Identification of critical chemical features for Aurora kinase-B inhibitors using Hip-Hop, virtual screening and molecular docking

Sugunadevi Sakkiah, Sundarapandian Thangapandian, Shalini John, Keun Woo Lee*

Division of Applied Life Science (BK21 Program), Environmental Biotechnology National Core Research Center (EB-NCRC), Plant Molecular Biology and Biotechnology Research Center (PMBBRC), Gyeongsang National University (GNU), 900 Gazwe-dong, Jinju 660-701, Republic of Korea

A R T I C L E   I N F O

Article history:
Received 26 July 2010
Received in revised form 31 August 2010
Accepted 31 August 2010
Available online 30 November 2010

Keywords:
Aurora kinase
Pharmacophore
Molecular docking
Hip-Hop
Homology modeling
Molecular dynamic simulation

A B S T R A C T

This study was performed to find the selective chemical features for Aurora kinase-B inhibitors using the potent methods like Hip-Hop, virtual screening, homology modeling, molecular dynamics and docking. The best hypothesis, Hypo1 was validated toward a wide range of test set containing the selective inhibitors of Aurora kinase-B. Homology modeling and molecular dynamics studies were carried out to perform the molecular docking studies. The best hypothesis Hypo1 was used as a 3D query to screen the chemical databases. The screened molecules from the databases were sorted based on ADME and drug like properties. The selective hit compounds were docked and the hydrogen bond interactions with the critical amino acids present in Aurora kinase-B were compared with the chemical features present in the Hypo1. Finally, we suggest that the chemical features present in the Hypo1 are vital for a molecule to inhibit the Aurora kinase-B activity.

© 2010 Published by Elsevier B.V.

1. Introduction

Aurora kinases play an important role in cell cycle which belongs to the serine/threonine kinase family and it is one of the potential targets in chemogenomics approaches [1]. In recent years Aurora kinases have been emerged as an important drug targets in several pharmaceutical companies and research industries since it plays a major role in regulating mitosis and cytokinesis [2]. Mitosis is a vital process for the regeneration of tissues and the genome, development of an individual, and functional integrity of a cell [3]. Three different types of Aurora kinases were reported for mammals, designated as Aurora kinases A, B and C, which all shares an high homologous of amino acid sequence but with distinct functions: (i) Aurora kinase-A in “polar kinase”, primarily associated with the centrosomes separation [4,5] (ii) Aurora kinase-B in “equatorial kinase”, is a chromosomal passenger protein [6] and (iii) Aurora kinase-C appears in the centrosome from anaphase to telophase and is highly expressed in the tests. All three kinases influence the cell cycle from its G2 phase through cytokinesis but it appears at specific locations during mitosis. Abnormalities of Aurora kinases have strong link with cancer and guide to the current growth of new classes of anti-cancer drugs which specifically target the ATP-binding domain of Aurora kinases. Aurora kinase-B and Aurora kinase-A (Aurora B/A) are over expressed in human tumors [7,8], inhibition of these kinases can lead to antitumor effects, hence it has an considerable interest in developing specific and novel anti-cancer drugs to achieve the selectivity between Aurora B/A [2]. Therefore, currently researchers express a great interest to design selective inhibitors for Aurora kinases as targets for therapeutic intervention, in particular for the treatment of cancer.

Our current approach focuses onto develop a pharmacophore model which can able to distinguish the Aurora kinase-B inhibitors from Aurora kinase-A. In this work we did not focuses on Aurora kinase-C because there is no any clear view about the function and structure of this protein is not reported. In order to find the selective 3D pharmacophoric features for Aurora kinase-B inhibitors, Hip-Hop module of Discover Studio v2.1 (DS, www.accelrys.com) was used to generate the hypothesis, the generated models are validated using cluster analysis and test set containing structurally diverse as well as the selective inhibitors of Aurora B/A. The resultant pharmacophore model from this study can be used as a computational tool in order to select or distinguish the Aurora kinase-B from Aurora kinase-A inhibitors.

1.1. Structure of Aurora kinase

Aurora kinases share a common sequence and structure, consisting of a highly conserved C-terminal catalytic domain and a short N-terminal domain which varies in size [9]. The N-terminal domain of Aurora A, B and C kinases have limited sequence similar-
ity and the catalytic domain of Aurora-family kinases contain short C-terminal tail sequence [10], which may aid proteosome degradation of the enzymes at the end of mitosis. The C-terminal of all Aurora-family kinases contains a conserved amino acid sequence, located in a structured loop region at the surface of the kinase domain, approximately 30 Å from the catalytic active site [11]. The active site cleft is bounded by the glycine-rich loop contains the consensus kinase sequence Gly-X-Gly-X-Gly and the activation loop. The catalytic domain of Aurora kinase-A has the typical bi-lobe kinase fold comprised of N-terminal β-strand and C-terminal α-helical domains and these domains are linked by hinge region, which plays an important role in forming the catalytic active site [11]. In Aurora kinase-B, N-terminal lobe is rich in β-strands, which implicated in nucleotide binding, interact with kinase regulators and the C-terminal lobe is mainly α-helical, serves as a docking site for substrates. In both kinases, the ATP binding pocket was present at the interface of these two domains (N- and C-terminal). Aurora kinase-A appears to play an important role during chromatins separation and formation of the cleavage furrow in anaphase and telophase [12]. Aurora kinase-B activity is maximal during prophase and the role of Aurora kinase-C is less clear.

