Pharmacophore-based screening and identification of novel human ligase I inhibitors with potential anti-cancer activity

Shagun Krishna*a,*, Deependra Kumar Singh*a,*, Sanjeev Meena,b Dipak Datta,b,c Mohammad Imran Siddiqia,c,#, Dibyendu Banerjeea,c,#

*Both authors contributed equally.
# Corresponding authors: email: mi_siddiqi@cdri.res.in, d.banerjee@cdri.res.in
a Molecular and Structural Biology Division, CSIR-Central Drug Research Institute, Lucknow-226031, India
b Biochemistry Division, CSIR-Central Drug Research Institute, Lucknow-226031, India
# Academy of Scientific and Innovative Research

Abstract

Human DNA ligases are enzymes that are indispensable for DNA replication and repair processes. Among the three human ligases, ligase I is attributed to the ligation of thousands of Okazaki fragments that are formed during lagging strand synthesis during DNA replication. Blocking ligation therefore can lead to the accumulation of thousands of single strands and subsequently double strand breaks in the DNA, which is lethal for the cells. The reports of the high expression level of ligase I protein in several cancer cells (versus the low ligase expression level and the low rate of division of most normal cells in the adult body) support the belief that ligase I inhibitors can target cancer cells specifically with minimum side-effects to normal cells. Recent publications showing exciting data for a ligase IV inhibitor exhibiting anti-tumor activity in mouse models also strengthens the argument for ligases as valid anti-tumor targets. Keeping this in view, we performed a pharmacophore based screening for potential ligase inhibitors in the Maybridge small molecule library and procured some of the top ranking compounds for enzyme based and cell based in vitro screening. We report here the identification of novel ligase I inhibitors with potential anticancer activity against a colon cancer cell line.

Keywords: Pharmacophore-Based Screening; Ligase Inhibitor; Molecular Docking; Anti-Cancer Agents; Molecular Modelling.

1.0. Introduction

DNA ligases are enzymes that join DNA molecules together. They do so by joining the adjacent 3'-hydroxyl and 5'-phosphoryl termini to form a phosphodiester bond in duplex DNA. They were first reported by the Gellert, Lehman, Richardson and Hurwitz laboratories in 1967. Inhibiting the function of ligases can have potentially serious consequences for DNA replication and repair events, processes that are most often targeted in cancer therapy. DNA ligases are grouped into 2 families of ATP-dependent and NAD+-dependent ligases, according to the cofactor required for ligase-adenylate formation. Human ligases fall into the ATP-dependent family of ligases. The reaction mechanism for ATP-dependent ligases involves the formation of a covalent enzyme-AMP intermediate from the cleavage of ATP to AMP and pyrophosphate. The adenylate group from AMP is then transferred to the 5'-phosphate of the nicked DNA molecule. Finally, the DNA ligase seals the gap by phosphodiester bond formation, via the displacement of the AMP residue with the 3'-hydroxyl group from the adjacent DNA strand.

There are three known human DNA ligases (hLigI, III, and IV) performing some vital functions in the body. Among them, DNA ligase I has an indispensable role in
replication. During replication, the two anti-parallel strands of DNA are copied differently. One strand is copied in a continuous manner (and called the leading strand) whereas the other strand, due to its opposite orientation, is copied discontinuously in small stretches called Okazaki-fragments, after its discoverer Reiji Okazaki. These fragments then have to be sealed together to make a continuous strand of DNA called the lagging strand. This important function during replication is performed by DNA ligase I. It is also involved in other important DNA repair processes such as nucleotide excision repair (NER)\textsuperscript{13,14} and Base excision repair (BER)\textsuperscript{15-17} and micro-homology mediated end-joining (MHEJ)\textsuperscript{18}. The hLigIII is involved in backup function of Okazaki fragment joining as well as in NER, BER and MHEJ\textsuperscript{14-19}. The hLigIV enzyme is mostly restricted to double strand break repair pathways like non-homologous end-joining (NHEJ)\textsuperscript{20,21}. Consequently severe cell lethality, increased genomic instability and hypersensitivity to DNA damage can be caused by deficiency in DNA ligation. Tomkinson \textit{et al} has mentioned that an individual with an inherited mutation in the hLigI gene exhibited retarded growth, development and immunodeficiency\textsuperscript{12}. Microcephaly and immunodeficiency are associated with DNA ligaseIV deficiency (LIG4 syndrome) and severe combined immunodeficiency with microcephaly, growth retardation, and sensitivity to ionizing radiation due to NHEJ1 deficiency (NHEJ1 syndrome)\textsuperscript{22}. No patients with ligase III deficiency have been identified till date.

