



## Computational approach to the identification of novel Aurora-A inhibitors

Mohammad Neaz Morshed<sup>a,b</sup>, Yong Seo Cho<sup>a</sup>, Seon Hee Seo<sup>a</sup>, Ki-Cheol Han<sup>a</sup>, Eun Gyeong Yang<sup>a</sup>, Ae Nim Pae<sup>a,\*</sup>

<sup>a</sup>Neuro-Medicine Center, Life Sciences Division, Korea Institute of Science and Technology, PO Box 131, Cheongryang, Seoul 130-650, Republic of Korea

<sup>b</sup>School of Science, Korea University of Science and Technology, 52 Eoeun dong, Yuseong-gu, Daejeon 305-333, Republic of Korea

### ARTICLE INFO

#### Article history:

Received 14 October 2010

Revised 25 November 2010

Accepted 30 November 2010

Available online 4 December 2010

#### Keywords:

Aurora-A

Anticancer

Docking

Pharmacophore

Virtual screening

Tumor cell

Inhibitor

### ABSTRACT

Aurora kinase A has been emerging as a key therapeutic target for the design of anticancer drugs. For the purpose of finding biologically active and novel compounds and providing new ideas for drug-design, we performed virtual screening using commercially available databases. A three-dimensional common feature pharmacophore model was developed with the HipHop program provided in the Catalyst software package, and this model was used as a query for screening the databases. A recursive partitioning (RP) model was developed as a filtering system, which was able to classify active and inactive compounds. Eventually, a step-wise virtual screening procedure was conducted by applying the common feature pharmacophore and the RP model in succession to discover novel potent Aurora-A inhibitors. A total of 68 compounds were selected for testing of their in vitro anticancer activities against various human cancer cell lines. Based on the activity data, we have identified fifteen compounds that warrant further investigation. Several compounds have a high inhibition rate (above 80% at 10  $\mu$ M) and a  $GI_{50}$  lower than 5  $\mu$ M for the cell lines DU145 and HT29. Enzyme assay for these compounds identified hits with micro molar activity. Compound **C11** has the highest activity ( $IC_{50}$  = 5.09  $\mu$ M). The hits obtained from this screening scheme could be potential drug candidates after further optimization.

© 2010 Elsevier Ltd. All rights reserved.

### 1. Introduction

The Aurora kinase family consists of highly related serine/threonine kinases that are involved in the regulation of mitosis.<sup>1</sup> Aurora kinases have been conserved throughout eukaryotic evolution and have evolved into three related kinases known as Aurora-A, Aurora-B, and Aurora-C in mammalian cells. All Aurora kinases contain a variable N-terminal domain followed by a conserved catalytic domain, and a short C-terminal extension. Despite significant sequence homology, the localization and functions of these kinases are largely distinct from one another. Aurora-A localizes to centrosomes during early S phase and is involved in centrosome maturation and separation, bi-polar spindle assembly, mitotic entry, and mitotic exit. Aurora-B kinase belongs to the chromosome passenger protein family. Less is known about Aurora-C kinases, which are specifically expressed at high levels in the testis and show centrosomal localization from anaphase to telophase.

Overexpression of Aurora-A has been shown to lead to centrosome amplification and aneuploidy, which usually results from incomplete cytokinesis and can be a driving force in genomic instability and tumorigenesis.<sup>2–5</sup> Aurora-A also has been implicated in the regulation of cell cycle checkpoints. Checkpoint

defects may ultimately contribute to genomic instability and carcinogenesis.<sup>6–8</sup> Recent studies have shown that overexpression of Aurora-A in cultured cells induces several cancer-related phenotypes, including increased cell proliferation and colony formation and inhibition of UV- or cisplatin-induced apoptosis.<sup>9</sup> Overexpression of Aurora-A can also transform rat-1 and NIH3T3 cells and form tumors in null mice.<sup>3,5</sup> Collectively, this evidence indicates that Aurora-A acts as an oncogene and plays an important role in cell cycle progression and carcinogenesis. Hence, the development of small molecule inhibitors of Aurora-A that target the ATP binding site is emerging as a new anticancer target-based therapeutic approach. Moreover, in vivo studies with the Aurora inhibitors **I** (MK-0457/VX-680),<sup>10,11</sup> **II** (PHA-739358),<sup>12</sup> **III** (MLN8054),<sup>13</sup> and **IV** (AZD1152)<sup>14,15</sup> (Fig. 1) in various animal models have shown tumor regression and are now in different stages of clinical development for various cancers.

The aim of this study was to find Aurora-A inhibitors with novel chemical scaffolds. We applied ligand and receptor-based virtual screening approaches. Initially HipHop pharmacophore-based virtual screening was performed, and this was then followed by additional screening models. This approach allows large libraries to be screened rapidly and would be expected to lead to more reliable enrichment. Figure 2 shows the screening strategy. The HipHop model, generated from a small set of known inhibitors, was the first screening filter. Compounds selected based on fit value criteria (fit value between 4.0 and 5.0) were screened in

\* Corresponding author. Tel.: +82 2 958 5185; fax: +82 2 958 5189.

E-mail address: [anpae@kist.re.kr](mailto:anpae@kist.re.kr) (A.N. Pae).

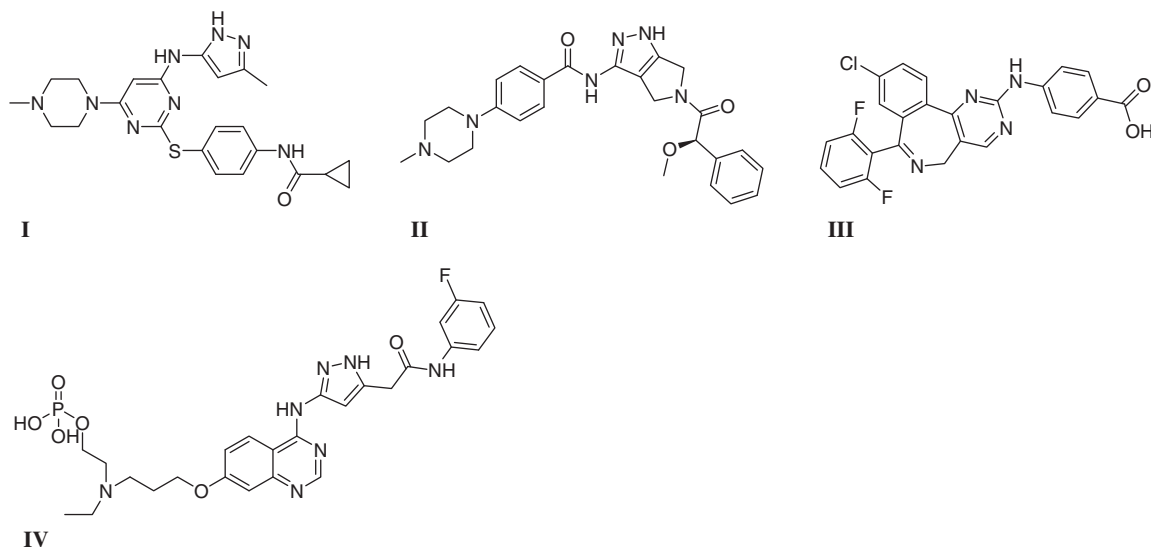


Figure 1. Aurora kinase inhibitors I–IV in clinical testing as anticancer treatment.