1.2. A valid target in cancer therapeutics – Aurora kinases

Aurora kinase-B was one potential targets in the cancer research field. Aurora B/A mapped to a region in human chromosome 20q13.2-13.3 and 17p13, respectively, amplified in cancer cell lines and primary tumors. Expression levels of Aurora B/A kinases were elevated in several cancer cells relative to normal cells. Thus, the inhibition of one or more Aurora kinases might be a novel chemotherapeutic strategy against cancer.

Aurora kinase-A (also known as Aurora-2, IAK1) which is localize at the centrosome of the mitotic spindle pole and proposed to function in late anaphase, promoting spindle elongation and centrosome separation. Aurora kinase-A activity is regulated by a protein called TPX2, required for spindle assembly. Dysregulation of Aurora kinase-A is thought to be oncogenic and result in the production of multiple centrosomes and aneuploidy [13,14]. The oncogenicity of Aurora kinase-A results from two distinct functions of the protein, namely (a) chromosome segregation and control of genomic stability (b) regulation of entrance into mitosis.

Aurora kinase-B (also known as Aurora-1, AIM1) is a chromosomal passenger protein which phosphorylates serine 10 of histone H3. Inhibition of Aurora kinase-B activity induces anti-proliferative phenotype, indicating that Aurora kinase-B is an attractive anti-cancer drug target [9]. Sequence and structural similarities among these three human Aurora kinases suggest that it will be a challenge to design small molecules with absolute specificity for a single Aurora kinase family member. Inhibition of Aurora B/A yields distinct phenotypes; hence it may present two avenues for anti-cancer drug discovery [15]. These recent findings implicate Aurora B/A kinases as an important regulator of both genomic integrity and cell cycle progression in cancer cells. Thus Aurora-A and Aurora-B kinases is an attractive targets for anti-cancer drug development. Despite the overall conservation, there are several unique features of the active site that can be exploited for designing selective inhibitors of Aurora kinase-B.

2. Methods and materials

2.1. Training and test sets preparation

Four known inhibitors were selected as a training set [16–18] (Fig. 1) to generate a common feature pharmacophore (Hip-Hop) model for Aurora kinase-B. The chemical features for the generation of the hypothesis were selected by using the feature mapping/DS. For hypothesis generation, initially the 2D structure of all molecules were built using MDL ISIS Draw v2.5 and imported into Discovery Studio v2.1 (DS, www.accelrys.com Accelrys, San Diego, CA, USA) for 3D conversion and minimization. Maximum of 255 conformations were generated for each compound using the Best Conformation model generation method based on CHARMm force field applying Poling algorithm. To ensure the energy-minimized conformer of the molecules, the conformations with energy

![Fig. 1. Structure of Aurora kinase-B antagonist used as training set, activity value (IC50) was indicated in bracket.](image-url)
more than 20 kcal/mol from the global minimum were rejected and molecules with their lowest energy conformations were submitted to DS for hypothesis generation [19,20]. The generated molecular conformations, not only used for hypothesis generation but also to find how well the molecules were able to fit into the hypothesis. In order to validate the generated hypothesis, test set was prepared which comprises of 19 structurally diverse molecules and selectivity profiles (good IC50 values) to Aurora B/A as well as these compounds are not included in training set. The test set compounds were prepared using the same protocol as that of the training set.

2.2. Generation of pharmacophore models

In the past years, ligand based pharmacophore modeling programs clearly demonstrated that many successfulchronicles in the field of medicinal chemistry. Hip-Hop, one of the powerful tools widely used in ligand-based approach and rational drug design, produce the hypothesis based on the common features present in the training set molecules. Hip-Hop generation follow two strategies, first one assume that all compounds are important and contain significant features, furthermore, differences in activities are related to the differences in other relevant factors like conformational energies, but not due to the absence of any important features required for binding. In contrast, the hypothesis generation in the second strategy biased to the most active compounds, assuming that they contain important features [21].

Common feature hypothesis for Aurora B was generated by implementing Hip-Hop algorithm in order to find the common chemical features shared by a set of selective inhibitors of Aurora kinase-B. The qualitative pharmacophore generation using Hip-Hop/DS is performing in three-steps [22]: (i) Conformation models are generated for each molecule in the training set (ii) Each conformer is examined for the presence of certain chemical features (iii) Three-dimensional configuration of chemical features common to the input molecules is determined. DS provides a dictionary of chemical features which are important in drug-enzyme/receptor interactions that includes hydrogen bond acceptor (HBA), hydrogen bond donor (HBD), hydrophobic (HYD), ring aromatic (RA), positive ionizable (PI) and negative ionizable (NI) groups. Different types of chemical features for each molecule in the training set were examined using Feature Mapping module from DS. By consolidating all the available features, four common features were used to generate the hypothesis: HBA, HBD, HYD and RA.