The levels of hLigI have been found to be high in several cancers compared to normal cells\textsuperscript{23-25} and hence there is a high probability that hLigI may serve as a good target for cancer therapy. Accumulation of single strand breaks followed by double strand breaks in subsequent cycles of replication is the hallmark of cells deficient in hLigI protein. Several inhibitors against hLigI have been reported but there is a need for the development of more potent specific inhibitors that can be used in therapy.

During the past few years, virtual screening (VS) has been coming out as a potent and established tool for identification of novel hits. The identification of compounds having a desired activity that structurally depart from known active reference compounds and represent new chemotypes is the primary goal of VS\textsuperscript{26}. This includes both ligand-based screening approaches through which test compounds structurally unrelated to a or a set of known active compound can be identified and structure-based virtual screening methods through which molecules that complement the target protein active site can be identified\textsuperscript{27}.

In the present study, we have employed a pharmacophore-based virtual screening approach for searching novel hLigI inhibitors within commercially available Maybridge database containing more than 56000 compounds. On the basis of previously reported inhibitors of the targeted protein, pharmacophore model was generated. The hit molecules identified through the pharmacophore-based virtual screening were subjected to docking into the protein active site, and further their electrostatic binding free energy was calculated and the shortlisted compounds were subjected to biological assay.

2.0. Materials and Methods

In the sections below, we report the pharmacophore modelling and \textit{in silico} screening of the Maybridge compound library for human DNA ligase inhibitors and report the identification of a new DNA ligase inhibitors that show specific anti-cancer activity in colon cancer cell line.

2.1. Generation of Pharmacophore Models and Selection of Dataset

For the generation of pharmacophore model, top three inhibitors of hLigI were selected from the work of Tomkinson \textit{et. al.}\textsuperscript{28}. All compounds were drawn by using sketch module of the molecular modeling suite Sybyl7.1\textsuperscript{29}. The geometries of these compounds
were then optimized by Sybyl 7.1 using the Tripos force field and Gasteiger–Huckel charges. The energy minimization was done using Powell method with an energy convergence gradient of 0.001 kcal mol$^{-1}$. A set of pharmacophore models was generated using GASP module of Sybyl7.1. GASP uses a genetic algorithm (GA) for the superimposition of sets of flexible molecules. GA is used for determining the association between functional groups in the superimposed molecules and the alignment of these groups in the common geometry for receptor binding. In GA the molecules are treated as a chromosome that trains about flexible bonds and mapping between hydrogen-bond donor proton, acceptor lone pair and ring centre features in pairs of molecules. The fitness function employed in GA is actually a weighted combination of the number and similarity of feature that have been aligned, the volume integral of the overlaid molecules and the van der Waals energy of the conformation of molecules defined by the torsional angles encoded in the chromosomes.

Table 1 shows the structure and IC$_{50}$ values of three inhibitors that have been used for the generation of pharmacophore used in this study. A UNITY query was prepared by mining the chemical features and applying the distance constraints of the GASP generated pharmacophore model. The query was consequently modified to optimize both selectivity and specificity. The subset of World of Molecular Bioactivity (WOMBAT) database is considered here as decoy or inactive set and a set of known inhibitors was put in the test dataset considered as active set for the validation of the pharmacophore model.

