52, 2192−2203

However, the overall accuracies of the VTS system in identifying inhibitor−kinase combinations with K_d < 10 μM are 64%, 72%, and 71% for the top-200 hits, top-20 hits, and Boltzmann hits, respectively. (Overall accuracy is calculated from VTS hits and nonhits, upper-left and lower-right values, compared to all 860 data points.)

Results for Approved Drugs. For assessing the approved drugs in our test set, we assumed that an ideal profile would consist of a relatively low number of overall protein hits while retaining a high percentage of kinase hits. It is worth noting that staurosporine (while not an approved drug), a known pan-kinase inhibitor,[50] yielded the top percentage of kinase hits among the 20 small molecules in all 3 protein hit criteria. (See Table S7 and Figure S8 in the Supporting Information for kinase hit data for all 20 kinase inhibitors.) Drugs included in the Fabian set of kinase inhibitors tested are Nexavar (BAY-43-9006), Gleevec, Tykerb (GW-2016), Iressa, Sutent (SU112448), Tarceva, and Zactima (ZD-6474). The top-20 average yielded the highest percentage of kinase hits among the approved drugs in our study as summarized in Tables 1 and 2. For the purpose of this kinase inhibitor study, an enrichment factor (EF) was calculated considering any protein kinase in our database to be a hit. Since 19% of the VTS protein library is a protein kinase, our enrichment factor is defined as the ratio of kinase hit percentage to 19. EF values for approved drug top-20 average hits (EF-20) are included in Table 1. (EF values for all 20 kinase inhibitors and protein hit criteria are in Supporting Information Table S7.) While showing a low number of protein hits, the VTS system was able to identify these small molecules as having significant specificity toward kinase binding with high percentages of kinase hits. Zactima's EF-20 value is lowest, 1.7, with only 4 out of 12 unique protein hits being kinases. The other protein hits include a Subtilisin, Coagulation Factors X and XA, Thrombin, Plasminogen Activator Inhibitor Type 1, Thyroid Hormone Receptor α1, and Histamine Methyltransferase. The first is a bacterial protease from *B. subtilis*; the next 4 are related to blood clotting. Primary targets (Table 2) for each of the seven drug molecules were found (if present in the VTS library) as top-20 average protein hits. Interestingly, Zactima (Caprelsa) prescribing information includes monitoring thyroid-stimulating hormone for risk of hypothyroidism and warnings of

Table 1. VTS top-20 average protein kinase hit percentages for approved drugs[a]

		protein kinase hit percentage			
		top-200	top-20	Boltzmann	EF-20
approved drugs	Nexavar (BAY-43-9006)	38	**45**	13	2.4
	Gleevec (Imatinib)	57	**77**	58	4.0
	Tykerb (GW-2016)	29	**41**	35	2.2
	Iressa (Gefitinib)	48	**55**	50	2.9
	Sutent (SU11248)	60	**67**	47	3.5
	Tarceva (Erlotinib)	50	**67**	0	3.5
	Zactima (ZD-6474)	28	**33**	17	1.7

[a]EF-20 values are calculated by dividing kinase hit percentages from the top-20 criteria by the percentage of kinases in the VTS protein library: 19%.

Table 2. VTS Top-20 Average Protein Hit Data Summary for Approved Drugs in Kinase Inhibitor Study

approved drug	primary target	primary target hit?	kinase hits[a] (% of total)	total hits[a]
Nexavar (BAY-43-9006)	RAF1	N/A[b]	5 (45%)	11
Gleevec	ABL, KIT, PDGFR	yes, yes, N/A	10 (77%)	13
Tykerb (GW-2016)	EGFR, ERBB2, ERBB4	yes, N/A, N/A	20 (41%)	49
Iressa	EGFR	yes	12 (55%)	22
Sutent (SU11248)	VEGFR2, PDGFR, FLT3, KIT	N/A, N/A, N/A, yes	14 (67%)	21
Tarceva	EGFR	yes	2 (67%)	3
Zactima (ZD-6474)	VEGFR2, EGFR	N/A, yes	4 (33%)	12

[a]A high percentage of kinase hits with a low number of total hits demonstrates the VTS system's robust ability to predict a small molecule to be a potential lead candidate. [b]N/A = structure not in VTS protein library.

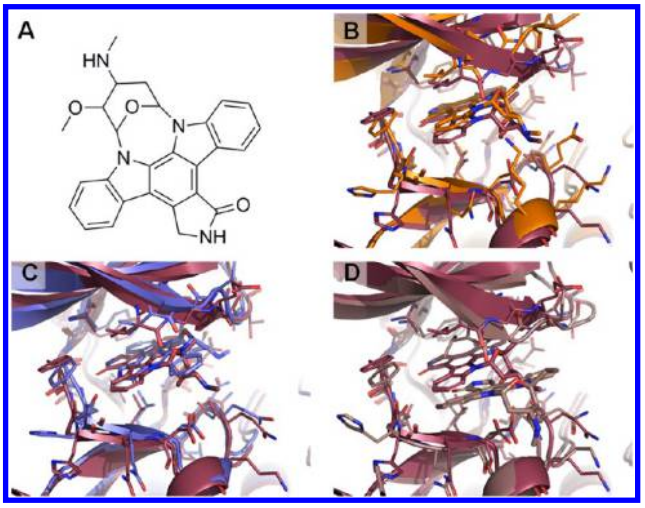


Figure 8. (A) Line structure of staurosporine and select VTS CDK2 docking poses. Protein kinases and docked poses are compared to PDB 1AQ1 (maroon). Examples include a top-20 hit (B, PDB 2BPM, orange), a top-200 hit (C, PDB 3DDP, blue), and a nonhit (D, PDB 1PF8, brown). *Z*-scores and ligand rmsd values are, respectively, −3.02 and 0.71 Å, −2.08 and 3.72 Å, and 1.22 and 6.04 Å. Staurosporine and residues within 4 Å are in stick representation. Also, rmsd values of the aligned kinase structures to PDB 1AQ1 are 1.70 (2BPM), 1.62 (3DDP), and 1.23 (1PF8). Subtle differences in the active site dramatically affected docking poses.

