Cyclin Dependent Kinase 2 Inhibitors an Artificial Neural Network Regression QSAR Study

Claudiu N. Lungu*

Department of Chemistry, Faculty of Chemistry and Chemical Engineering, Babes-Bolyai University, Arany Janos Str. 11, 400028, Cluj, Romania

Corresponding author: Claudiu N. Lungu, Department of Chemistry, Faculty of Chemistry and Chemical Engineering, Babes-Bolyai University, Arany Janos Str. 11, 400028, Cluj, Romania, Tel: +40-742255099; E-mail: lunguclaudiu5555@gmail.com

Received date: February 28, 2018; Accepted date: March 14, 2018; Published date: March 23, 2018

Abstract

Cyclin dependent kinase (CDK) plays a major role in regulating the cell dynamic. Kinases are present in all known eukaryotes and their regulatory function pathway was evolutionary conserved, suggesting this pathway plays a dominant role in cell growth control and its disruption may result in cell death. Three-dimensional quantitative structure-activity relationships (3D-QSAR), molecular docking and molecular dynamics (MD) simulation strategies were applied to investigate the molecular interaction between active ligands and cyclin dependent kinase 2 (CDK2). A QSAR model was computed using artificial neural network (ANN) regression, with a good predictive ability in internal and external validation. Results were compared with the prototype inhibitor of CDK, Staurosporine. A mixed analysis embodying the QSAR model, molecular docking and molecular dynamics, enabled the pharmacophore high definition. Data provided by MD were highly consistent with the findings of 3D-QSAR model.

Keywords: 3D-QSAR; Low molecular dynamics simulation; Conformational analysis; Molecular docking; CDK2 inhibitors; Staurosporine

Introduction

Screening a large domain of chemical space for a certain compound shape proved to be a valuable tool in identifying compounds with no evident structural similarity but with affinity for the same binding site of a proteic structure. Compounds resulted after such a process can, by functionalization and derivatization, be “shaped” to mimic structural similarity, thus becoming, by the QSAR paradigm, same or, hopefully, better bioactive, or enhanced in drug-like properties. A QSAR model can be used first to find a representative compound in a data set used to build the model. This compound looked at as the least common multiple (i.e., common pharmacophore) can further be used as a lead structure in similar shape screening. The dissociation constant Kd of the complex ligand-protein can be predicted by the QSAR model. It is also possible that one structure to act on multiple receptors, or to interact with multiple binding sites, located on the same or distinct molecules, such as Staurosporine, able to inhibit protein kinases of different types. Staurosporine (antibiotic AM-2282 or STS) was first isolated in Streptomyces Staurosporus.

It’s main biological activity is the competitive binding to protein kinase, against ATP, being a model of ATP competitive kinase inhibitor; however, its great affinity for many CDKs, in other words, its lack of specificity, prevented Staurosporine from clinical use [1]. Despite the progress in discovering more drug-like inhibitors of CDK2, there is still some chemical space available for potent and selective inhibitors of CDK2. Difficulties lie in achieving isomorphic specificity [2], targeting to specific cells or tissues and ensuring correct degree of inhibition [3]. So far, the interaction between CDK2 and ligands is not completely understood, and related mechanism not yet explained. The large variation in binding affinities of these compounds with CDK2 and the relation of biological activity with the flap motion of the enzyme, as well as, with conformational changes in the catalytic site of CDK2, were investigated using a combined approach including docking and molecular dynamics simulations. The activity data were retrieved from an original series of 264 compounds [4], obtained by isothermal titration calorimetry [5] (ITC) against CDK2; compound structures were curated as follows: only highly active compounds were considered, with Kd <10,000 nM; structural isomers were not considered. Following these steps, 26 final compounds resulted. The above selection was done for obtaining a model based on non-congeneric compounds in order to expand the chemical space; however, the inactive compounds were not considered in the construction of the model. CDK molecules contain a Walker A motif [6] or a P loop associated with a phosphate group. The Walker A motif is present in ATP and GTP binding proteins. The manner of how this loop is important in binding a compound and is investigated in this study.