**Table 1. Structure of the three compounds and their IC$_{50}$ values used in generation of pharmacophore model.**

<table>
<thead>
<tr>
<th>S.No</th>
<th>Structure</th>
<th>IC$_{50}$ (µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><img src="image1.png" alt="Structure 1" /></td>
<td>0.6±0.5</td>
</tr>
<tr>
<td>2</td>
<td><img src="image2.png" alt="Structure 2" /></td>
<td>4±2</td>
</tr>
<tr>
<td>3</td>
<td><img src="image3.png" alt="Structure 3" /></td>
<td>5±2</td>
</tr>
</tbody>
</table>
2.2. Pharmacophore Validation

Validation of a pharmacophore model was performed with the aim to verify its ability to discriminate between active molecules from the inactive in the database. The quality of pharmacophore model was validated using test set and decoy set methods. Various statistical parameters such as percentage yield of actives, percent ratio of actives, Enrichment Factor (EF), false negatives, false positives and Goodness of hit score (GH) were further calculated.

2.2.1. Test Set

The test set is used to check whether the generated pharmacophore model is able to retrieve active compounds. The geometry of the entire test set molecules was generated in the same manner as that for the compounds used in pharmacophore model generation.

2.2.2. Decoy Set

WOMBAT database as decoy set is used to analyze the discriminating power of the best model by calculating goodness of hit score (GH) and enrichment factor (EF).

In particular, for determining the discriminatory power of pharmacophore models, Goodness of Hit List (GH) is calculated as follows formulae:

\[ GH = \frac{(H_a)(3A + H_t)}{4H_t A} \times \left(1 - \frac{H_t - H_a}{D - A}\right) \]

Where \( H_a \) is the number of active hit molecules screened, \( H_t \) is the number of total hit molecules screened, \( A \) is the number of active molecules in the database, and \( D \) is the total number of molecules in the database.

The enrichment factor (EF) shows how many times more active molecules are identified by the model than by a random selection of the same number of compounds and it is given by the formulae:

\[ EF = \frac{(H_a \times D)}{(H_t \times A)} \]

2.3. Database Searching and Ranking of Hit Molecules

The validated pharmacophore model was used as a three dimensional (3D) structural query for retrieving compounds from Maybridge database using UNITY module incorporated with Sybyl7.1. UNITY is a database screening and scrutinizing tool that is proficient to perform 2D, 3D as well as flexible searches. By default; UNITY 3D database screening performs a prefiltering of database compounds discarding all molecules that do not match Lipinski’s rule of five. The compounds identified by the search were further ranked according to UNITY score.

2.4. Target Structure Preparation and Molecular Docking

The crystal structure of hLigI was retrieved from Protein Data Bank (PDB Identifier 1X9N) and this structure is a complex with co-crystallized deoxyribonucleotide moieties. The deoxyribonucleotides were removed from the complex. Hydrogen atoms were then added followed by energy minimization using Sybyl7.1 with default values. The resulting structure was then used for docking of hit molecules identified by virtual screening, performed on a Silicon Graphics Origin300 server running under IRIX6.5 operating system with the use of the FlexX module of Sybyl7.1. The docking method implemented in FlexX is based on an incremental construction algorithm that first splits the compound into its basic fragments and automatically selects the base fragment by the use of a pattern recognition technique called pose clustering and positions it into an active site followed by the incremental building of the remaining of the portion onto the active site. The conformational flexibility of the ligand is included by generating multiple conformations for each fragment and placement of the ligand is scored to estimate the free energy of binding of protein–ligand interactions.
2.5. Electrostatic Binding Free Energy Calculation and Rescoring of Top Hits

The Adaptive Poisson-Boltzmann Solver (APBS) software was used for computing the electrostatic contribution of binding free energy of ligand and protein complex. It helped to select the final ranked set of top hits. It works with the combination of molecular mechanics with continuum solvation models. The ligand conformations obtained after the docking process with hLigI, were used as an input for binding energy calculations. The PDB format of hLigI and ligands complex was converted to PQR format by using PDB2PQR program. The AMBER force field parameters were assigned to the hLigI protein structure.