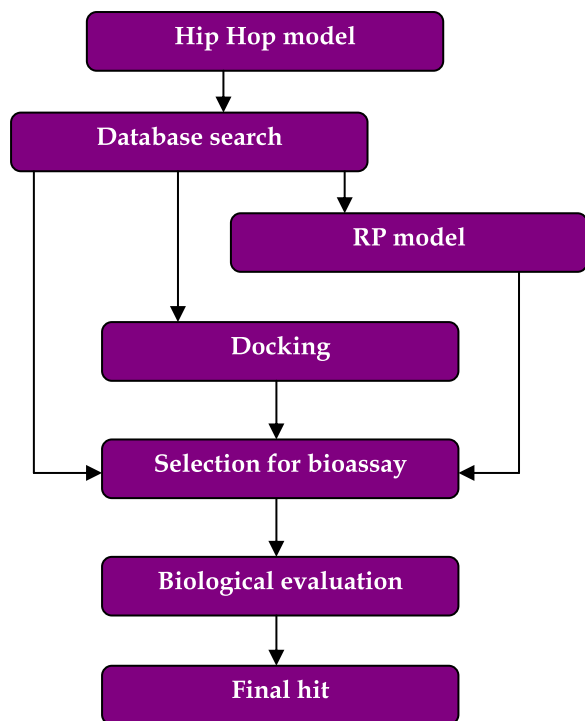


Figure 2. Flowchart of virtual screening process.

the second step. Three types of filtering approaches with different selection criteria were applied in this step. In the first approach, the compounds from the first step were docked into the active site. The compounds were selected by considering a predicted docking pose and by observing the important interactions necessary to be active. In the second approach compounds were selected considering the diversity among the top fit value hits from the first filter. The final filtering approach was the application of a recursive partitioning (RP) model. Following these filtering methods, hits were further validated via the measurement of biological activity. In summary, the initial screening based on the pharmacophore model screened the database by checking essential functional features in each molecule, a docking based approach refined the

hits considering the essential interactions in the active site of the target, and the RP-based virtual screening discriminated between active and inactive compounds by evaluating the molecular topological pattern. The RP model is very attractive for several reasons: it utilizes property descriptors with continuous value ranges, it transforms these descriptors into a binary classification scheme, and it generates the resulting decision tree. This makes it easy for even non-statisticians to distinguish between active and inactive compounds.

## 2. Materials and methods

### 2.1. Ligand preparation

A total of 1,047,671 compounds in the database were prepared by adding hydrogen atoms and rectification of wrong valences. The LigPrep module implemented in the Schrödinger package was used to prepare pharmacophore-search-based 3863 hits for molecular docking simulation. This two-dimensional (2D) to three-dimensional (3D) conversion program generates accurate energy minimized molecular structures with their tautomeric structures, ionization states, ring conformations, and stereoisomers to produce broad chemical and structural diversity from a single input structure. The compounds were only restricted to 3D structure generation and their ionization states at physiological pH by retaining their original chirality.

### 2.2. Generation and validation of the common feature pharmacophore model

A pharmacophore model, according to IUPAC definition, is 'an ensemble of steric and electronic features that is necessary to ensure the optimal intermolecular interactions with a specific biological target and to trigger (or block) its biological response'. Typically, pharmacophore features include hydrophobic, aromatic, hydrogen bond acceptor, hydrogen bond donor, positive and negative ionizable. In the application, common feature pharmacophores are generated using the HipHop algorithm. HipHop identifies configurations or three-dimensional spatial arrangements of chemical features that are common to molecules in a training set. The configurations are identified by a pruned exhaustive search, starting with small sets of features and extending them until no larger

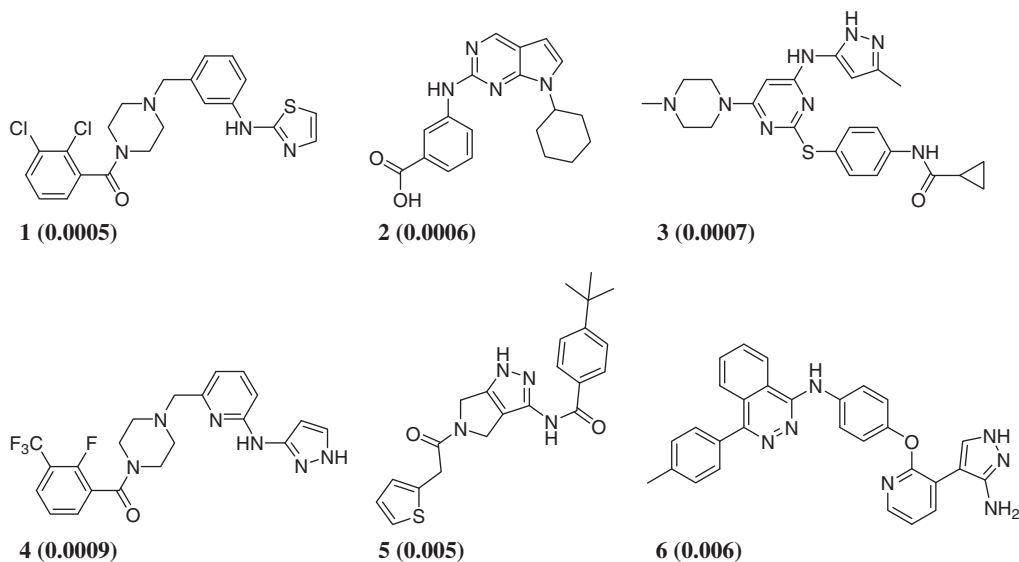
common configuration is found. Training set members are evaluated on the basis of the types of chemical features they contain, along with the ability to adopt a conformation that allows those features to be superimposed on a particular configurations. HipHop also maps partial features of molecules in the alignment set. This provision gives the option to use partial mapping during the alignment. Partial mapping allows identifying larger, more diverse, and more significant hypotheses and alignment models without the risk of missing compounds that do not map to all of the pharmacophore features. The resultant pharmacophores are ranked as they are built. The ranking is a measure of how well the molecules map onto the proposed pharmacophores, as well as the rarity of the pharmacophore model. If a pharmacophore model is less likely to map to an inactive compound, it will be given a higher rank; the reverse is also true.