Methods

A data set of 26 molecules with measured Kd(nM) on human cyclin dependent kinase 2 were used to compute a QSAR model by ANN regression method. Target variable was set as Kd(nM). Dependent variables were as follows: potential energy (kJ/mol), molecular weight (MW), AlogP, polar surface area (PSA), molecular radius (MR), molecular polarity, first Zagreb index (ZI1) [7], Wiener index [8], Xu index [9], Gutman topological descriptor (GTM=ZI2) [10], Eccentricity (ECC) [11], Lipinski’s rule of five values [12], and also total connectivity index. Docking study was performed using AutoDock package [13]. 1QA1 PDB id was chosen as a model for CDK2 receptor. Docking site was retrieved from literature and from the model mentioned above [14]. Furthermore, the structure was prepared by the Protein Preparation Wizard in the Schrödinger software suite, including hydrogen atoms, correcting partial charges using the OPLS 2005 force field and generate protonation states, and in vacuum structure energy minimization. Co-crystal ligand was removed, and the resulting protein model was used as the receptor...
model in the studies. The 26 structures with experimental determine activities (Kd data) are shown in Table 1. Ligprep module incorporated in Phase was used to prepare the compounds, and compute of the probable ionization state at pH 7.4 ± 0.2. Phase 3.0 package was used to generate the pharmacophore for CDK2 inhibitors. Desmond was used to performed the molecular dynamic study using OPLS_2005 force field. The prepared structures were used to generate a pharmacophore model using their biological activity measured in Nm. Ligands were set as active

The data set was divided in a training and test set, respectively, considering a uniform distribution of Kd values in training and test sets. The pharmacophore of these compounds was rationalized from a set of six pharmacophore characteristics: negatively charged group (N) positively charged group (P) and aromatic ring (R) hydrogen bond acceptor (A), hydrogen bond donor (D), hydrophobic group (H). Each pharmacophore feature was defined by a set of chemical structure characteristics. Pharmacophore was explain by the type, location and directionality of each component [15]. Several pharmacophore hypothesis were generated. The hypotheses were ranked according to the following scores: alignment, vector score and a volume score respectively. The predictive power of the QSAR model was validated by the test set compounds externally and internally. Chemical space of activity was also represented in parallel for the observed dissociation constant (Kd) and predicted dissociation constant (Kdp). Balaban index [16], total valence connectivity, Log P and polar surface area (PSA) were used to generate a 3D wire frame surface plot.

Two compounds were chosen for a comparative study: (i) the compound with the best experimentally determine Kd and (ii) Stauroporine. Conformational analysis of the two molecules together with low molecular dynamics applied to CDK2 complex with the compounds was performed. Conformational analysis was used to characterize the regions of the compound exposed to the solvent and present high mobility comparative with the rest of the molecule. In this respect, by rotating the atoms residue around the molecule with 1 degree of freedom, -180 + 180 degrees, a graph containing conformational energies of the structure is obtained. This graph is used to make assumptions regarding compound binding. Low molecular dynamic [17,18] is a technique for observing discrete changes in structure of proteins especially regarding the residues present in binding sites. In this technique, a molecular dynamic is performed on a molecule having the amino acids of the active zone fully flexible, the neighboring residues fixed and distant residues being inert. Two steps were adopted: (i) molecular 3D-similarity indices were computed (for the 26 structures whose activity had been experimentally reported in Table 1) in view of 3D-QSAR study; (ii) docking and molecular dynamics (MD) simulation were performed in order to explore favorable coordinates of the CDK2 inhibitors in docking as well as to understand the large variation in the binding affinities of those compounds with CDK2.

Results

20 hypotheses were produced and analyzed. After analyzing active ligands alignment and the generated hypotheses, hypothesis AADRR19 was selected. The selected hypothesis contained one hydrogen bond donor (D4), two hydrogen bond acceptors (A2 and A3), and two aromatic rings (R10 and R12), as shown in Figure 1.

Figure 1: (a) Common Pharmacophore for active ligands. Pharmacophore features are color coded: blue H-donor, pink H-acceptor, orange aromatic ring; (b) Distance between pharmacophore features are shown in Ångstroms.