The method employed in calculating binding free energy describes the mechanism of binding process into two components. First desolvating the opposing surfaces of both ligand and receptor and then allowing the charges of the two molecules interact. It is then possible to disentangle the change in electrostatic free energy on molecular association into three components. First is the change in electrostatic desolvation free energies of the ligand upon binding, second is the change in electrostatic desolvation free energies of the receptor upon binding, and third is the ligand receptor interaction energy in the presence of the surrounding solvent. The electrostatic binding free energy is defined as the following equation:

\[
\Delta G_{\text{ele}}^{\text{bind}} = \Delta G_{\text{desol}}^L + \Delta G_{\text{desol}}^R + E_{LR}
\]

The loss of electrostatic interaction between the solvent and receptor (or ligand) upon binding is characterized the electrostatic contribution to the desolvation energy of the receptor, and is calculated in the two step method used by Perez et al.

2.6. In vitro DNA ligation assay

DNA ligation assay was done essentially as described by Chen et al. Three different single strand DNA oligos, 52-mer, 25-mer, and 27-mer, were annealed to form a double-stranded nicked substrate for the ligase enzymes. The 27-mer oligo was labelled at the 5′ end with a fluorescent dye cyanin3 (Cy3) for easy detection on the GE Life Sciences LAS4010 detector. The reaction mixture (20µL) contained 1 pmol of labelled DNA substrate and 0.2 pmol of purified hLigI in a ligation buffer containing Tris-Cl (50mM, pH 7.5), MgCl2 (10mM), BSA (0.25 mg/ml), NaCl (100mM) and ATP (500µM). The double-stranded nicked DNA was incubated with the purified ligase in the absence or presence of inhibitors at 37°C for 30 min. Reactions were stopped by adding 10µl of stop buffer (90% formamide and 10% of 50mM EDTA). The ligated DNA molecule would be larger in size and run higher up in denaturing gel containing 7M urea and 12% acrylamide. If an inhibitor is added to the reaction mixture, this would lead to an inhibition of ligation and a corresponding loss of labelled ligated product in the gel. We would then be able to calculate the percentage of inhibition of ligation by estimating the density (by image quant LAS 4010) of ligated product in the lane without inhibitor and comparing it with the lanes containing different inhibitor molecules or with the same inhibitor molecule at different concentrations.

2.7. Electrophoretic Mobility Shift Assay (EMSA)

EMSA was performed with a double stranded, fluorescein dye (FAM) labelled oligo as substrate for ligase I protein. Of the three oligos used to make the substrate, one was a 5′-FAM labelled 27-mer oligo with a dideoxy modified 3′ end (5′-/56-FAM/GTACTATCTCCTACTGACATACATAGAC/3ddc/-3′). Other two oligos are 25-
mer(5′-CTGATCTACCAATCGATCGACGTAC-3′) and 52-mer (5′-GTACGTGATCGATTGAGTACATCAGGTTCTATGTATGTACGAGATAGTAC-3′). These oligos were annealed to obtain a non-ligatable nicked substrate for ligase I protein. We incubated 10 pmole of hLigI with 50, 100, 200µM inhibitor and 2 pmole non-ligatable single nicked substrate DNA in a ligation buffer containing Tris-Cl (50mM, pH 7.5), MgCl₂ (10mM), BSA (0.25 mg/ml), NaCl (100mM) and ATP (500µM), in a reaction volume of 20µl, for two hours on ice. After the addition of 10µl of native gel buffer (tris (pH7.5) 50mM, 20% glycerol, 0.05% bromophenol blue), samples were separated by 6.5% native PAGE and bands were detected by image quant LAS4010.