In the Hip-Hop modeling, the highest weight was assigned by assigning a value of 2 (ensures that all the chemical features in the compound will be considered in pharmacophore generation) and 0 (force the compounds to map all features) in the principal and maximum omitting features columns, respectively, for the most active compound. A value of 1 for the principle and maximum omitting feature columns ensure that at least one mapping for each of generated hypotheses will be found and assure that all but one feature can miss from the input parameters, respectively. All other parameters were kept at the default settings. Completion of the Hip-Hop run gives ten common feature hypotheses based on the ranking score corresponding with its critical chemical features. The ranking score for each hypothesis was calculated based on a ranking formula and the default definition of the “FIT” of a molecule to the hypothesis. The highest score represents that the molecules in the training set fit the hypothesis not by a chance.

2.3. Selection of the best hypothesis

Some of the strategies adopted to select a best pharmacophore model of Aurora kinase-B from the ten generated hypotheses: First, clustering analysis was performed, to categorize the obtained ten pharmacophore models into several clusters based on their ranking score and composition similarities. Second, the training set was used to validate the lustiness and selectivity of top ranked hypothesis from each cluster using the “Ligand Pharmacophore Mapping”. When a compound is mapped to a hypothesis, the quality of the mapping was determined by the “fit-value” and this value is defined as the weight (f) X [1-SSE (f)], where f is the mapping features, SSE (f) is the sum over location constraints c on f of [D(c)/ T(c)]2, D is the displacement of the feature from the center of the location constraint, and T (tolerance) is the radius of the location constraint sphere for the feature. The maximum fit value for a perfectly fitting compound is the sum of the weight values for all features and the minimum value should be 0. In this study, the fit value of every molecule mapped to pharmacophore models were calculated and the hypothesis was selected which establishes a highest fit value from the training set compounds. Third, using the test set which includes the selective inhibitors of Aurora B/A to identify the best hypothesis that can precisely distinguish Aurora kinase-B from Aurora kinase-A inhibitors as well as to evaluate the potential selectivity of the obtained hypothesis. Finally, we assessed the mapping situation of the best fit molecules on the selected hypothesis to figure out the key features that contributed the selectivity of the best hypothesis model.

2.4. Database screening

Virtual screening of database serve for two main purposes: (i) quality of the generated hypothesis was validated by selective detections of compounds with known inhibitory activity value and (ii) to find the novel, potential leads suitable for further drug development. The best pharmacophore model was used as 3D query to screen the Maybridge and Chembridge chemical databases consisting of 60,000 and 50,000 structurally diversified molecules, to retrieve new compounds which could be a selective and novel scaffold of Aurora kinase-B inhibitors and sorted based on the highest fit value. The ADMET functionality implemented in DS, estimate the values of Blood Brain Barrier (BBB) penetration, solubility, Cytochrome P450 (CYP450) 2D6 inhibition, Hepatotoxicity, Human intestinal adsorption (HIA), Plasma Protein Binding (PPB) and access a broad range of toxicity measure of the ligands. Among all the criteria’s we mainly focused on BBB, solubility, and HIA, the cut off value was 3, 3 and 0, respectively. In general, when the molecules posses Log P less than 5, number of hydrogen bond acceptor less than 10, number of hydrogen bond donor less than 5, molecular weight less than 500 and the number of rotatable bond should not exceeding 10 [23]. These are some of the important criteria for a compound to be a good oral bioavailability drug. Hence the sorted molecules were further filtered based ADME (Absorption, Distribution, Metabolism and Excretion) and Lipinski’s rule of five [24] properties which were calculated using ADMET and Prepare Ligand module implemented in DS, respectively. Finally, these sorted molecules were subjected for molecular docking studies to find the suitable binding orientation of the molecules in active site of the Aurora kinase-B receptor.

2.5. Comparative modeling: 3D structure generation of Aurora kinase-B

Comparative modeling (homology modeling) is one of the most accurate computational methods to generate reliable tertiary protein structure from its sequence and is routinely used in many biological applications. Recently, there are four crystal structures of Xenopus laevis Aurora kinase-B were deposited in Protein Data Bank (PDB, www.rcsb.org) which shows high similarity and identity with human Aurora kinase-B, but till date, there is not any X-ray crystal structure (3D-structure) of human Aurora kinase-B.
Hence, the homology model for human Aurora kinase-B was carried out to find its tertiary structure. The identity and similarity between the target and template protein determine the quality of the homology model structure. Human Aurora kinase-B primary sequence was retrieved from Swiss-Prot Protein Database (Accession ID: Q96GD4) which has 344 amino acids. To find a suitable template for human Aurora kinase-B a similarity search against PDB was performed using BLAST [http://www.ncbi.nlm.nih.gov] server Aurora kinase-B from Xenopus laevis (PDB ID: 2VXR, Resolution: 1.86 Å) was selected as a best template to construct human Aurora kinase-B using MODELLER algorithm in DS. The final model was checked using the PROCHECK program [25,26], to search for deviations from normal protein conformational parameters.