The common feature hypothesis, HipHop pharmacophore model, is the automated tool within Catalyst that is based on the alignment of common features present in highly potent compounds. For HipHop pharmacophore analysis for published Aurora-A inhibitors, six compounds (Fig. 3) were carefully selected by considering the activity, structural rigidity, and diversity.<sup>10,11,16–18</sup> All the compounds here are highly active and have different central moiety. For example, compounds **2** and **3** both have pyrimidine moiety but one forms pyrrolo-pyrimidine (compound **2**) and other remains as pyrimidine itself. So this difference will generate distinct feature for pharmacophore model. In similar way, pyrazole ring in compounds **3** and **5** can cause the diversity of structure and generate variety of features in pharmacophore model. The conformational models of selected compounds (having up to 250 conformers) were built in Catalyst using the 'best conformer generation' method with a 20 kcal/mol energy cutoff.<sup>19</sup> HipHop pharmacophore models were derived by comparing a set of conformational models and a number of 3D configurations of chemical features shared among the training set molecules. The parameter settings of 'MaximumOmittedFeatures', 'Misses', and 'CompleteMisses' were varied to generate multiple hypotheses in which some compounds may or may not fit all features. Four chemical functions with hydrogen bonding acceptor, hydrogen bonding donor, aromatic ring, and hydrophobic group were used as the pharmacophore feature. The hypothesis generation process in Catalyst returned 10 possible pharmacophore hypotheses with a different arrangement of the constituent features or ranking

score. After deleting the redundant hypotheses that had the same chemical characteristics and nearly identical distances between these functions, diverse configurations of hypotheses were selected according to the ranking and fitting scores. The best hypothesis was determined by the potential of discriminating between active and inactive compounds. As an active data set, all known 141 Aurora-A inhibitors were collected from the Integrity database of Prous ([www.prous.com](http://www.prous.com)) and were as added to the Catalyst database.

### 2.3. Generation of the recursive partitioning model

The Recursive partitioning (RP) model was generated using the CART algorithm implemented in the Cerius<sup>2</sup> program. RP is a statistical method for multivariable analysis. It selects variables automatically, is non-parametric and captures nonlinear relationships automatically. Classification and regression trees (CART) is a non-parametric Decision tree learning technique that produces either classification or regression trees, depending on whether the dependent variable is categorical or numeric, respectively. The CART methodology involves RP of a dataset matrix to make a hierarchical decision tree (rows = compounds and columns = descriptors, X and activities, Y). The decision tree is constructed by one question per node, creating a binary (yes/no) split and resulting in two statistically distinct subsets or nodes. The splitting criteria are determined by a statistical analysis of each variable and the assigned categorical activity of the compounds. The splitting process continues until no more significant nodes are obtained or when a minimum number of samples per node are reached (a specified parameter) and a class prediction is made on the terminal node. The RP model is found to be sensitive to the descriptors used, and diversity of the data sets can radically change the property of the decision tree. The method is applicable to structurally unique compounds with activity data to uncover substructural rules that govern the biological activity. The RP classification tree is often of great interest to visualize the distribution of potencies at the node and to see how a split at a node divides the potencies at two daughter nodes. This method has been repeatedly used by researchers of bioinformatics and chemoinformatics, either to classify genes or to differentiate active and inactive compounds. However, the limitation of the RP method is its inability to extrapolate beyond the range of observed responses. The main objective of



**Figure 3.** The structures and activities (IC<sub>50</sub>, μM) of Aurora-A inhibitors used as training set for pharmacophore generation.

incorporating the RP method in the virtual screening process is to rapidly classify unknown compounds based on a small number of readily interpretable descriptors; therefore, for screening compounds.

The recursive partition decision tree model was constructed using a QSAR module of Cerius<sup>2</sup> version 4.10.17. The RP tree was constructed by E-state key and the topological descriptors based on the chemical graph theory. The activity classes were weighted equally, and the splits were scored using the Gini Impurity scoring function. The pruning factor values were varied between 3 and 6. The value of 1/1000 of the samples was considered as the minimum number of samples in any node. The various values were used for maximum tree depth (layers  $\leq 10$ ), and the default values were accepted for a maximum number of generic splits (30) and the number of knots per variable (20).

The true positive Aurora-A inhibitors were set as the active class, and the inactive compounds obtained from the literature and in-house library were selected as the inactive class. A total of 377 compounds were classified into two categories: the active class (1), which includes the compounds having an activity range below or equal to 1  $\mu$ M; and the inactive class (0), which covers the activity range of more than 1  $\mu$ M. There were 174 active and 203 inactive compounds by this activity range. Training set consists of 341 (156 active and 185 inactive) compounds. Two-dimensional and three-dimensional descriptors of Cerius<sup>2</sup> were used for the RP tree generation. The descriptors were optimized by means of removing those with constant values and 95% of the zero values, while some of the descriptors were deleted on the basis of the correlation threshold  $\geq 0.9$ . In the RP study, we defined the activity class (1 or 0) column as a dependent variable (Y) and the descriptors used as independent variables (X).

## 2.4. Molecular docking

To perform interaction-based selection of the hits obtained from screening, a molecular docking study was carried out in the ligand binding site of 2bmc using GLIDE. All crystallographic water molecules and co-crystal ligand were removed from the crystal structure. Hydrogen atoms and charges using OPLS\_2005 force field were added and brief minimization to relieve steric clashes was performed using the protein preparation module in Maestro with the 'preparation and refinement' option. This restrained partial minimization was terminated when the root-mean-square deviation (rmsd) reached a maximum value of 0.3 Å. A grid file was created within the area of 6.5 Å around the co-crystallized ligand, and the core interactions (Glu 211 and Ala 213) were defined in the grid. The additional interaction with Lys 162 present in this crystal structure was also included in grid file generation. Hydrogen bonds were used as a constraint while performing the docking of the compounds. The standard-precision (SP) mode of docking was performed, and the Glide scoring function (G-Score) was used to select the final poses for each ligand.