Table 3. CDK2 Hits for Each of the 20 Kinase Inhibitors in the Test Set

	KINASE INHIBITORS[a]																			
HIT TYPES	bay	birb	ci	ekb	fla	gle	gw	ire	ly	mln	ros	sb202	sb203	sp	sta	su	tar	vat	vx	zd
Top-200	22	27	55	18	66	30	33	9	6	2	87	33	19	92	47	45	2	10	0	3
Top-20	3	4	18	3	5	19	5	0	2	0	51	4	0	67	33	27	0	1	0	0
Boltzmann	0	0	6	4	13	8	8	1	2	0	36	1	3	38	21	17	0	1	0	1

[a]Approved drugs (gray) and compounds with the highest number of CDK2 hits (black) are shaded.

dx.doi.org/10.1021/ci300073m | J. Chem. Inf. Model. 2012, 52, 2192−2203

Table 4. Average Ligand rmsd Values for 546 Cross-dockings and 11 Self-dockings[b]

		TOTAL	Top-20 Hits		Top-200 Hits		Average Hits		Non-Hits		R	
Kinase	Inhibitor	No.	No.	Average RMSD	No.	Average RMSD	No.	Average RMSD	No.	Average RMSD	Z vs. L	K_d (μM)
ABL1	Gleevec	28	7	3.84 ± 4.67	0	--- ± ---	21	13.20 ± 1.51	0	--- ± ---	0.97	0.0022
CDK2	Roscovitine	196[a]	76	3.36 ± 1.23	61	4.44 ± 2.11	59	6.25 ± 1.95	0	--- ± ---	0.57	2.9
CDK2	Staurosporine	101	9	1.80 ± 1.10	19	4.54 ± 1.65	21	4.69 ± 1.46	52	8.26 ± 1.91	0.79	0.0081
CDK5	Roscovitine	1	1	2.97 ± ---	0	--- ± ---	0	--- ± ---	0	--- ± ---		2
CSK	Staurosporine	2	2	1.14 ± 0.64	0	--- ± ---	0	--- ± ---	0	--- ± ---	-1.00	0.44
EGFR	Iressa	6	3	3.43 ± 0.27	3	1.95 ± 0.45	0	--- ± ---	0	--- ± ---	-0.92	0.0018
EGFR	Tarceva	2	0	--- ± ---	0	--- ± ---	2	6.36 ± 3.04	0	--- ± ---	-1.00	0.0014
FYN	Staurosporine	1	1	0.47 ± ---	0	--- ± ---	0	--- ± ---	0	--- ± ---		0.051
JNK1	SP600125	1	1	0.35 ± ---	0	--- ± ---	0	--- ± ---	0	--- ± ---		0.1
KIT	Gleevec	1	1	3.82 ± ---	0	--- ± ---	0	--- ± ---	0	--- ± ---		0.83
KIT	Sutent	2	2	9.55 ± 3.32	0	--- ± ---	0	--- ± ---	0	--- ± ---		0.00071
LCK	Gleevec	4	1	12.21 ± ---	3	12.30 ± 0.68	0	--- ± ---	0	--- ± ---	-0.39	0.062
LCK	Staurosporine	4	4	8.86 ± 0.26	0	--- ± ---	0	--- ± ---	0	--- ± ---	0.73	0.02
P38A	Nexavar	56	20	5.77 ± 4.03	12	10.46 ± 4.27	20	10.21 ± 3.85	4	8.16 ± 0.50	0.40	0.26
P38A	BIRB796	28	9	3.01 ± 2.70	4	9.85 ± 1.62	12	10.76 ± 3.10	3	11.42 ± 3.95	0.78	0.00024
P38A	Gleevec	28	6	4.39 ± 0.37	4	10.30 ± 4.00	17	11.22 ± 1.88	1	13.35 ± ---	0.74	>10
P38A	SB203580	56	18	9.81 ± 6.07	18	5.75 ± 3.14	20	9.62 ± 7.41	0	--- ± ---	0.08	0.017
PIM1	LY333531	12	1	0.88 ± ---	3	3.31 ± 3.50	7	4.32 ± 1.89	1	6.48 ± ---	0.58	0.055
PIM1	Staurosporine	12	1	1.01 ± ---	3	4.05 ± 1.91	7	4.90 ± 2.21	1	2.40 ± ---	0.42	0.015
SRC	Gleevec	14	8	11.87 ± 1.21	4	12.31 ± 2.80	2	9.91 ± 0.03	0	--- ± ---	0.01	>10
STK16	Staurosporine	1	1	0.40 ± ---	0	--- ± ---	0	--- ± ---	0	--- ± ---		0.2
SYK	Gleevec	1	1	4.44 ± ---	0	--- ± ---	0	--- ± ---	0	--- ± ---		>10
SYK	Staurosporine	1	1	0.70 ± ---	0	--- ± ---	0	--- ± ---	0	--- ± ---		0.007

[a]Six dockings were not successful. [b]Dockings with $N < 10$ are in gray. The four main columns denote mutually exclusive subsets of the analyzed dockings. For example, the top-200 hits contain no top-20 hits, and for this table, hits obtained using the mean docking score for the calibration data is also reported. The non-hits column depicts those dockings that scored worse than the calibration means. The last two columns have correlation coefficients for *Z*-scores vs ligand RMSD values and K_d values reported by Fabian et al.[50]

bleeding.[59] Common top-20 average nonkinase hits (listed in Supporting Information Table S12) include Subtilisin, blood clotting-related proteins, Albumin, Carboxyesterases, Epoxide Hydrolases, HSP-90α, and Histamine Methyltransferase.