2.6. Refinement of homology model using molecular dynamics simulation

Molecular dynamics (MD) simulation was performed to refine the side chain orientations and also to gain a better relaxation as well as more correct arrangement of the atoms in Aurora kinase-B model. The GROningen MACHine for Chemical Simulations (GROMACS) V3.3 package [27,28] was used to solvate a model in a cubic the side chain orientations and also to gain a better relaxation as well as more correct arrangement of the atoms in Aurora kinase-B model. The GROningen MACHine for Chemical Simulations (GROMACS) V3.3 package [27,28] was used to solvate a model in a cubic box of dimension 1 nm by applying GROMACS force field [29]. The SPC water model was used in order to create the aqueous environment. Particles mesh Ewald (PME) [30] electrostatic and periodic boundary conditions were applied in all directions. The system was neutralized by adding eight Cl− counter ions by replacing eight water molecules. It was subjected to a steepest descent energy minimization until a tolerance of 1000 kJ/mol reached, therefore the system can be get rid of the high energy interactions and steric clashes. All the bond lengths were constrained with [31] the LINCS method the energy minimized system was treated for 100 ps equilibration run. The pre-equilibrated system was consequently subjected to 5 ns production MD simulation, with a time-step of 2 fs at constant temperature (300 K), pressure (1 atm) and without any position restraints [32]. Snapshots were collected every 5 ps and all the analyses of the MD simulation were carried out by GROMACS analysis tools. From the 5 ns MD simulation, average structure with low RMSD value was selected as the best model for molecular docking studies.

2.7. Molecular docking protocol

In computer aided drug design process, pharmacophore based molecular docking was one of the most reputable method which was used to find the accuracy of binding orientation (poses) of the ligands into the protein active site. In this paper, LigandFit module was used to find the suitable orientation of the molecules in the active of Aurora kinase-B. The quality of the receptor structure plays a central role in determining the success of docking calculations [33,34]. The model Aurora kinase-B was used as a receptor protein for molecular docking studies. There are three stages in LigandFit protocol: (i) Docking: attempt is made to dock a ligand into a user defined binding site (ii) In-Situ Ligand Minimization and (iii) Scoring: various scoring functions were calculated for each pose of the ligands. For docking study, initially protein was prepared by removing water molecules and hydrogen was added based on the CHARMM force field.

After the protein preparation, the active site of the protein has to be identified for docking the small molecules. The active site of the protein can be represent as binding site; it is a set of points on a grid that lie in a cavity. Two methods are available to define a protein binding site: (i) identify site based on the shape of the receptor using “eraser” algorithm and (ii) volume occupied by the known ligand pose already in an active site. For this study, we preferred the second method to find the active cavity of Aurora kinase-B (Homology model) by comparing the inhibitor present in the template structure. In order to check whether the LigandFit and the selected parameters are suitable for Aurora kinase-B docking studies, we sketched one of the co-crystal structures and docked with its receptor and the RMSD was calculated between the docked pose and the bound conformation of the small molecules (co-crystal). From this we can say that the selected parameters are worthy for producing the suitable orientation of the small molecules in the active site of the protein. Then the specific Aurora kinase-B inhibitors and hit molecules were docked into the active site of Aurora kinase-B. During the docking process top 10 ligand conformations for each compound was saved based on the best orientation of the molecules in the active site of the protein as well as the highest dock score values which was calculated after the energy minimization of each molecules using the smart minimizer method, begins with steepest descent method followed by the conjugate gradient method. Single docking score may be fails to fetch the active molecules hence the consensus scoring method was applied which consist of LigScore1, LigScore2, PMP, PLP1, PLP2. Post-docking filter based on H-bond network was constructed to distinguish between active and inactive compounds. The molecules were selected based on the consensus scoring method which was subjected to visualization process to find whether these molecules are able to form all the possible hydrogen bond (H-bond) interactions with target based on the occupancy of the ligands in the space close to Ala157 for Aurora kinase-B.

3. Results and discussions

The accurate prediction of binding affinities and biochemical activities of the inhibitors are one of the major challenges in computational drug design approaches. Ligand and receptor based approaches are used to determine structure–activity relationship and also to develop the new compounds with greater activity and a better selectivity index for a specific target [35].

3.1. Selection of pharmacophore features for Aurora kinase-B

Based on the training set molecules (Fig. 1), top 10 common feature hypotheses (Hypo1–Hypo10) were generated which have shown the ranking scores from 35.410 to 25.042 by Hip-Hop program. Direct hit, partial hit mask value of ‘1’ and ‘0’ for all hypotheses indicates that all the molecules present in the training set are able to map all features present in the hypothesis and there is no partial mapping or missing of any chemical features in the hypothesis, specified in Table 1. The top ranked hypothesis Hypo1, consist