QSAR, obtained independently of pharmacophore model generated using Schrodinger software, has the following statistical parameters: The training set Pearson R2=0.9838, with a mean standard error MSE=0.446 (see the plot in Figure 2). Descriptors that have high impact on model building were PSA of compounds followed by potential energy of the compounds and AlogP respectively. Topological descriptors used have a modest contribution to the model. From all MW and polarity had the modest contribution in model architecture. General model equation $y=0.8674x+270.92$ where y is the target variable, x a certain dependent variable.

![Figure 2](scatter_plot.png)

Scatter plot of predicted dissociation constant versus experimental dissociation constant

<table>
<thead>
<tr>
<th>Kd</th>
<th>MW</th>
<th>AlogP</th>
<th>PSA</th>
<th>MR</th>
<th>P</th>
<th>ZM1</th>
<th>W</th>
<th>Xu</th>
<th>Ram</th>
<th>GMT</th>
<th>ECC</th>
<th>PE</th>
</tr>
</thead>
<tbody>
<tr>
<td>MW</td>
<td>0.862</td>
<td>0.9</td>
<td>0.976</td>
<td>0.973</td>
<td>0.977</td>
<td>0.963</td>
<td>0.981</td>
<td>0.922</td>
<td>0.978</td>
<td>0.978</td>
<td>0.28</td>
<td></td>
</tr>
<tr>
<td>AlogP</td>
<td>0.862</td>
<td>0.604</td>
<td>0.821</td>
<td>0.844</td>
<td>0.856</td>
<td>0.828</td>
<td>0.808</td>
<td>0.908</td>
<td>0.834</td>
<td>0.755</td>
<td>0.42</td>
<td></td>
</tr>
<tr>
<td>PSA</td>
<td>0.9</td>
<td>0.604</td>
<td>0.89</td>
<td>0.839</td>
<td>0.83</td>
<td>0.916</td>
<td>0.894</td>
<td>0.712</td>
<td>0.917</td>
<td>0.963</td>
<td>0.12</td>
<td></td>
</tr>
</tbody>
</table>
Table 1: Descriptors correlation matrix: Kd-dissociation constant (nM), MW-molecular weight, AlogP-partition constant, PSA-polar surface area, P-polarity, ZM1-first Zagreb index, W-Wiener index, Xu-Xu index, Ram-ramification index, GMT1-Gutman molecular topological descriptor1, ECC-Eccentricity, PE-potential energy.

Discret equation of descriptors for the target variable (y) correlation:
y = PE * 0.00234096 + 0.11584,
y = MW * 0.00252934 - 0.388785,
y = AlogP * 0.178285 + 0.128918,
y = PSA * 0.00655576 - 0.278792,
y = MR * 0.0101536 - 0.457824,
y = P * 0.020245 - 0.361666,
y = ZM1 * 0.00655738 - 0.359016,
y = W * 0.000183697 + 0.0461768,
y = Xu * 0.0411207 - 0.471759,
y = GMT1 * 4.16536e-05 + 0.0495158,
y = ECC * 0.00186047 - 0.0469767.

Predicted activities of the observed and test set molecules are also listed in Table 2.
As shown in Figure 1, the presence of two H-acceptor groups, two aromatic rings and one H-donor group is mandatory for the inhibitory effect on CDK2. By superimposing the common pharmacophore model on the best hit (compound #9) one observes that both coordinations of H-acceptor groups are the same. Also by superimposing the best hit #9 over #26, with the lowest Kd, one observes that #9 has six H acceptor groups, four H-donor groups, two hydrophobic groups and two hydrogen bound forming groups, compared with three H acceptor groups, two H donor groups, two hydrophobic groups, and two aromatic groups, for #26. By increasing the number of H acceptor groups and the distance between hydrophobic and aromatic groups, larger Kd values are observed. Also orientation/alignment of ligand in the binding site is crucial. Figure 3 illustrates superposition of #9 over the common pharmacophore comparative to #26; one observes a difference of 90° between orientation of #9 and #26, respectively.