Multiple CDK2 Structures. The VTS library contains 146 structures of Cyclin-dependent kinase 2 (CDK2). Of these, 101 grid files dock ligands into/near the ATP binding site. Despite the high number, varying numbers of CDK2 hits occurred (from 0 to 92; Table 3) for the 20 kinase inhibitors among the protein hit criteria. Within the top-200 average VTS hits, the three inhibitors that yielded the highest number of CDK2 hits were SP600126, roscovitine, and flavopiridol (92, 87, and 66, respectively), which are higher than any of the approved drugs, ranging from 2 to 45 CDK2 hits. Fabian reported only these 3 molecules and staurosporine (47 hits, top-20 criteria) with $K_d <$ 10 μM against CDK2. The primary target for roscovitine and flavopiridol is CDK2. Compound SP600125, its primary target being JNK, has been reported to act independently of JNK inhibition in its anticancer activity[60,61] with reported IC_{50} values of 3.96 and 22.2 μM for CDK2 bound to Cyclins A and E, respectively.[60] When comparing known cocrystal structures against corresponding kinase structures, it is observed that the average CDK2 rmsd = 1.22 ± 0.66 (maximum of 2.48) accounting for an average ligand rmsd of 5.14 ± 2.50 (maximum of 13.16).

Our CDK2 results suggest that having multiple structures of the same protein in the VTS protein library is beneficial, as might be expected since supplementing the library with conformational diversity should allow for less "induced fit effect" error in any virtual screening method. Figure 8 presents 3 CDK2 structures of protein rmsd values from 1.23 to 1.70 producing cross-dockings with ligand rmsd values from 0.71 to 6.04, representing a top-20 hit, a top-200 hit, and a nonhit.

Multiple protein structures and the "apparent goodness of virtual screening results" have been discussed in detail by Sheridan et al.,[62] Rockey et al.,[63] and Cavasotto et al.[64] Sheridan emphasized that in consideration of diverse small molecule structures, docking is sensitive to a single crystal protein structure. Furthermore, adding more crystal structures to a docking run will increase its accuracy. Instead of increasing the number of protein structures, flexible (protein) docking algorithms may be used; however, they are relatively slow compared to standard docking techniques, especially in terms of thousands of dockings per molecule as required for our VTS protein structure library. Having multiple protein structures in the VTS library increases conformational diversity without using costly simulation methods.

Kinase Inhibitor Pose Analysis. In order to assess how well VTS kinase hits compared to known inhibitor binding modes (or poses), X-ray cocrystallized structures matching 23

Table 5. X-ray Co-crystal Structures for Ligand Conformation Analysis of VTS Kinase Inhibitor Study

	kinase	inhibitor	co-crystal (PDB ID)						
1	ABL1	Gleevec	3K5 V	1OPJ	3MS9	3PYY	3MSS	1IEP	2HYY
2	CDK2	Roscovitine	2A4L	3DDQ					
3	CDK2	Staurosporine	1AQ1						
4	CDK5	Roscovitine	1UNL						
5	CSK	Staurosporine	1BYG						
6	EGFR	Iressa	2ITO	2ITY	2ITZ				
7	EGFR	Tarceva	1M17						
8	FYN	Staurosporine	2DQ7						
9	JNK1	SP600125	1UKI						
10	KIT	Gleevec	1T46						
11	KIT	Sutent	3G0E						
12	LCK	Gleevec	2PL0						
13	LCK	Staurosporine	1QPJ						
14	P38A	Nexavar	3HEG	3GCS					
15	P38A	BIRB796	1KV2						
16	P38A	Gleevec	3HEC						
17	P38A	SB203580	3GCP	1A9U					
18	PIM1	LY333531	2J2I						
19	PIM1	Staurosporine	1YHS						
20	SRC	Gleevec	2OIQ	3OEZ					
21	STK16	Staurosporine	2BUJ						
22	SYK	Gleevec	1XBB						
23	SYK	Staurosporine	1XBC						

unique (35 in total) kinase−inhibitor combinations (Table 5) were used as reference structures to compare ligand poses with VTS hit data, and Gscores were normalized to *Z*-scores. Cross-dockings (546) and 11 self-dockings were analyzed with these known cocrystals to correlate ligand rmsd values with VTS docking scores. Somewhat surprisingly, except for LCK−staurosporine (R = 0.73), all statistics generated from less than 10 dockings (gray area, Table 4) yielded negative correlation coefficients. However, only EGFR−Tarceva yielded no top-20 hits (compared to K_d = 0.0014 μM) in its two dockings: one cross-docking (PDB 1XKK[65]) and one self-docking (PDB 1M17[66]). Inspection of each PDB structure revealed reported water molecules (which were deleted during preparation for VTS) hydrogen bonding a threonine residue to Lapatinib (1XKK) and Tarceva (1M17) explaining the inability to reproduce the binding mode of Tarceva in EGFR and recognize it as a protein hit. Conversely, Fabian et al. reported K_d > 10 μM for Gleevec with p38α, SRC, and SYK whereas VTS predicted all as top-20 average protein hits. A literature search yielded the following K_d values: 34.0 μM for p38α, Namboodiri et al.;[67] 31 μM for SRC, Seeliger et al.;[68] and 5.0 μM for SYK, Atwell et al.[69] Karaman et al. reported binding results of K_d = 3.8 μM for SRC and K_d = 6 μM for SYK against Gleevec.[70] Also found were IC_{50} values, ∼1.2−7.5 μM, for Nexavar inhibiting EGFR in four different hepatocellular carcinoma cell lines.[71]