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Features*</th>
<th>Rankingb</th>
<th>Direct hitc</th>
<th>Partial hitd</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypo1</td>
<td>RA, HYD, HBA, HBA</td>
<td>35.410</td>
<td>1111</td>
<td>0000</td>
</tr>
<tr>
<td>Hypo2</td>
<td>R, HYD, HYD, HBA</td>
<td>32.914</td>
<td>1111</td>
<td>0000</td>
</tr>
<tr>
<td>Hypo3</td>
<td>RA, HYD, HYD, HBA</td>
<td>32.914</td>
<td>1111</td>
<td>0000</td>
</tr>
<tr>
<td>Hypo4</td>
<td>HYD, HYD, HBA, HBA</td>
<td>32.772</td>
<td>1111</td>
<td>0000</td>
</tr>
<tr>
<td>Hypo5</td>
<td>RA, HYD, HBA</td>
<td>31.757</td>
<td>1111</td>
<td>0000</td>
</tr>
<tr>
<td>Hypo6</td>
<td>HYD, HYD, HYD, HBA</td>
<td>29.704</td>
<td>1111</td>
<td>0000</td>
</tr>
<tr>
<td>Hypo7</td>
<td>RA, RA, HBA</td>
<td>26.177</td>
<td>1111</td>
<td>0000</td>
</tr>
<tr>
<td>Hypo8</td>
<td>HYD, HBA, HBA</td>
<td>25.675</td>
<td>1111</td>
<td>0000</td>
</tr>
<tr>
<td>Hypo9</td>
<td>RA, HYD, HBA</td>
<td>25.043</td>
<td>1111</td>
<td>0000</td>
</tr>
<tr>
<td>Hypo10</td>
<td>HYD, HBA, HBA</td>
<td>25.042</td>
<td>1111</td>
<td>0000</td>
</tr>
</tbody>
</table>

* RA = ring aromatic; HBA = hydrogen bond acceptor; HBD = hydrogen bond donor; HYD = hydrophobic.

b Higher the ranking score, lesser the probability of chance correlation. The best hypothesis shows the highest value.

c DH, PH indicates whether (1) or (0) a training set molecule mapped every feature of the hypothesis and mapped to all but one feature in the hypothesis. The numbers from (right to left) correspond to the compounds (from top to bottom).
of one **HYD**, one **RA** and two **HBA** features which was shown in Fig. 2, the remaining all hypotheses demonstrated a lesser score when compared with **Hypo1**. Comparing all the hypotheses, **Hypo1**–**Hypo6** and **Hypo7**–**Hypo10** show four and three chemical features pharmacophore, respectively. Except **Hypo7** all other hypotheses having one **HBA** and **HYD** features, which indicates the importance of this two functional groups in Aurora kinase-B.

### 3.2. Validation and best pharmacophore model selection for Aurora kinase-B

The difference between the chemical features and its locations as well as the composition can be evaluated and categorized using the Hypothesis clustering module using Catalyst v4.11 ([www.accelrys.com](http://www.accelrys.com)). This process will give different numbers of cluster, but in this paper we focused on the features similarities between hypotheses and picked three clusters as presented in Table 2. Cluster I contains four hypotheses (2-three and 2-four features hypotheses) which contains the following features: **RA, HYD**, and **HBA** groups and the ranking score ranges from 35.41 to 25.04. Cluster II includes five hypotheses (4-four and 1-three features hypotheses): **RA, HYD** and **HBA** features and the ranking score of these hypotheses from 31.75 to 29.70. Cluster III has only one three features hypothesis contains **RA** and **HBA** features with ranking score of 26.17.

Examine all three Clusters reveal that the four features hypotheses shows good ranking score (above 30) when compared with three features hypotheses whose ranking score was nearly 25, from this, we concluded that the four features pharmacophore will be the best hypothesis. Cluster III has only one 3-feature hypothesis as well as it misses **HYD** group, one of the important chemical feature to block the activity of Aurora kinase-B which was already mentioned, hence more attention was given to cluster I and II. All the three features hypotheses from Cluster I (**Hypo8** and **Hypo10**) and Cluster II (**Hypo5**, **Hypo9**, and **Hypo6**) were eliminated. In cluster II, two hypotheses (**Hypo3** and **Hypo2**) have shown the same arrangement and occupied the same location in 3D space because both features were retrieved from the same chemical functional groups of the training set and also the ranking score was same (32.914). Hence one hypothesis from Cluster II (**Hypo3**) and two hypotheses (**Hypo1** and **Hypo4**) from cluster I was consider for further validation, which was selected based on the ranking score and the number of features. As we mentioned already, one **HBA** and one **HYD** chemical groups are vital for the inhibition of Aurora kinase-B, when one **HYD** and one **RA** groups are added with two vital chemical features (**HBA**, **HYD**) the ranking score was 32.914 (**Hypo3**). The presence of **RA** group was changed to **HBA** (**Hypo4**) the ranking score value was slightly decrease (32.772). From the above analyzes it was clearly indicated that the presence of one **HBA** and **RA** groups play a major role in ranking score of the hypothesis. This was further confirmed, when **Hypo1** was

### Table 2

<table>
<thead>
<tr>
<th>Hypothesis&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Ranking score&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Features&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Cluster&lt;sup&gt;d&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hypo1</strong></td>
<td>35.410</td>
<td><strong>RA, HYD, HBA, HBA</strong></td>
<td>I</td>
</tr>
<tr>
<td><strong>Hypo4</strong></td>
<td>32.772</td>
<td><strong>HYD, HYD, HBA, HBA</strong></td>
<td>I</td>
</tr>
<tr>
<td><strong>Hypo8</strong></td>
<td>25.675</td>
<td><strong>HYD, HBA, HBA</strong></td>
<td>II</td>
</tr>
<tr>
<td><strong>Hypo10</strong></td>
<td>25.042</td>
<td><strong>HBD, HBA, HBA</strong></td>
<td>II</td>
</tr>
<tr>
<td><strong>Hypo5</strong></td>
<td>31.757</td>
<td><strong>RA, HYD, HBA</strong></td>
<td>II</td>
</tr>
<tr>
<td><strong>Hypo2</strong></td>
<td>32.914</td>
<td><strong>RA, HYD, HBA</strong></td>
<td>II</td>
</tr>
<tr>
<td><strong>Hypo9</strong></td>
<td>32.914</td>
<td><strong>RA, HYD, HBA</strong></td>
<td>II</td>
</tr>
<tr>
<td><strong>Hypo6</strong></td>
<td>29.704</td>
<td><strong>HYD, HYD, HBA</strong></td>
<td>II</td>
</tr>
<tr>
<td><strong>Hypo10</strong></td>
<td>26.177</td>
<td><strong>RA, RA, HBA</strong></td>
<td>III</td>
</tr>
</tbody>
</table>