2.8. DNA intercalation assay

To determine whether the ligation inhibition occurs due to DNA intercalation, we performed the DNA intercalation assay. In this case, we incubated 100 ng of linearized pUC18 plasmid with increasing concentrations (50, 100, 200µM) of the inhibitors for 30 min at 37°C before running the mixture in an agarose gel. Doxorubicin (positive control) and ampicillin (negative control) were used as controls for DNA intercalation. Reaction products were resolved on a 1% agarose gel at 5.3 V/cm. Gel was visualized by ethidium bromide staining.

2.9. SRB assay

Various cancer cell lines were obtained from American Type Culture Collection (ATCC), USA. These cells were grown in recommended media supplemented with 10% FBS and PenStrep in a 5% CO₂ humidified atmosphere at 37 °C. No cultures beyond 25 passages were used for the study. A standard colorimetric SRB (sulforhodamine B) assay was used for the measurement of cell cytotoxicity. In brief, 5000 cells were added to each well of 96-well plate and treated with 10/100 μM of test compounds and untreated cells received the same volume of DMSO as vehicle control. After 48 h of exposure, cells were fixed with ice-cold TCA, stained with SRB in 1% acetic acid, washed and bound dye was dissolved in 10mM Tris base and absorbance were measured at 510 nm on a plate reader (Epoch Microplate Reader, Biotek, USA). The cytotoxic effects of compounds were calculated as % inhibition in cell growth as per the formula [100-(Absorbance of compound treated cells/ Absorbance of untreated cells)] X 100.

2.10. Determination of IC₅₀ values

IC₅₀ was determined by incubating the cells in 5 incremental concentrations of the inhibitor from 2-10µM concentrations. For this purpose, 5000 cells of DLD-1 were plated onto 96-well plates and let adhere for 12 hours. After 12 hours, inhibitors were added and incubated for 48 hours. SRB assay was then performed and the concentration of the inhibitor that inhibited 50% cellular growth of cells was determined using the Prism Graphpad software.

3.0. Result and Discussion

3.1. Pharmacophore Generation

The overall computational process employed in the study is shown in Figure 1. A set of four pharmacophore models was generated using GASP with the help of selected inhibitors. The four models generated are shown in Figure 2. Various statistical parameters were calculated for these models and are discussed in the next section. Model 4 did not identify any molecule from the test set or decoy set, hence it was not considered for statistical evaluation. The pharmacophore models included in statistical evaluation consist of three features generated by GASP alignment: one hydrophobic feature and two acceptor atoms.
"Figure 1. Flow chart showing the methodology applied in the present study."
3.2. Pharmacophore Validation

A good pharmacophore model should be able to enrich maximum biologically active molecules from a structurally diverse dataset as well as to reject most of inactive molecules. To validate the consistency of the pharmacophore models, they were subjected to screening the test set containing 33 compounds and a subset of WOMBAT consisting of 1204 molecules, which are considered here as inactive for hLigI since they are active against other proteins.

Enrichment factor (EF), goodness of hit (GH) and other statistical values have been calculated for the three selected pharmacophore models using test set and WOMBAT database. Model1 showed an EF value of 24.50 by mapping 17 active compounds from the 26 screened compounds from the database. The EF value for Model2 and Model3 were 9.37 and 5.26 as they have retrieved 13 and 8 active compounds respectively. The high is the EF value the greater is the ability of a pharmacophore in identifying the active compounds. Model1 has GH value of 0.614 while Model2 and Model3 have GH value 0.277 and 0.160. This validation result which is tabulated in Table 2 showed that Model1 (Fig: 2A) with minimum false positive and false negative, good EF value and GH score was capable enough to be used in the further
screening of Maybridge database to identify novel leads. The Model1 selected for further study is although very simple and have only three pharmacophoric features, yet it has been statistically validated. In the past also there are studies reported in literature wherein 3-point pharmacophore models were generated and used for the identification of novel lead molecules against several protein targets.53,54

“Figure 3. Selected active compounds (A) HTS01682 and (B) NRB00556 aligned on pharmacophore model 1.”