R values (Table 4) for p38α−SB203580 and SRC−Gleevec are notably low: 0.08 and 0.01, respectively. For all seven relevant SRC structures, the DFG loop[72] is in its active position ("DFG-in") not allowing Gleevec to be docked as in its cocrystals, PDB IDs 2OIQ[68] and 3OEZ[73] where the DFG loop is in the inactive position ("DFG-out"). The two reference structures for p38α−SB203580 are PDB IDs 1A9U[74] and 3GCP.[75] The DFG loop is in two different conformations for each of these structures while SB203580 retains the same binding mode in the hinge region. In PDB 3GCP, β-octylglucoside occupies the DFG-in position, while the DFG loop is in the inactive DFG-out position. Further, the SB203580 (3GCP) is involved with Pi-stacking between Phe 169 of the DFG loop and Tyr 35 of the glycine-rich loop. In 1A9U, there is only Pi-stacking between SB203580 and Tyr 35 while the DFG-loop is in the active DFG-in position. Correlation coefficients were calculated for two groupings of p38α structures, DFG-in and DFG-out, yielding R values of 0.61 and −0.09, respectively. SB203580 docking modes in the p38α/DFG-out group were frequently in the allosteric site. The docking modes in the p38α/DFG-in group depended on conformation of the glycine-rich loop and its Tyr 35 as well as the location of Phe 169 of the DFG loop (Figure 9).

Protein Hit Criteria Analysis. Of the 557 kinase−inhibitor dockings analyzed for the inhibitor pose analysis, there were 169 unique kinase structures. *Z*-scores were calculated, from this subset of the VTS library, for protein hit criteria using calibration statistics. These average *Z*-scores are $Z = -1.62 \pm 0.12$ for the top-200 average, $Z = -2.60 \pm 0.26$ for the top-20 average, and $Z = -3.15 \pm 0.49$ for the Boltzmann average. (See Supporting Information Figure S11.) These average Z-scores correspond to the approximate top 5%, top 0.5%, and top 0.1%, respectively. When the same statistics were evaluated for the entire VTS protein library (normally distributed, Supporting Information Table S16), the resulting values were -1.81 ± 0.51, -2.90 ± 0.73, and -3.37 ± 1.07, roughly equivalent to the top 5%, top 0.2%, and top 0.05%. These values were lower than the subset of VTS kinase entries. That is, the calibration on the NCI Diversity Set, a set of weak or better binders, yielded average hit criteria with lower (better) benchmarks than the subset of kinases. The top-200 hits consisted of approximately 3300 (nonunique) docking poses, 11% of all 29 400 dockings.

There were 890 top-20 hits, 3% of all dockings. However, the average normalized docking score for the top-200 average protein hits, −2.35, is approximately the top 1%. Therefore, the top-200 hits of the kinase inhibitors scored, on average, in the 98th percentile, which is reasonable since our test set is a group

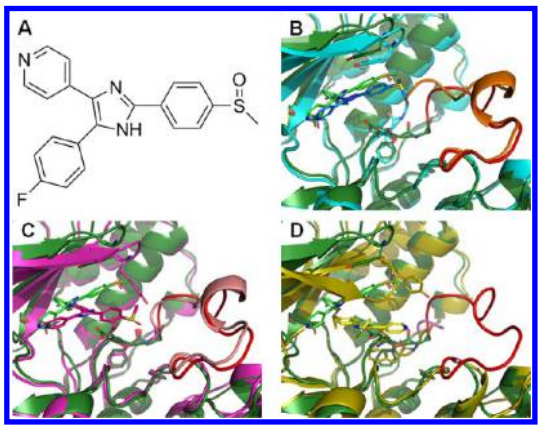


Figure 9. (A) Line structure of SB203580 and select VTS p38α docking poses. Protein kinases and docked poses are compared to PDB 3GCP (green). Examples include a top-20 hit (B, PDB 1OZ1, blue), a top-200 hit (C, PDB 1DI9, purple), and a nonhit (D, PDB 2BAL, yellow). *Z*-scores and ligand rmsd values are, respectively, −2.66 and 1.79 Å, −1.99 and 3.49 Å, and −0.12 and 9.81 Å. SB203580, the DFG loop (DFG-in position), and Tyr35 are in stick representation. The activation loop is off-colored with PDB 2BAL missing 12 residues (D, pink). Note the influence of Tyr35 on the glycine-rich loop above the docked poses.

of known inhibitors. It is not surprising that the top-200 criteria, the approximate top 5% cutoff, produced too many hits with a kinase percentage of 50%, compared to 60% in the top-20 hits, unable to adequately distinguish protein hits from the "noise." This noise influenced the top-200 hits such that the false negatives and false positives rates were higher than the top-20 and Boltzmann hits as mentioned earlier. For future studies, analysis of these (and perhaps other) statistical measures would help ensure more robust criteria for VTS to effectively produce protein hits.