<sup>a</sup> Numbers for the hypothesis are consistent with the numeration as obtained by the hypothesis generation.

<sup>b</sup> The higher the ranking score, the less likely is it that the molecules in the training set fit the hypothesis by a chance correlation. The best hypotheses have the highest ranks.

<sup>c</sup> **RA** = ring aromatic; **HYD** = hypophobic group; **HBA** = hydrogen bond acceptor group; **HBD** = hydrogen bond donor group.

<sup>d</sup> Cluster assembly is adopted from Catalyst’s “hypotheses clustering” analysis result based on the composition similarity between hypotheses.

---

**Fig. 2.** Geometric constraint of **Hypo1**. Green color indicates hydrogen bond acceptor, **HBA**; brown indicates ring aromatic, **RA**; cyan indicates hypophobic, **HYD**. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
compared with Hypo4 and Hypo3, the presence of one HBA and RA groups increase the ranking score of the hypothesis (from 32 to 35). From the overall results, we strongly suggest that the RA, 2-HBA and HYD features (Hypo1) play a decisive role in selectivity between Aurora kinase-B and Aurora kinase-A inhibitors. Hence Hypo1 was selected as a best qualitative hypothesis for Aurora kinase-B.

Fig. 3 represents the mapping of Hypo1 with one of the training set compound (Compound B1) which shows a maximum fit value of 3.73. During the fit process the conformations of compounds were calculated within the 20 kcal/mol energy threshold to minimize the distance between Hypo1 features and mapped atoms of Compound B1. Hypo1 has four features hence, the maximum fit value of any ligand alignment with this model is 4. Alignment of the training set was performed and found to give fit score ranging from 3.73 to 1.74.

The test set consists of 19 structurally distinct compounds from training set and also has good specificity and selectivity against Aurora B/A kinases was used to validate the Hypo1. Hypo1 was used to screen the test set compounds, which is not only to validate whether the hypothesis has an ability to distinguish the active and inactive inhibitors of Aurora kinase-B but also to find the selective features which can clearly differentiated Aurora kinase-B from Aurora kinase-A inhibitors. While screening the test set compounds using Hypo1, it shows maximum fit value of 3.6 for the compounds having good activity values for selective inhibitors of Aurora kinase-B. But it shows the maximum fit values of 2.9 for the Aurora kinase-A selective inhibitors which indicate that Hypo1 can be the best hypothesis to differentiate the Aurora kinase-B from Aurora kinase-A inhibitors. When comparing these fit values with the activity values (IC50) of Aurora kinase-B and Aurora kinase-A specific inhibitors, it distinctly indicates that Hypo1

![Fig. 3. Compound B1 shows best alignment with Hypo1 hypothesis.](image)
establish maximum fit value for a selective Aurora kinase-B inhibitors (fit value >3.0) when compared with Aurora kinase-A inhibitors (fit value <2.9) and the result was indicated in Table 3.

Still there is a question whether the RA or HBA group plays a major role in discriminate the selective inhibitors of Aurora kinase-B and Aurora kinase-A. when this Hypo1 is compared with already reported hypothesis of Aurora kinase-B [15] reveals that RA group should be important for Aurora kinase-B inhibition. In order to determine whether RA group is necessary in Aurora kinase-B inhibitor selectivity, we abolished this feature from Hypo1, which represent as HypoB (RA feature in Hypo1 was removed) and used to screen the test set, but HypoB fails to differentiate Aurora kinase-B inhibitors from Aurora kinase-A was confirmed by its fit values (Table 4). All the compounds present in the test set shows a fit values in the range of 2.8–2.9, from this we concluded that the HypoB was not able to discriminate between the specific inhibitors of Aurora kinase-B and Aurora kinase-A. But in the presence of RA chemical features it shows a different fit values for specific inhibitors of Aurora kinase-B and Aurora kinase-A.

We observed that, RA group will be a key feature which can differentiate the Aurora kinase-B from Aurora kinase-A inhibitors, when this group present in Hypo1 it shows a good fit value for Aurora kinase-B selective inhibitors but in the absence of this feature shows the fit values of Aurora kinase-B equal to that of Aurora kinase-A inhibitors. From the above analyzes, it was concluded that Hypo1 pharmacophore model consist all the essential features necessary for compounds to be highly active and selective towards the Aurora kinase-B receptor. The four features hypothesis model (1-RA, 1-HYD and 2-HBA) are the most excellent one to identify the selective inhibitors towards Aurora kinase-B. Hence, the pharmacophore model (Hypo1) from this study can be used as a computational tool to design selective Aurora kinase-B inhibitors.