## 2.5. Virtual screening using an external library

Virtual screening was carried out by combining the ligand and receptor-based approaches to obtain new compounds with desired activity profiles. The commercial library Asinex (AsinexGold, 229,398 compounds, AsinexPlatinum, 125,231 compounds, Asinex, Moscow, Russia, [www.asinex.com](http://www.asinex.com)) and ChemDiv (693,042 compounds, ChemDiv, Inc. California, USA, [www.chemdiv.com](http://www.chemdiv.com)) have been utilized for virtual screening.

The best pharmacophore model was used for the virtual screening experiment, by selecting the fast flexible database search option. The first screening selected new compounds with similar functional and spatial properties defined in 3D pharmacophore query.

Compounds were selected by fit value (fit value between 4.0 and 5.0) and were screened in the second step considering various criteria. In the second step, three types of filtering approach/selection criteria were applied. One of them selected compounds considering the predicted docking pose. In this approach the compounds from the first step were docked into the protein active site for the selection by observing the important interactions necessary for activity. The second approach allowed the selection of compounds by considering the diversity among the top fit value hits from the first filter. The final approach in the second step was the application of recursive partitioning (RP) model. The final hit compounds were selected by following the criteria of all three methods applied in the second step of screening.

## 2.6. Biological evaluation

### 2.6.1. MTT assay

Cytotoxic activity of the screened compounds against human cancer cell lines was investigated using the MTT assay. Human colon adenocarcinoma (HT-29) and human prostate cancer (DU145) cell lines were supplied by the Korean Cell Line Bank, Seoul National University. All cell lines were grown in RPMI 1640 (Gibco BRL) supplemented with 10% (V/V) heat-inactivated fetal bovine serum (FBS) and maintained at 37 °C in a humidified atmosphere with 5% CO<sub>2</sub>. The cells ( $3 \times 10^3$  cells/well) were seeded into 96-well plates. Various concentrations of samples were added to each well in duplicate and then incubated at 37 °C with 5% CO<sub>2</sub> for two days so that time cells would be in the exponential phase of growth at the time of drug addition. 15  $\mu$ L of the Dye Solution (Promega, CellTiter96) was added to each well. The plate was incubated at 37 °C for up to 4 h in a humidified 5% CO<sub>2</sub> atmosphere. After incubation, 100  $\mu$ L of the Solubilization Solution/Stop Mix (Promega, CellTiter96) were added to each well. The plate was allowed to stand overnight in a sealed container with a humidified atmosphere at room temperature to completely solubilize the formazan crystals. The optical density at 570 nm was measured using a microplate reader (Versamax, Molecular Devices), and the anti-cancer effective concentration was expressed as a GI<sub>50</sub>.

### 2.6.2. Kinase assay

Kinase assays were performed at Reaction Biology Corporation ([www.reactionbiology.com](http://www.reactionbiology.com), Malvern PA) using the 'HotSpot' assay platform. Assay protocol is supplied in [Supplementary data](#).

## 3. Results and discussion

### 3.1. Generation of the best common feature model

For the HipHop pharmacophore analysis, highly active Aurora-A inhibitors were chosen as described in the materials and methods section. All active ligands do not have the same functional groups, and therefore, the active ligands adopted different binding modes to maximize their interactions with the active site. Thus, all possible pharmacophore models were inspected by altering the parameter options to determine whether some compounds reflected all of the features. In the HipHop pharmacophore generation, the molecules are used to construct the pharmacophore configuration space that can be specified through setting the 'principal' compound parameter. This parameter can adopt the values of 0, 1, or 2. A value of 2 means this molecule will be used to build the configuration space. A value of 0 means this molecule will be ignored when building the configuration space. A value of 1 indicates to the program that the molecule can be used to build the configuration space, but whether this molecule is actually used will depend on the setting of other parameters ('Misses' and 'CompleteMisses').



**Table 1**

The common feature hypotheses and corresponding hit rates against active compounds

Pharmacophores	Pharmacophore composition	Actives (141)
A	HHHDAA	88 (63%)
B	RHHDA	100 (70%)
C	RHHDA	106 (75%)
D	RHHDA	110 (78%)

'Misses' means that hypotheses that fail to map completely to more than one training compound will be disallowed. 'Complete-Misses' specify the number of molecules that do not have to map to any features in the hypothesis. The 'Misses' and 'CompleteMisses' were varied from 1 to 3, respectively. The value for 'MaxOmitFeat' was set to 2 so that a pharmacophore could be maintained, even if a specific molecule is a complete miss for that pharmacophore. Table 1 reports the diverse pharmacophore model compositions collected from the repeated trials and their hit rate against active compounds.

Pharmacophore D showed a superior active hit rate relative to pharmacophore A, B, and C. Therefore, we selected the pharmacophore D as the best pharmacophore model for virtual screening. Pharmacophore D also appeared to give more effective discrimination against poorly potent inhibitors. The best pharmacophore query was comprised of one hydrogen bond acceptor, one hydrogen bond donor, one ring aromatic, and two hydrophobic features. The distances between pharmacophoric sites of the model D is shown in Figure 7.

### 3.2. Recursive partitioning tree for post-filtering

To classify the hit compounds from the common feature pharmacophore-based virtual screening into active or inactive classes, the recursive partitioning model was developed using the two-dimensional descriptors implying molecular shape information. The statistical results for the best RP model are demonstrated in Table 2. The definition, 'Class%ObsCorr', is a measure of the number of compounds correctly predicted to belong to a class as a percentage of the total number of compounds observed to be in each class. It provides information on false negatives as well as false positives. Here we can see 80.13% active and 90.27% inactive are classified correctly by the model shown in Figure 5. The measure of 'Overall%PredCorr' represents the total number of compounds correctly classified divided by the number of compounds predicted to belong to each class. It provides information on the accuracy of the prediction. The enrichment factor for a specific class is the ratio of the 'Overall%PredCorr' to the original percentage of compounds belonging to that class.

Application of the resultant RP model following initial virtual screening makes it possible to efficiently decrease the number of candidate compounds for testing in the experimental assay.