CONCLUSIONS AND FUTURE DIRECTIONS

Our VTS system results on the kinase inhibitor test set have shown promise in the ability to characterize pan-kinase inhibitors, research compounds, and approved drugs. Our calibration procedure, though admittedly not fully optimized, was able to accurately predict inhibitor−kinase binding affinities when K_d < 10 μM and $K_d \geq$ 10 μM are both considered (72% accuracy in the best case). Thus our VTS system is able to robustly discriminate protein binders from nonbinders. It must be emphasized that a viable VTS system must be able to produce a reasonable signal-to-noise ratio which we submit we have accomplished with the top-20 average criteria. It is insufficient to simply find true protein hits if the concomitant false positive and false negative hit rates are high. Thus, when one inspects our VTS results for both bona fide hits and nonhits, the predictions using the top-20 average criterion are correct 72% of the time. We believe the inclusion of multiple copies of a protein structure contributed to VTS accuracy. Taken together, VTS offers a relatively rapid and accurate prediction of a given MOI's potential to bind to proteins that may not have been previously considered as one of its targets.

It is clear that the success of our VTS methodology relies on the calibration dockings with the NCI Diversity Set I and addition of multiple copies of protein structures. We conclude that the rigid body bias inherent to virtual screening using a single protein structure is decreased in our VTS system for two reasons. First, the structures used in the calibration set are likely to be influenced by protein conformational bias in a manner similar to the MOI. In general, molecules in the NCI Diversity Set with structural features similar to the MOI are likely to dock similarly and thus score similarly, reducing error in a systematic manner. Second, by incorporating multiple structures for proteins, conformational diversity may be increased among entries in the VTS protein database. Our CDK2 study revealed that even with 101 relevant protein structures (36 CDK2 grid files were not centered about the ATP binding site), known CDK2 binders were ranked as hits from 0 to 91% of the CDK2 proteins. Also, careful consideration is necessary when choosing more structures. For example, adding p38α structures having the DFG-out conformation should help enhance p38α dockings. Finally, it is known that docking scores and activity of a congeneric series of molecules in a specific protein target can correlate well (e.g., the work of Pauly et al.[76]). However, using only a set of closely related proteins would be counterproductive in the VTS context. On the other hand, using rankings of an MOI relative to average scores generated from the prior calibration of each protein has proved to be an effective strategy in separating signal from noise in VTS. This can be further enhanced by statistics on the normalized docking scores to determine proper hit criteria.

Although the top-20 average is optimal in determining protein hits over the top-200 and Boltzmann averages, future work will test the proposed optimization of protein hit criteria. MOI ligand similarity and ligand efficiencies of the calibration dockings will be investigated in order to minimize false positives and false negatives. Moreover, conformational diversity should be increased by adding more available PDB structures per protein, when available. MD simulations or other methods for generating ensembles of protein conformations could also be employed. Careful additions to the protein structure library as well as further small molecule test sets will be used to test VTS efficacy in determining toxicity, promiscuity, and narrowing focused libraries. Further validation of the VTS system in comparison with other available experimental data besides kinases is in progress.

ASSOCIATED CONTENT

Supporting Information

VTS library composition and kinase inhibitor information including protein hit listings as well as ligand pose analysis data. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Phone: (813)745-6047. E-mail: wesleybrooks@usf.edu (W.H.B.) or wayne.guida@moffitt.org (W.C.G.).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank the H. Lee Moffitt Cancer Center & Research Institute, Tampa, FL, for support in development of the VTS system. This research was supported in part by the U.S. Army Medical Research and Materiel Command, National Functional Genomics Center project, under award number W81XWH-08-2-0101. Opinions, interpretations, conclusions and recommendations are those of the author(s) and are not necessarily endorsed by the U.S. Army. The 1990 compound

NCI Diversity Set was provided by the Developmental Therapeutics Program, Division of Cancer Treatment and Diagnosis, National Cancer Institute, 6130 Executive Blvd., Room 8020, Rockville, MD 20852. The authors also thank the reviewers for many helpful suggestions and Dr. Kandethody M. Ramachandran for his discussion of statistics. Lastly, the authors acknowledge NIH (1K22HL088341- 01A1) and the University of South Florida (start-up) for funding.