“Table2. Statistical parameters obtained after the screening of test set molecules for Pharmacophore model.”

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Model1</th>
<th>Model2</th>
<th>Model3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total no of molecule in Database (D)</td>
<td>1237</td>
<td>1237</td>
<td>1237</td>
</tr>
<tr>
<td>Total no of actives in Database (A)</td>
<td>33</td>
<td>33</td>
<td>33</td>
</tr>
<tr>
<td>Total Hits (Ht)</td>
<td>26</td>
<td>52</td>
<td>57</td>
</tr>
<tr>
<td>Active Hits (Ha)</td>
<td>17</td>
<td>13</td>
<td>8</td>
</tr>
<tr>
<td>% yield of actives (H_a/H_t × 100)</td>
<td>65.38%</td>
<td>25.00%</td>
<td>14.03%</td>
</tr>
<tr>
<td>% ratio of actives (H_a/A × 100)</td>
<td>51.51%</td>
<td>39.39%</td>
<td>24.24%</td>
</tr>
<tr>
<td>Enrichment Factor (EF)</td>
<td>24.50</td>
<td>9.37</td>
<td>5.26</td>
</tr>
<tr>
<td>False negative (A-Ha)</td>
<td>16</td>
<td>20</td>
<td>25</td>
</tr>
<tr>
<td>False Positive (Ht-Ha)</td>
<td>9</td>
<td>39</td>
<td>49</td>
</tr>
<tr>
<td>Goodness of Hit Score (GH)</td>
<td>0.614</td>
<td>0.277</td>
<td>0.160</td>
</tr>
</tbody>
</table>
3.3. Database Searching and Scoring

The database searching was performed using UNITY with Model 1. In Unity search, the compounds present in the database are subjected to a conformational flexible 3D alignment based on the direct tweak’s algorithm. In this algorithm, low energy conformers are generated for each compound in the database that possesses the shortest distance to the model features. Thus, there is no need of a conformational model for database compounds and the VS is hastened by evading fitting of a large number of compound conformers. The partial matching of pharmacophore features is also allowed in UNITY.55

As a result of flexible search; we found 16000 hits out of 56000 molecules that were mapped to the pharmacophore model. Figure 3(A) and 3(B) shows the active compounds mapped on pharmacophore model1. The compounds identified by the search were further ranked according to UNITY score. The ranking of hit list was done using the simple scoring function implemented in UNITY, where $R_{\text{energy}}, R_{\text{bond}}$ and $R_{\text{rms}}$ are the ranks of the hits by energy, number of rotatable bonds and RMS correspondingly. The smaller the score the better the hit will be. The score is calculated from the following formula: 56

$$
\text{Score} = R_{\text{energy}}\sqrt{R_{\text{bond}} + R_{\text{rms}}}
$$

3.4. Molecular Docking

A putative DNA binding site within the DBD was identified in the crystal structure of hLigI (PDB ID: 1X9N) which is co-crystallized with DNA. For the docking procedure, we focused on four residues, His337, Arg449, Arg451 and Gly453 that are located in the central region of the DBD and make direct contacts with the DNA substrate.24 This region was inspected meticulously to identify the residues that are involved in DNA binding and then those residues were selected for the active site which are engaged in making hydrogen bond with DNA. Initially we docked the three inhibitors28 to this active site to optimize the parameters of the docking program and also to validate our selected active site. All these inhibitors docked well at this site. The top 3000 compounds obtained by scoring of the identified hits were subjected to docking into the above mentioned active site. The binding modes were generated according to the FlexX scoring scheme (Table 3), which is represented by the structure with the most favourable binding free energy ($\Delta G_{\text{bind}}$). The top scoring docked conformations of ninety-nine compounds; having energy greater than -25.00 kcal/mole, in the active site of hLigI were retrieved and analysed in terms of the favoured mode of binding of the molecules and key residues employed in the interaction.