**Table 2**

Statistical results produced from the final RP model

	# of compounds (%) <sup>a</sup>	Class%ObsCorr <sup>b</sup>	Overall%PredCorr <sup>c</sup>	Enrichment <sup>d</sup>
Inactive class	185 (54.25)	90.27	84.34	1.55
Active class	156 (45.75)	80.13	87.41	1.91

<sup>a</sup> The number of samples in each class.

<sup>b</sup> Interclass prediction.

<sup>c</sup> Overall prediction.

<sup>d</sup> The enrichment factor: Overall%PredCorr divided by the original percentage of compounds belonging to that class (%)

Figure 5 displays the optimized RP model by the encouraged descriptors to extract active Aurora-A inhibitors. First descriptor to split compounds is S<sub>ssCH2</sub>. At the second step the descriptor Atype N 73 splits the active compounds and ADME\_Absorption\_T2\_2D splits the inactive compounds. E-state key values, one topological index and other descriptors were required to achieve the observed performance. The first key descriptor is the electrotopological value (E-state key) computed for each atom in a molecule which encodes information about both the topological environment of that atom and the electronic interactions due to all of the other atoms in the molecule. That is, the information on the electron accessibility at the atom and the degree of adjacency or topological state of the atom was provided by the E-state key. The meaning of the E-state symbols in the Cerius<sup>2</sup> implementation is as follows: S, sum of numerical value for following atom type; s, single bond; d, double bond; t, triple bond; and a, aromatic bond. Here S<sub>ssCH2</sub> stands for the sum of intrinsic values for the –CH<sub>2</sub>– atom type with two single bonds. The other main topological descriptor used to split the data set is the Zagreb, which characterizes the degree of atomic branching in a molecule. These 2D descriptors are beneficial in that the 3D shape query directed from the compounds is more restricted. An explanation corresponding to each descriptor is provided in Table 3.

### 3.3. Virtual screening using an external library

All the databases contained 1,047,671 compounds as mentioned in method materials section. To satisfy the requirement of pharmacophore-based virtual screening, 2D structures (in SD format) of all small molecules in the dataset were converted into 3D structures with multi-conformers (in Catalyst 3D format) using the 'Build 3D Database' protocol in Discovery Studio 2.5; the maximum number of conformers generated for each molecule was set to 250. The common feature pharmacophore was applied as the first filter to screen compounds according to higher fit values (above 4.0 out of maximum 5.0) and gave 3863 hits. Second, the initial hit compounds were screened using three filtering processes/methods. The reason for applying various criteria was to use ligand- and receptor-based approaches for the final selection independently and to extract compounds of diverse scaffolds, which can be modified to get better activity as well as selectivity. Following this scheme can make a difference in a way that the ranking of compounds are different for the models developed in different approaches. In one process 3863 hits were screened by docking to select compounds based on their interactions with the receptor active site. Forty-nine compounds were selected by Glide docking score (G-score). The second filter was the selection of compounds from the pharmacophore hits (having top fit value)

**Table 3**

Summary of the descriptors that were found to be useful in decision making in the RP model

Descriptor	Illustration
S <sub>ssCH2</sub>	Sum descriptor for carbon with two single bonds
Atype N 73	Ghosh and Crippen atom type descriptor
Zagreb	The Zagreb index is defined as the sum of the squares of vertex valencies
S <sub>ssNH</sub>	Sum descriptor for nitrogen with two single bonds
Atype C 32	Ghosh and Crippen atom type descriptor
S <sub>aaN</sub>	Sum descriptor for nitrogen with two aromatic bonds
S <sub>dsN</sub>	Sum descriptor for nitrogen with one double and one single bond
ADME_Absorption_T2_2D	Multivariate distance among nitrogen, oxygen atoms and attached hydrogen atoms from van Der Waals polar surface area

by considering the diversity of the scaffold as well as the mapping of features. Number of hetero atom, aromatic/non-aromatic ring,  $n$  ( $n = 5, 6, 7$ )-member ring, polycyclic ring, alkyl substituent and their position etc. were observed to determine diversity of

compounds. Four compounds were selected by following the criteria of the second filter. This filtering method is helpful in identifying compounds with different functionalities but conserving the similar properties to targeted hits. In the third method, 3863 hits

**Table 4**

Biological test results of virtual screening hits for Aurora-A

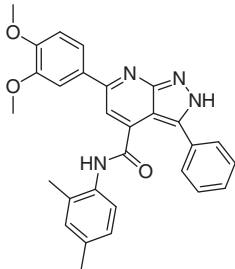
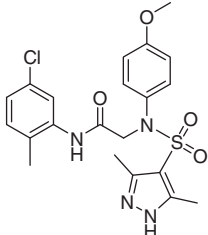
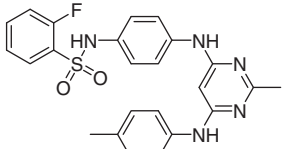
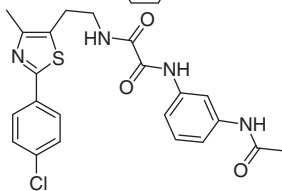
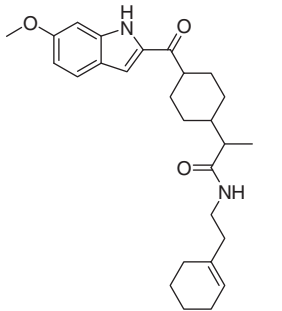
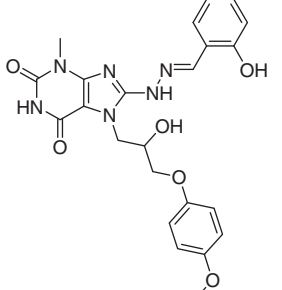
No.	ID	Structure	% Inhibition @ 10 $\mu$ M		GI <sub>50</sub> ( $\mu$ M)		IC <sub>50</sub> ( $\mu$ M)
			DU 145	HT-29	DU 145	HT-29	
C1	C202-1930		92	91	2.06	1.65	>100
C2	C517-2449		74	87	5.98	3.98	NA
C3	D430-1310		38	83	23.42	5.12	NA
C4	G569-0267		32	54	>100	16.47	>100
C5	L075-0758		49	68	15.86	9.36	86.15
C6	BAS 0039 0218		23	58	27.73	7.94	18.4

Table 4 (continued)