REFERENCES

(1) Yang, X. X.; Hu, Z. P.; Chan, S. Y.; Zhou, S. F. Monitoring drug-protein interaction. *Clin. Chim. Acta* **2006**, *365* (1−2), 9−29.
(2) Zhang, J. X.; Huang, W. J.; Zeng, J. H.; Huang, W. H.; Wang, Y.; Zhao, R.; Han, B. C.; Liu, Q. F.; Chen, Y. Z.; Ji, Z. L. DITOP: drug-induced toxicity related protein database. *Bioinformatics* **2007**, *23* (13), 1710−1712.
(3) Gao, Z.; Li, H.; Zhang, H.; Liu, X.; Kang, L.; Luo, X.; Zhu, W.; Chen, K.; Wang, X.; Jiang, H. PDTD: a web-accessible protein database for drug target identification. *BMC Bioinformatics* **2008**, *9*, Article No. 104.
(4) Goede, A.; Dunkel, M.; Mester, N.; Frommel, C.; Preissner, R. SuperDrug: a conformational drug database. *Bioinformatics* **2005**, *21* (9), 1751−1753.
(5) Scheiber, J.; Jenkins, J. L.; Sukuru, S. C. K.; Bender, A.; Mikhailov, D.; Milik, M.; Azzaoui, K.; Whitebread, S.; Hamon, J.; Urban, L.; Glick, M.; Davies, J. W. Mapping Adverse Drug Reactions in Chemical Space. *J. Med. Chem.* **2009**, *52* (9), 3103−3107.
(6) Bleakley, K.; Yamanishi, Y. Supervised prediction of drug-target interactions using bipartite local models. *Bioinformatics* **2009**, *25* (18), 2397−2403.
(7) Nettles, J. H.; Jenkins, J. L.; Bender, A.; Deng, Z.; Davies, J. W.; Glick, M. Bridging chemical and biological space: "Target fishing" using 2D and 3D molecular descriptors. *J. Med. Chem.* **2006**, *49* (23), 6802−6810.
(8) Zoete, V.; Grosdidier, A.; Michielin, O. Docking, virtual high throughput screening and in silico fragment-based drug design. *J. Cell. Molec. Med.* **2009**, *13* (2), 238−248.
(9) Feher, M. Consensus scoring for protein-ligand interactions. *Drug Discovery Today* **2006**, *11* (9−10), 421−428.
(10) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28* (1), 235−242.
(11) Hert, J.; Willett, P.; Wilton, D. J. Comparison of fingerprint-based methods for virtual screening using multiple bioactive reference structures. *J. Chem. Inf. Comput. Sci.* **2004**, *44* (3), 1177−1185.
(12) Neugebauer, A.; Hartmann, R. W.; Klein, C. D. Prediction of protein-protein interaction inhibitors by chemoinformatics and machine learning methods. *J. Med. Chem.* **2007**, *50*, 4665−4668.
(13) Guvench, O.; MacKerell, A. D., Computational Fragment-Based Binding Site Identification by Ligand Competitive Saturation. *Plos Comput. Biol.* **2009**, *5*, (7).
(14) Marcou, G.; Rognan, D. Optimizing fragment and scaffold docking by use of molecular interaction fingerprints. *J. Chem. Inf. Model.* **2007**, *47* (1), 195−207.
(15) Zavodszky, M. I.; Stumpff-Kane, A. W.; Lee, D. J.; Feig, M. Scoring confidence index: statistical evaluation of ligand binding mode predictions. *J. Comput.-Aided Molec. Des.* **2009**, *23* (5), 289−299.
(16) Nissink, J. W. M. Simple Size-Independent Measure of Ligand Efficiency. *J. Chem. Inf. Model.* **2009**, *49* (6), 1617−1622.
(17) Vigers, G. P. A.; Rizzi, J. P. Multiple active site corrections for docking and virtual screening. *J. Med. Chem.* **2004**, *47* (1), 80−89.
(18) Gohlke, H.; Klebe, G. Statistical potentials and scoring functions applied to protein-ligand binding. *Curr. Opin. Struct. Biol.* **2001**, *11* (2), 231−235.
(19) Li, Y. Y.; An, J.; Jones, S. J. A computational approach to finding novel targets for existing drugs. *PLoS Comput. Biol.* **2011**, *7* (9), e1002139.
(20) Boguski, M. S.; Mandl, K. D.; Sukhatme, V. P. Repurposing with a Difference. *Science* **2009**, *324* (5933), 1394−1395.
(21) 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* (2), 197−206.
(22) Kellenberger, E.; Foata, N.; Rognan, D. Ranking targets in structure-based virtual screening of three-dimensional protein libraries: Methods and problems. *J. Chem. Inf. Model.* **2008**, *48* (5), 1014−1025.
(23) Keiser, M. J.; Setola, V.; Irwin, J. J.; Laggner, C.; Abbas, A. I.; Hufeisen, S. J.; Jensen, N. H.; Kuijer, M. B.; Matos, R. C.; Tran, T. B.; Whaley, R.; Glennon, R. A.; Hert, J.; Thomas, K. L. H.; Edwards, D. D.; Shoichet, B. K.; Roth, B. L. Predicting new molecular targets for known drugs. *Nature (London)* **2009**, *462* (7270), 175.
(24) O'Connor, K. A.; Roth, B. L. Finding new tricks for old drugs: An efficient route for public-sector drug discovery. *Nat. Rev. Drug Discovery* **2005**, *4* (12), 1005−1014.
(25) Valente, A. P.; Miyamoto, C. A.; Almeida, F. C. L. Implications of protein conformational diversity for binding and development of new biological active compounds. *Curr. Med. Chem.* **2006**, *13* (30), 3697−3703.
(26) Ashburn, T. T.; Thor, K. B. Drug repositioning: Identifying and developing new uses for existing drugs. *Nat. Rev. Drug Discovery* **2004**, *3* (8), 673−683.
(27) Li, H. L.; Gao, Z. T.; Kang, L.; Zhang, H. L.; Yang, K.; Yu, K. Q.; Luo, X. M.; Zhu, W. L.; Chen, K. X.; Shen, J. H.; Wang, X. C.; Jiang, H. L. TarFisDock: a web server for identifying drug targets with docking approach. *Nucleic Acids Res.* **2006**, *34*, W219−W224.
(28) Chen, Y. Z.; Zhi, D. G. Ligand-protein inverse docking and its potential use in the computer search of protein targets of a small molecule. *Proteins−Struct. Funct. Genet.* **2001**, *43* (2), 217−226.
(29) Zahler, S.; Tietze, S.; Totzke, F.; Kubbutat, M.; Meijer, L.; Vollmar, A. M.; Apostolakis, J. Inverse in silico screening for identification of kinase inhibitor targets. *Chem. Biol.* **2007**, *14*, 1207−1214.
(30) Kortagere, S.; Krasowski, M. D.; Ekins, S. The importance of discerning shape in molecular pharmacology. *Trends Pharmacol. Sci.* **2009**, *30* (3), 138−147.
(31) Rockey, W. M.; Elcock, A. H. Progress toward virtual screening for drug side effects. *Proteins−Struct. Funct. Genet.* **2002**, *48* (4), 664−671.
(32) Chen, Y. Z.; Ung, C. Y. Prediction of potential toxicity and side effect protein targets of a small molecule by a ligand-protein inverse docking approach. *J. Molec. Graph. Modell.* **2001**, *20* (3), 199−218.
(33) Swaan, P. W.; Ekins, S. Reengineering the pharmaceutical industry by crash-testing molecules. *Drug Discovery Today* **2005**, *10* (17), 1191−1200.
(34) Steffen, A.; Thiele, C.; Tietze, S.; Strassnig, C.; Kamper, A.; Lengauer, T.; Wenz, G.; Apostolakis, J. Improved cyclodextrin-based receptors for camptothecin by inverse virtual screening. *Chem.—Eur. J.* **2007**, *13*, 6801−6809.
(35) *Maestro*, 8.0; Schrödinger, LLC: New York, NY, 2007.
(36) *LigPrep*, 1.6; Schrödinger, LLC: New York, NY, 2005.
(37) Holbeck, S. L. Update on NCI in vitro drug screen utilities. *Eur. J. Cancer* **2004**, *40* (6), 785−793.
(38) *LigPrep*, 2.2; Schrödinger, LLC: New York, NY, 2008.
(39) Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C. MacroModel - an integrated software system for modeling organic and bioorganic molecules using molecular mechanics. *J. Comput. Chem.* **1990**, *11* (4), 440−467.
(40) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: A new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* **2004**, *47* (7), 1750−1759.
(41) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47* (7), 1739−49.