3.5. Binding Energy Analysis

An imperative constraint for a successful VS experiment is to precisely predict the binding energies of the docked conformation. In this perspective, our VS protocol applied the use of binding free energy calculation to propose the final ranked set of top thirty hits. Now a days, it has been validated as a VS refining tool to prioritize active hits.57 Table 3 shows the APBS binding free energy to that of FlexX for the top five complexes selected for biological assay. There are differences in the binding energy of FlexX and APBS; this is because FlexX does not calculate columbic contribution from all of atoms in protein like APBS. As our initial goal was to identify compounds with better binding affinity, in the course of this we have reranked our top thirty docked complexes to achieve the final set of proposed hit based on their electrostatic binding free energy. The binding energy of the top thirty docked complexes is given in supplementary Table 1.
Table 3. Docking energy, Electrostatic binding energy (BE) and ligase inhibition activity of five compounds subjected to ligation assay.

<table>
<thead>
<tr>
<th>S.No</th>
<th>Name</th>
<th>Structure</th>
<th>Electrostatic BE(kcal/mol)</th>
<th>Docking energy by FlexX (kJ/mol)</th>
<th>Inhibition of ligase activity (IC₅₀)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>HTS07250</td>
<td></td>
<td>-28.9307</td>
<td>-28.444</td>
<td>&gt;100µM</td>
</tr>
<tr>
<td>2</td>
<td>NRB00556</td>
<td></td>
<td>-37.363</td>
<td>37.56±6.97µM</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>HTS01682</td>
<td></td>
<td>-25.6239</td>
<td>-28.548</td>
<td>24.93±3.71 µM</td>
</tr>
<tr>
<td>4</td>
<td>HTS07014</td>
<td></td>
<td>-17.9523</td>
<td>-28.897</td>
<td>&gt;100µM</td>
</tr>
<tr>
<td>5</td>
<td>TG00130</td>
<td></td>
<td>-16.4692</td>
<td>-28.897</td>
<td>&gt;100µM</td>
</tr>
</tbody>
</table>

The five compounds were purchased among the top thirty compounds after analysing their binding energy as predicted by FlexX and APBS and also the mode of interaction as predicted by docking simulations. The compounds HTS07250, NRB00556, HTS01682, HTS07014 and TG00130 were purchased as per their availability and subjected to DNA ligation assay and we have found that HTS01682 and NRB00556 significantly inhibit the ligation process accomplished by hLigI. The binding energy was found to be -28.0327 kcal/mol and -25.6239 kcal/mol for the active Compound NRB00556 and compound HTS01682 respectively.
"Figure 4. Inhibition of DNA ligase I activity for compounds HTS01682, HTS07250, TG00130, NRB00556 and HTS07014 at 100µM concentrations. HTS01682 showed maximum inhibition of ligation of about 91.22 ± 4.18% while NRB00556 showed 86.5 ± 4.82% inhibition and TG00130 showed 56.75 ± 8.51% inhibition of ligase I activity, respectively (lanes 3, 6 and 7). HTS07250 and HTS07014 did not inhibit hligI significantly (lanes 4 and 5). Lane 1 was a control with no ligase and lane 2 contained only the ligase I protein but no inhibitor.”

3.6. In vitro screening of putative hligI inhibitor

Out of the 30 compounds selected from in silico screening, 5 compounds were tested compounds (HTS01682, HTS07250, HTS07014, NRB00556 and TG00130) in our in vitro studies. We first performed fluorescent based ligation assays at 100µM concentration for hligI (Figure 4). HTS01682 demonstrated the maximum inhibition of ligation of 91.22 ± 4.18% while NRB00556 with 86.5 ± 4.82% inhibition and TG00130 with 56.75 ± 8.51% inhibition respectively (lanes 3, 6 and 7) also showed significant inhibition of ligation. The compounds HTS07250 and HTS07014 however, did not inhibit
hligI significantly (lanes 4 & 5). Lane 1 was a control with no ligase and lane 2 contained only the hligI and DMSO but no inhibitor.