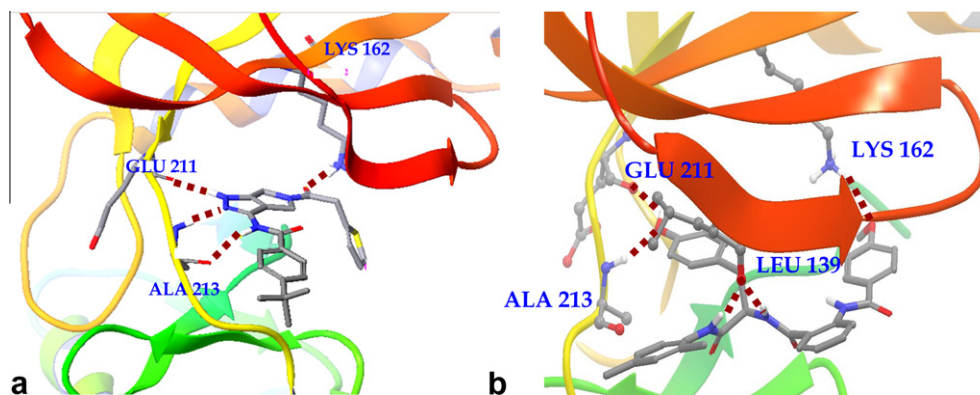
No.	ID	Structure	% Inhibition @ 10 $\mu$ M		GI <sub>50</sub> ( $\mu$ M)		IC <sub>50</sub> ( $\mu$ M)
			DU 145	HT-29	DU 145	HT-29	
C7	BAS 0041 6436		80	89	36.89	29.14	16.72
C8	BAS 0102 4712		87	54	3.93	9.48	11.94
C9	BAS 0345 0373		84	85	4.26	7.08	NA
C10	BAS 0357 1837		57	62	>100	7.91	98.7
C11	BAS 0102 4719		90	73	2.73	5.23	5.09
C12	BAS 1536 4720		49	89	17	2.04	>100
C13	ASN 0377 6845		49	78	14.06	6.33	NA

(continued on next page)

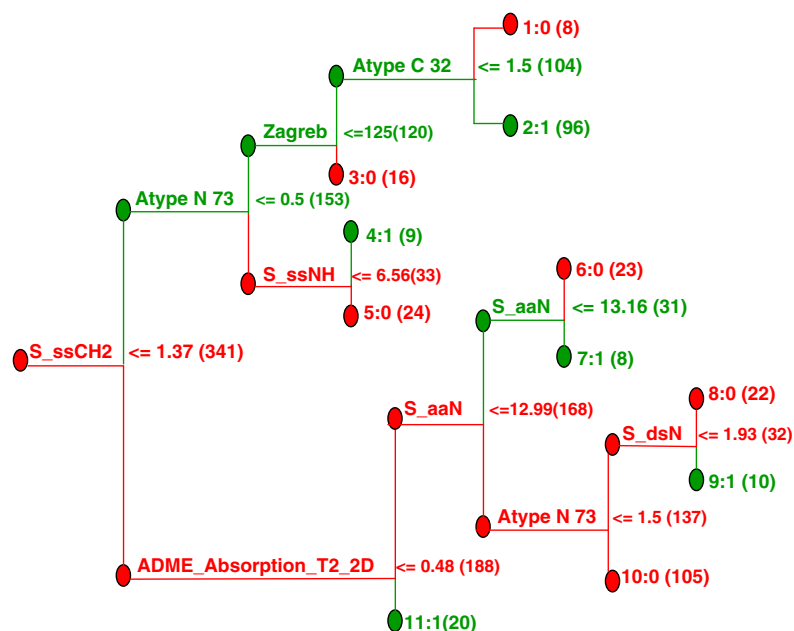
**Table 4** (continued)

No.	ID	Structure	% Inhibition @ 10 $\mu$ M		GI <sub>50</sub> ( $\mu$ M)		IC <sub>50</sub> ( $\mu$ M)
			DU 145	HT-29	DU 145	HT-29	
C14	ASN 0741 0292		80	88	5.81	3.64	NA
C15	E018-1724		34	64	26	5.36	NA

NA indicates no inhibition or compound activity could not be fit to an IC<sub>50</sub> curve.

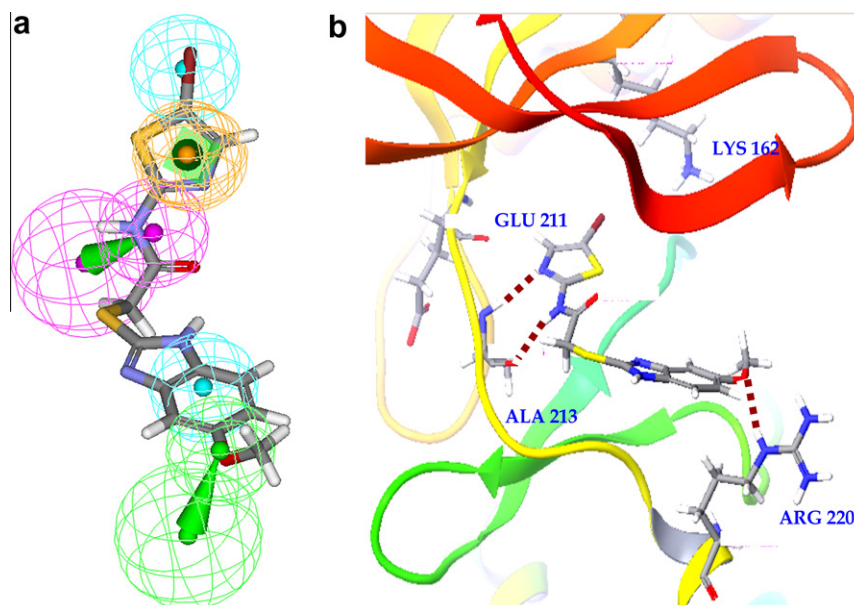


**Figure 4.** (a) Docking pose of the hiphop training set compound (5) in the active site of Aurora-A crystal structure. (b) Interactions with the compound (C11) from screening in the active site.