(42) *Glide*, 5.6; Schrödinger, LLC: New York, NY, 2010.
(43) Delano, W. L. *The PyMol Molecular Graphics System*, 0.99; Delano Scientific: Palo Alto, CA, USA, 2002.
(44) Halgren, T.; Bush, B., The Merck molecular force field (MMFF94). Extension and application. *Abstr. Pap. Am. Chem. Soc.* **1996**, *212*, 2-COMP.
(45) Halgren, T. Merck molecular force field 0.5. Extension of MMFF94 using experimental data, additional computational data, and empirical rules. *J. Comput. Chem.* **1996**, *17* (5−6), 616−641.
(46) Halgren, T.; Nachbar, R. Merck molecular force field 0.4. Conformational energies and geometries for MMFF94. *J. Comput. Chem.* **1996**, *17* (5−6), 587−615.
(47) Halgren, T. Merck molecular force field 0.3. Molecular geometries and vibrational frequencies for MMFF94. *J. Comput. Chem.* **1996**, *17* (5−6), 553−586.
(48) Halgren, T. Merck molecular force field 0.2. MMFF94 van der Waals and electrostatic parameters for intermolecular interactions. *J. Comput. Chem.* **1996**, *17* (5−6), 520−552.
(49) Halgren, T. Merck molecular force field 0.1. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **1996**, *17* (5−6), 490−519.
(50) Fabian, M. A.; Biggs, W. H.; 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* (3), 329−336.
(51) Jorgensen, W. L.; Tiradorives, J. The OPLS potential functions for proteins—energy minimizations for crystals of cyclic-peptides and crambin. *J. Am. Chem. Soc.* **1988**, *110* (6), 1657−1666.
(52) Jorgensen, W.; Maxwell, D.; TiradoRives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **1996**, *118* (45), 11225−11236.
(53) Hendrickson, W. A.; Horton, J. R.; LeMaster, D. M. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J.* **1990**, *9* (5), 1665−72.
(54) *Maestro*, 9.1; Schrödinger, LLC: New York, NY, 2010.
(55) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47* (7), 1739−1749.
(56) Miller, B. T.; Singh, R. P.; Klauda, J. B.; Hodoscek, M.; Brooks, B. R.; Woodcock, H. L. CHARMMing: a new, flexible web portal for CHARMM. *J. Chem. Inf. Model.* **2008**, *48* (9), 1920−9.
(57) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **2009**, *30* (10), 1545−614.
(58) MacKerell, A. D., Jr.; Bashford, D.; Bellot, M.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; McCarthy, D. J.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102* (18), 3596−3616.
(59) AstraZeneca Caprelsa(R) (vandetanib) Tablets. http://www1.astrazeneca-us.com/pi/caprelsa.pdf (June 16, 2012).
(60) Du, L. H.; Lyle, C. S.; Obey, T. B.; Gaarde, W. A.; Muir, J. A.; Bennett, B. L.; Chambers, T. C. Inhibition of cell proliferation and cell cycle progression by specific inhibition of basal JNK activity - Evidence that mitotic Bcl-2 phosphorylation is JNK-independent. *J. Biol. Chem.* **2004**, *279* (12), 11957−11966.
(61) Kim, J. A.; Lee, J.; Margolis, R. L.; Fotedar, R. SP600125 suppresses Cdk1 and induces endoreplication directly from G2 phase, independent of JNK inhibition. *Oncogene* **2010**, *29* (11), 1702−1716.
(62) Sheridan, R. P.; McGaughey, G. B.; Cornell, W. D. Multiple protein structures and multiple ligands: effects on the apparent goodness of virtual screening results. *J. Comput. Aided Mol. Des.* **2008**, *22* (3−4), 257−65.
(63) Rockey, W. M.; Elcock, A. H. Structure selection for protein kinase docking and virtual screening: homology models or crystal structures? *Curr. Protein Pept. Sci.* **2006**, *7* (5), 437−57.
(64) Cavasotto, C. N.; Abagyan, R. A. Protein flexibility in ligand docking and virtual screening to protein kinases. *J. Mol. Biol.* **2004**, *337* (1), 209−25.