Fig. 4. Sequence alignment of template and target sequences using ClustalW.

Fig. 5. Cartoon representation of Aurora kinase-B.
3.3. Virtual screening

Another objective of this study was to identify the novel scaffolds of Aurora kinase-B inhibitor hence, the best hypothesis Hypo1 was employed as 3D query to screen Maybridge and Chembridge databases. As an initial filtration, the predictive Hypo1 screened 26,049, 22,051 compounds, from Maybridge and Chembridge chemical databases that consist of 60,000 and 50,000 small molecules, respectively, which was sorted to 3350 and 4741 based on the maximum fit value of the compounds. Secondly, the screened hit molecules were sorted based on the ADME properties and Lipinski’s Rule of Five [24]. Mainly BBB, solubility and absorption criteria’s were focused in ADME, if the molecules have the level of 3 and 0 for solubility and absorption, respectively, these values represents that the molecules have good solubility and absorption. The drug should not cross the BBB, hence the level ‘3’ was selected means low penetration of BBB. The sorted molecules were satisfied the Lipinski’s rule of five such as Log P less than 5, number of HBA less than 10, number of HBD less than 5, molecular weight less than 500 and number of the rotatable bond not exceeding 10. The number of rotatable bonds was considered to reduce the flexibility of the molecules. These are the important criteria to screen a good orally-bioavailability drug. Considering all the above values, finally, 265 compounds from Maybridge and 390 compounds from Chembridge have the drug like properties and hence these molecules were subjected to molecular docking process.

3.4. Aurora kinase-B 3D structure generation using comparative modeling method

Aurora kinase-B plays a critical role in chemogenomic approaches, unfortunately the 3D structure of the protein is not determined experimentally; hence we generated homology model for human Aurora kinase-B. The structure of Xenopus laevis Aurora
kinase-B (PDBID: 2VRX, Resolution 1.86 Å) was selected as a best template for human Aurora kinase-B which has shown the best sequence identity of 77 percentage. Asp157, Glu155 and Lys106 are the critical amino acids [32] responsible for the catalytic activity in both template and human Aurora kinase-B (amino acids numbering was based on template). The sequence alignment between the template and target protein (Homo sapiens Aurora kinase-B) by ClustalW program (Fig. 4). Fig. 5 reports the cartoon representation of human Aurora kinase-B structure, which clearly shows the N- and C-terminal domains and the ATP binding cleft, was found between these two domains. All β-sheets and α-helices and the backbone structure resembling the same alignment that had been found in Xenopus laevis Aurora kinase-B. Leu122, Glu171 and Asp173 plays a critical role in its function, these are the conserved residues in Aurora kinase-B. Ala173, Glu171, Glu177, Lys122 and Ala157, Glu155, Glu161, Lys106 are the critical amino acids present in the template [33] and target proteins, respectively. The final model was validated using the PROCHECK, to search for deviations from normal protein conformational parameters.

3.5. Validation of homology model

Ramachandran’s plot is a protein structure validation tool for checking the detailed residue-by-residue stereo chemical quality of a protein structure [36]. The phi and psi distribution of Ramachandran’s plot of non-glycine, non-proline residues are summarized in Fig. 6. A good homology model should have >90% of the residues in the favorable region, in our homology model 99.1% of the residues were present in favored and allowed regions as well as with a relative low percentage of residues having general torsion angles which affirms that human Aurora kinase-B model was successfully predicted as accurate. None of the active site residues are present in the disallowed region and also all the bond distances and angles lie within the allowable range about that standard dictionary values which indicated that Aurora kinase-B model is reasonably good in geometry and stereochemistry. The root mean square deviation (RMSD) between the template and the target structure is 0.073 Å (Fig. 7). The main chain parameters plot for the model is shown in Fig. 8 which indicates that the structure compares with well-refined structures at a similar resolution. The six properties plotted: (i) Ramachandran’s plot quality, (ii) peptide bond planarity, (iii) bad non-bonded interactions, (iv) Cα tetrahedral distortion, (v) main chain hydrogen bond energy and (vi) the overall G factor which measures the overall normality of the structure. In brief, the geometric quality of the backbone conformation, the residue interaction, the residue contact and the energy profile of the structure are well within the limits which established for reliable structures. All the evaluation suggests that a reasonable homology model for Aurora kinase-B has been obtained to allow for examination of protein-substrate interactions.

3.6. Molecular dynamic simulation

In order to obtain the energetically favorable stable receptor conformation for docking study, the model was subjected to MD simulation. The root mean square deviations (RMSD) of the protein backbone atoms are plotted as a function of time to check the stability of the system throughout the simulation. During the last 1 ns, the RMSD of the system tends to be converged, indicating the system is stable and well equilibrated. The relative flexibility of the model was characterized by plotting the root mean-square fluctuation (RMSF) relative to the average structure obtained from the MD simulation trajectories. Some of the analyzes like Root Mean Square Deviation (RMSD), Potential Energy (PE) and Root Mean-Square Fluctuation (RMSF) were carried out to check the stability of the model in explicit condition for 5 ns. Fig. 9 shows the overall RMSD analysis of Cα-atoms, which explains the protein structure deviation at atomic level from the initial structure with respect to the function of time. To examine the flexible regions of the model, RMSF plot was generated with respect to their individual residues. A representative average structure was obtained from the saved frames of the last 1 ns MD simulation a trajectory was used for further analyses.