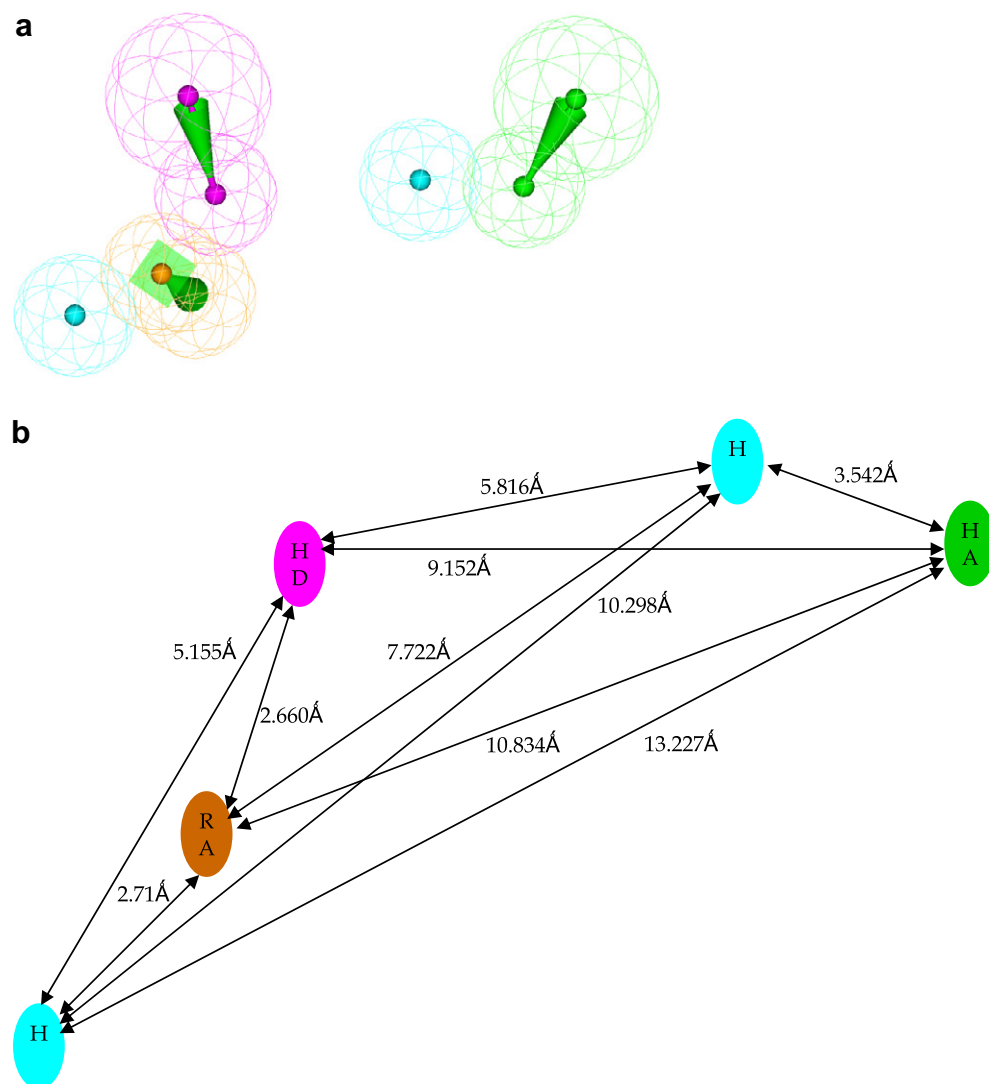


**Figure 5.** RP tree generated to discriminate active and inactive compounds. At each node (decision point), molecules were split into higher or lower responses according to the descriptors. The green color indicates active and red color for inactive class.





**Figure 6.** (a) Mapping of pharmacophore features (cyan: hydrophobic groups; green: hydrogen bond acceptor feature with a vector in the direction of the putative hydrogen donor; orange: aromatic ring; purple: hydrogen bond donor feature with a vector in the direction of the putative hydrogen acceptor) to the virtual screening hit compound (C12). (b) Docking pose of the hit compound (C12) in the active site of Aurora-A.



**Figure 7.** (a) The features of pharmacophore model D. (b) Distances between pharmacophore sites.

were put into the RP model as prediction set, and only 529 of the hit compounds were predicted to have an increased probability of being active, thereby eliminating a large number of compounds predicted to be inactive. We selected fifteen compounds by visual inspection from the total actives obtained by this method. Visual inspection focused on the number of hetero atom, hydrogen bond donor, acceptor, different types of aromatic, non-aromatic ring and alkyl group. Finally, a total of 68 compounds were purchased for biological testing.

Several hit compounds having diverse structure and in vitro activity (percent inhibition at 10  $\mu$ M of test compounds) against different human cell lines are shown in Table 4. Based on the percent inhibition values, fifteen compounds were selected for further testing and their  $GI_{50}$  values were determined (Table 4). Kinase assay was carried out for all of these compounds to determine  $IC_{50}$  value. Some compounds with average cell line activity had acceptable  $IC_{50}$  value (Table 4). For example, compound **C8** has  $IC_{50}$  value 11.94  $\mu$ M, whereas,  $GI_{50}$  values for DU 145 and HT-29 are 3.93 and 9.48  $\mu$ M respectively. On the other hand, compounds having high  $GI_{50}$  value did not show satisfactory result for  $IC_{50}$ . The hit compound **C12** is visualized in its respective pharmacophore model and the most generically stable docking pose in Figure 6. It fulfills the best pharmacophore, by perfectly matching one hydrogen bond acceptor and one donor. The docking pose and conformation of the compound aligns well with the results from the ligand-based approach. But it had  $IC_{50}$  value more than 100  $\mu$ M. The most active compound **C11** among the hits from the virtual screening has hydroxyl and methoxy group at the terminal position in its scaffold. The docking pose of this compound (Fig. 4b) in the active site reveals the interaction pattern. To compare the binding mode, the inhibitor **5** (Fig. 3) was docked into the active site of crystal structure (2bmc.pdb) using Glide. Two hydrogen bonds were formed with Ala 213 (Fig. 4a). One is a hydrogen bond donor from the amide hydrogen, and the other is an acceptor from the pyrazolo nitrogen. The second important hydrogen bond was formed between the hinge residue Glu 211 and the pyrazole ring hydrogen. An additional hydrogen bond was observed with the residue Lys162. The compound **C11** makes two hydrogen bonds by hydroxyl group with Glu 211 and Ala 213. The methoxy group forms other hydrogen bond with Lys 162. Two additional hydrogen bonds are formed with oxygen of Leu 139. It is clearly shown that the hit compound identified by this virtual screening scheme has the hydrogen bonds necessary for activity. It gives opportunity for further optimization to identify more active compound.