(65) Wood, E. R.; Truesdale, A. T.; McDonald, O. B.; Yuan, D.; Hassell, A.; Dickerson, S. H.; Ellis, B.; Pennisi, C.; Horne, E.; Lackey, K.; Alligood, K. J.; Rusnak, D. W.; Gilmer, T. M.; Shewchuk, L. A unique structure for epidermal growth factor receptor bound to GW572016 (Lapatinib): relationships among protein conformation, inhibitor off-rate, and receptor activity in tumor cells. *Cancer Res.* **2004**, *64* (18), 6652−9.
(66) Stamos, J.; Sliwkowski, M. X.; Eigenbrot, C. Structure of the epidermal growth factor receptor kinase domain alone and in complex with a 4-anilinoquinazoline inhibitor. *J. Biol. Chem.* **2002**, *277* (48), 46265−72.
(67) Namboodiri, H. V.; Bukhtiyarova, M.; Ramcharan, J.; Karpusas, M.; Lee, Y.; Springman, E. B. Analysis of imatinib and sorafenib binding to p38alpha compared with c-Abl and b-Raf provides structural insights for understanding the selectivity of inhibitors targeting the DFG-out form of protein kinases. *Biochemistry* **2010**, *49* (17), 3611−8.
(68) Seeliger, M. A.; Nagar, B.; Frank, F.; Cao, X.; Henderson, M. N.; Kuriyan, J. c-Src binds to the cancer drug imatinib with an inactive Abl/c-Kit conformation and a distributed thermodynamic penalty. *Structure* **2007**, *15* (3), 299−311.
(69) Atwell, S.; Adams, J. M.; Badger, J.; Buchanan, M. D.; Feil, I. K.; Froning, K. J.; Gao, X.; Hendle, J.; Keegan, K.; Leon, B. C.; Müller-Dieckmann, H. J.; Nienaber, V. L.; Noland, B. W.; Post, K.; Rajashankar, K. R.; Ramos, A.; Russell, M.; Burley, S. K.; Buchanan, S. G. A novel mode of Gleevec binding is revealed by the structure of spleen tyrosine kinase. *J. Biol. Chem.* **2004**, *279* (53), 55827−32.
(70) Karaman, M. W.; Herrgard, S.; Treiber, D. K.; Gallant, P.; Atteridge, C. E.; Campbell, B. T.; Chan, K. W.; Ciceri, P.; Davis, M. I.; Edeen, P. T.; Faraoni, R.; Floyd, M.; Hunt, J. P.; Lockhart, D. J.; Milanov, Z. V.; Morrison, M. J.; Pallares, G.; Patel, H. K.; Pritchard, S.; Wodicka, L. M.; Zarrinkar, P. P. A quantitative analysis of kinase inhibitor selectivity. *Nat. Biotechnol.* **2008**, *26* (1), 127−32.
(71) Ezzoukhry, Z.; Louandre, C.; Trécherel, E.; Godin, C.; Chauffert, B.; Dupont, S.; Diouf, M.; Barbare, J. C.; Mazière, J. C.; Galmiche, A. EGFR activation is a potential determinant of primary resistance of hepatocellular carcinoma cells to sorafenib. *Int. J. Cancer* **2012**, DOI: 10.1002/ijc.27604. The PubMed ID (PMID) is 22514082.
(72) Knighton, D. R.; Zheng, J. H.; Ten Eyck, L. F.; Ashford, V. A.; Xuong, N. H.; Taylor, S. S.; Sowadski, J. M. Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. *Science* **1991**, *253* (5018), 407−14.
(73) Lovera, S.; Sutto, L.; Boubeva, R.; Scapozza, L.; Dölker, N.; Gervasio, F. L. The different flexibility of c-Src and c-Abl kinases regulates the accessibility of a druggable inactive conformation. *J. Am. Chem. Soc.* **2012**, *134* (5), 2496−9.
(74) Wang, Z.; Canagarajah, B. J.; Boehm, J. C.; Kassisà, S.; Cobb, M. H.; Young, P. R.; Abdel-Meguid, S.; Adams, J. L.; Goldsmith, E. J. Structural basis of inhibitor selectivity in MAP kinases. *Structure* **1998**, *6* (9), 1117−28.

(75) Simard, J. R.; Getlik, M.; Grütter, C.; Pawar, V.; Wulfert, S.; Rabiller, M.; Rauh, D. Development of a fluorescent-tagged kinase assay system for the detection and characterization of allosteric kinase inhibitors. *J. Am. Chem. Soc.* **2009**, *131* (37), 13286−96.
(76) Pauly, G. T.; Loktionova, N. A.; Fang, Q.; Vankayala, S. L.; Guida, W. C.; Pegg, A. E. Substitution of aminomethyl at the meta-position enhances the inactivation of O6-alkylguanine-DNA alkyltransferase by O6-benzylguanine. *J. Med. Chem.* **2008**, *51* (22), 7144−53.