Chemical space of activity, represented in topological terms, showed similar coordinates for Kd and Kdp (represented using Balaban index, a highly discriminating index) and total valence connectivity. Chemical space of drug like properties represented using polar surface area and Log P shows the same shape as the topological space (Figures 4 and 5).
Compound #9 in complex with CDK2 was compared with the crystallographic structure of the CDK2 complex with inhibitor Staurosporine.

Compound #9 forms (by giving away a proton from N 4582 with π electrons from Phe 80) an aren-H bound, O 4602 accepts a proton from Gln 85 and forms an H-bond, N4603 gives away a proton and forms with Leu 83 an H-bond. Then O4588, C4534, C 4544, C4549, O 4602 and O4604 respectively are exposed to the solvent. Staurosporin in complex with CDK2 was imported from 1AQ1 PDB model.

Compound #9 forms (by giving away a proton from N 4582 with π electrons from Phe 80) an aren-H bound, O 4602 accepts a proton from Gln 85 and forms an H-bond, N4603 gives away a proton and forms with Leu 83 an H-bond. Then O4588, C4534, C 4544, C4549, O 4602 and O4604 respectively are exposed to the solvent. Staurosporin in complex with CDK2 was imported from 1AQ1 PDB model.

By superimposing #9 with Staurosporine (Figure 7), one observes that difference in activity and consecutively in specificity is given by the regions exposed to the solvent and form close bonds with the amino acids present at the margin of the binding pocket. By superimposing #9 with Staurosporine (Figure 7), one observes that difference in activity and consecutively in specificity is given by the regions exposed to the solvent and form close bonds with the amino acids present at the margin of the binding pocket. Conformational analysis of these regions for both compounds showed that: (i) in case of Staurosporin the ramified exposed residues, when rotated around the binding pocket, have significantly favorable conformational energies compared with the hooked regions of # 9 that has higher conformational energies. It is observed that Staurosporin has lower energy values and a large energy domain in contrast with # 9. Regions with lower conformational energies suggest a stronger binding and a larger topological domain for forming favorable bonds. Some conformational energy values are shown in Figure 8 (see supplemental material for C(2265)-N(2264)-C(2234)-C(2233); O(3957)-S(3956)-C(3955)-C(3952); O(3957)-S(3956)-N(3958)-H(3970).

Figure 9a shows all the 26 inhibitors docked into the pocket of CDK2. All molecules were positioned relatively the same, shared a similar binding mode excepting some molecules due to low activity with unfavorable conformations (i.e orientation of the functional groups) for the binding. The binding modes of the most active compounds correspond with the binding mode of the prototype molecule Staurosporine. Compounds with poor activity (see Figure 3) bind differently from Staurosporine. This difference in the binding modes resulted in distinct activities. Molecular docking and dynamics Simulation were done to further characterized the binding mode of the most important functional groups to CDK2. In docking procedure ligands was treated as flexibly and the protein was held fixed. RMSD value for heavy atoms of the ligands, between the generated docked pose and the native pose was 0.82 Å (Figure 9a). Docking was able to reproduce the native conformation successfully (low RMSD value between the docked pose and the initial geometry) (Figure 9b).
Low molecular dynamic applied on CDK2 complex with a ligand 9 and Staurosporine, respectively, demonstrated the presence of P loop. Target temperature was set to be 310.15Ko. The final model (Figures 10a-c and 11) was retrieved after 9 picoseconds. One structure was maintained inert for analogy with second structure on which low molecular dynamics was applied. All operations were made on the minimized second structure.

In summary a strategy based on multiple computation techniques was used in order to explore the structure base inhibition process for a series of CDK2 inhibitors. Docking was utilized to generate hypothetical binding modes for ligands. Low molecular dynamic was used to evaluate the binding mode from the receptor perspective. Both Staurosporine and #9 have a lipophilic core made of coplanar aromatic rings that fit into the binding pocket. As suggested by the QSAR model shape is of curtail importance. Both Staurosporine and #9 have an external region that binds with marginal amino acids acting like a “hook” in fixing the compound at it's site. Furthermore the satisfactory agreement between experiment and theory and between QSAR model and pharmacophore model (build independently) suggested that the QSAR model has good correlation and predictive power and straitens suggested findings.

References