#### 4. Conclusion

In this work, we described the development of common feature-based pharmacophore models for Aurora-A inhibitors. Furthermore, we reported the virtual screening of several chemical databases using our best pharmacophore model, recursive partitioning model, docking study and visual inspection. This virtual screening workflow resulted in the selection of 68 compounds for biological testing. Based on the percent inhibition values, fifteen compounds were selected and the activity was evaluated by

measuring a  $GI_{50}$  value. Kinase assay result of these compounds identified a hit compound with  $IC_{50}$  value of 5.09  $\mu$ M. All the hits selected by the biological assay result have diverse scaffolds, which provide scope for modification to achieve better activity as well as selectivity.

#### Acknowledgments

This work was supported by Korea Institute of Science and Technology (KIST).

#### Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmc.2010.11.064.

#### References and notes

- Fu, J. Y.; Bian, M. L.; Jiang, Q.; Zhang, C. M. *Mol. Cancer Res.* **2007**, 5, 1.
- Zhou, H. Y.; Kuang, J.; Zhong, L.; Kuo, W. L.; Gray, J. W.; Sahin, A.; Brinkley, B. R.; Sen, S. *Nat. Genet.* **1998**, 20, 189.
- Bischoff, J. R.; Anderson, L.; Zhu, Y. F.; Mossie, K.; Ng, L.; Souza, B.; Schryver, B.; Flanagan, P.; Clairvoyant, F.; Ginther, C.; Chan, C. S. M.; Novotny, M.; Slamon, D. J.; Plowman, G. D. *EMBO J.* **1998**, 17, 3052.
- Meraldi, P.; Honda, R.; Nigg, E. A. *EMBO J.* **2002**, 21, 483.
- Wang, X.; Zhou, Y. X.; Qiao, W.; Tominaga, Y.; Ouchi, M.; Ouchi, T.; Deng, C. X. *Oncogene* **2006**, 25, 7148.
- Marumoto, T.; Hirota, T.; Morisaki, T.; Kunitoku, N.; Zhang, D. W.; Ichikawa, Y.; Sasayama, T.; Kuninaka, S.; Mimori, T.; Tamaki, N.; Kimura, M.; Okano, Y.; Saya, H. *Genes Cells* **2002**, 7, 1173.
- Jiang, Y.; Zhang, Y. K.; Lees, E.; Seghezzi, W. *Oncogene* **2003**, 22, 8293.
- Anand, S.; Penrhyn-Lowe, S.; Venkitaraman, A. R. *Cancer Cell* **2003**, 3, 51.
- Wang, X. X.; Liu, R.; Jin, S. Q.; Fan, F. Y.; Zhan, Q. M. *Cell Res.* **2006**, 16, 356.
- Cheetham, G. M. T.; Charlton, P. A.; Golec, J. M. C.; Pollard, J. R. *Cancer Lett.* **2007**, 251, 323.
- Harrington, E. A.; Bebbington, D.; Moore, J.; Rasmussen, R. K.; Ajose-Adeogun, A. O.; Nakayama, T.; Graham, J. A.; Demur, C.; Hercend, T.; Diu-Hercend, A.; Su, M.; Golec, J. M. C.; Miller, K. M. *Nat. Med.* **2004**, 10, 262.
- Fancelli, D.; Moll, J.; Varasi, M.; Bravo, R.; Artico, R.; Berta, D.; Bindi, S.; Cameron, A.; Candiani, I.; Cappella, P.; Carpinelli, P.; Croci, W.; Forte, B.; Giorgini, M. L.; Klapwijk, J.; Marsiglio, A.; Pesenti, E.; Rocchetti, M.; Roletto, F.; Severino, D.; Soncini, C.; Storici, P.; Tonani, R.; Zugnoni, P.; Vianello, P. *J. Med. Chem.* **2006**, 49, 7247.
- Manfredi, M. G.; Ecsedy, J. A.; Meetze, K. A.; Balani, S. K.; Burenkova, O.; Chen, W.; Galvin, K. M.; Hoar, K. M.; Huck, J. J.; LeRoy, P. J.; Ray, E. T.; Sells, T. B.; Stringer, B.; Stroud, S. G.; Vos, T. J.; Weatherhead, G. S.; Wysong, D. R.; Zhang, M. K.; Bolen, J. B.; Claiborne, C. F. P. *Natl. Acad. Sci. U.S.A.* **2007**, 104, 4106.
- Mortlock, A. A.; Foote, K. M.; Heron, N. M.; Jung, F. H.; Pasquet, G.; Lohmann, J. J. M.; Warin, N.; Renaud, F.; De Savi, C.; Roberts, N. J.; Johnson, T.; Dousson, C. B.; Hill, G. B.; Perkins, D.; Hatter, G.; Wilkinson, R. W.; Wedge, S. R.; Heaton, S. P.; Odedra, R.; Keen, N. J.; Crafter, C.; Brown, E.; Thompson, K.; Brightwell, S.; Khatri, L.; Brady, M. C.; Kearney, S.; McKillop, D.; Rhead, S.; Parry, T.; Green, S. J. *Med. Chem.* **2007**, 50, 2213.
- Wilkinson, R. W.; Odedra, R.; Heaton, S. P.; Wedge, S. R.; Keen, N. J.; Crafter, C.; Foster, J. R.; Brady, M. C.; Bigley, A.; Brown, E.; Byth, K. F.; Barrass, N. C.; Mundt, K. E.; Foote, K. M.; Heron, N. M.; Jung, F. H.; Mortlock, A. A.; Boyle, F. T.; Green, S. *Clin. Cancer Res.* **2007**, 13, 3682.
- Vianello, P. *Expert Opin. Ther. Pat.* **2007**, 17, 255.
- Moriarty, K. J.; Koblish, H. K.; Garrabrant, T.; Maisuria, J.; Khalil, E.; Ali, F.; Petrounia, L. P.; Crysler, C. S.; Maroney, A. C.; Johnson, D. L.; Gallemmo, R. A. *Bioorg. Med. Chem. Lett.* **2006**, 16, 5778.
- Fancelli, D.; Berta, D.; Bindi, S.; Cameron, A.; Cappella, P.; Carpinelli, P.; Catana, C.; Forte, B.; Giordano, P.; Giorgini, M. L.; Mantegani, S.; Marsiglio, A.; Meroni, M.; Moll, J.; Pittala, V.; Roletto, F.; Severino, D.; Soncini, C.; Storici, P.; Tonani, R.; Varasi, M.; Vulpetti, A.; Vianello, P. *J. Med. Chem.* **2005**, 48, 3080.
- Smellie, A.; Kahn, S. D.; Teig, S. L. *J. Chem. Inf. Comput. Sci.* **1995**, 35, 285.