3.7. Molecular docking of Aurora kinase

Molecular docking was performed using LigandFit module in DS. This study was performed to gain insight into the most probable binding conformation of the inhibitors. Molecular docking is a computational technique that samples conformations of small compounds in protein binding sites; scoring functions are used to assess which of these conformations are best complements to
the protein binding site. There are two main aspects to assess the quality of docking methods: (i) docking accuracy, which recognizes the true binding mode of the ligands to the target protein and (ii) screening enrichment which measures the relative improvement in the identification of true binding ligands using a docking method versus random screening.

Initially the template 3D structure was used as receptor and the bound ligand was sketched and minimized to dock into the active site of the protein and the result was analyzed to validate the LigandFit module and also to verify whether the selected parameters are able to produce the suitable orientation of the ligand. When compared the docked ligand pose with the bound pose its shows the similar orientation and the RMSD was 1.24 Å (Fig. 10) as well as the docked pose shows all the necessary interactions which was shown by the bound conformation. Hence, the specific inhibitors of Aurora kinase-B and hit compounds from Maybridge and Chembridge databases which satisfied drug like properties were docked into the active site of Aurora kinase-B, to confirm the interactions of each molecule with critical residues and the orientation of molecules. Applying the scoring functions like PLP1,
PLP2, Dock score, LigScore1 and LigScore2 to find the suitable scoring function to select the potent inhibitors of Aurora kinase-B. Sometimes, single validation may lead to pick some of the false positive and also to miss false negative hence consensus scoring was applied. Consensus scoring was applied to test the specific Aurora kinase-B inhibitors, it shows a valuable results. Then analyzing the interactions of the specific inhibitors through visualization, show good interactions with the critical residues like Ala157, Glu155, and Lys106 plays an essential role in Aurora kinase-B. Out of these residues Ala157 play a critical role in the inhibition of Aurora kinase-B. The pre-validated analysis process was used to select the potent inhibitors from the databases hit molecules. Initially by applying the consensus scoring function, 143 and 209 molecules were selected from Maybridge and Chembridge databases. Out of 143 and 209 molecules 49 and 34 molecules shows a very good H-bond interactions with critical residues like Ala157, Glu155, and Lys106. Totally, 83 hit molecules show very good interactions with the active residues of Aurora kinase-B when compared with specific inhibitors. One compound from each database as well as the two specific Aurora kinase-B inhibitors binding orientations in the active site of the Aurora kinase-B was shown in Fig. 11.

4. Conclusion

Hip-Hop algorithm was used to produce the best hypothesis for Aurora kinase-B based on a series of Aurora kinase-B inhibitors. The four features hypothesis (Hypo1) was selected as a best pharmacophore model for Aurora kinase-B based on the rank score, cluster process. When comparing all the 10 hypotheses HYD and RA groups were indicated the difference in Aurora kinase-B selectivity. The pharmacophore model was further validated using the external test set which consist of specific inhibitors of Aurora B/A kinases. By analyzing the overall results the HYD present in Hypo1 doesn’t show much difference in selectivity between Aurora kinase-B and Aurora kinase-A inhibitors. But the RA group demonstrates divergence in the inhibitor selectivity between Aurora B/A kinases inhibitors. The test set result reveals that Hypo1 has a good capability to separate the Aurora kinase-B from Aurora kinase-A and we have found that RA group plays a major role in Aurora kinase-B selectivity. In order to validate the key RA feature, we obliterated this group from Hypo1 hypothesis and validated with the external test set. But the Hypo1 (RA group was removed) hypothesis was failed to produced the result as Hypo1. Hence, we concluded that the RA group in Hypo1 plays a major role in Aurora kinase-B selectivity. Finally, Hypo1 was used to screen the Maybridge and Chembridge databases and the hit molecules from the databases were further sort out based on the fit value, drug like properties, and the sorted molecules were subject to molecular docking studies. Finally, 49 and 34 Maybridge and Chembridge database compounds show good interactions with the critical amino acids (Asp157, Lys106, and Glu155). Thus, Hypo1 will act as a valuable tool for retrieving structurally diverse, novel and selective compounds. Subjecting these molecules into in vitro studies may be a potent lead for the inhibition of Aurora kinase-B.
Acknowledgement

All students were supported by Basic Science Research Program (2009-0073267), Pioneer Research Center Program (2009-0081539), and Environmental Biotechnology National Core Research Center program (20090091489) through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (MEST). And all students were recipient of fellowship from the BK21 Program of MEST.

References


Fig. 11. The specific inhibitors and the hit molecule in the active site of Aurora kinase-B. (a) Test4B and Test9B. (b) HTS_08249 from Maybridge and Compound 26300 from Chembridge. Small molecules are colored in green, critical residues (Ala157, Lys106, and Glu155) are represented in ball and stick, hydrogen bonds in black are shown. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)