“Figure 5. Concentration dependent ligation inhibition activity for the active compounds HTS01682 and NRB00556 at 25, 50 and 100 µM concentrations against various ligase proteins. The top left panel clearly demonstrates that both the compounds are specific inhibitors of human ligase I protein. No significant inhibition was observed for the activity of hLigIIIβ (left panel, second from top), hLigIV/XRCC4 (left panel, third from top) or for the T4 DNA ligase protein (left panel, bottom figure). The graphs on the right also clearly show that both inhibitors inhibit human ligase IIIβ activity by 50% and 30% respectively.”

3.7. Specificity and kinetics of ligase inhibitors
We performed a concentration dependent ligation inhibition assay for the active compounds against various human and non-human DNA ligase proteins such as hligI, hligIIIβ, hligIV/XRCC4 and T4 DNA ligase. We also checked whether the inhibitors exhibited a concentration dependent (25, 50, 100µM) ligation inhibition activity. The active compounds (HTS01682 and NRB00556) were tested as shown in the Figure 5. We observed that these compounds demonstrated a dose dependent ligation inhibition activity against hligI with 50% ligation inhibition at 24.93±3.71µM for HTS01682 and 37.56 ± 6.97µM for NRB00556 respectively. Although the compounds have some overlapping activity against hligIIIβ, they are completely inactive against hligIV/XRCC4 and T4 DNA ligases. Hence these compounds specifically inhibit the activity of hligI with some overlapping activity against hligIIIβ.
3.8. Mechanism of inhibition of ligase inhibitors

There are several modes by which a compound can inhibit ligation. Two of the most common modes are by direct interaction with the ligase protein (hence occluding the binding of ligase from the substrate DNA) or by intercalation with DNA, thereby restricting the access of DNA from the ligase. To check direct binding, we performed Electrophoretic Mobility Shift Assays (EMSA). This experiment works on the principle that the ligase enzyme binds to DNA and forms a complex that runs higher up on a gel than the DNA when run alone (Figure 6A and B lane 1 versus lane 2). Now if an inhibitor binds, it can do so at either the DNA binding domain of the ligase protein (hence occluding its binding with DNA) or bind to an allosteric site in which case it may form a bigger ligase-DNA-inhibitor complex. As seen in Figure 6 A, HTS01682 reduces the hligI-nicked DNA complex at higher concentrations (lanes 3-5). Such a reduction in hligI-DNA complex can occur only due to competition between inhibitor and substrate DNA for binding with DBD of the ligase protein. In contrast, in Figure 6 B, at increasing concentrations (lanes 3-5), NRB00556 stabilizes and increases the hligI-nicked DNA complex. The explanation in this case is that NRB00556 stabilizes the ligase-DNA complex so that the ligase cannot complete the ligation step, although it maybe able to bind the DNA.
compared with the negative control ampicillin (lane 2), the vehicle control DMSO (lane 3) or DNA alone (lane 1). However, the addition of the known DNA intercalator doxorubicin to the same DNA led to a hindrance in migration as seen by the higher migrating band in lane 6 (positive control).”

To rule out other modes of inhibition of ligation by these inhibitors, we also checked the DNA intercalation property of these inhibitors. This experiment works on the principle that any molecule that intercalates with DNA will pose a hindrance to the movement of DNA in agarose gel, as compared to when there is no intercalator in the DNA. In Figure 7, unlike the known DNA intercalator doxorubicin, there was no hindrance seen in the movement of DNA in the presence (even at 200µM concentration) of hligI inhibitors. Hence we can infer from these experiments that the inhibition of ligation shown by the active compounds is through a direct interaction between the compound and the protein (HTS01682) or protein-DNA complex (NRB00556) and not through DNA intercalation.

“Figure 8. (A) The two active compounds HTS01682 (Yellow) and NRB00556 (Pink) in the hLigI. DNA is shown in cyan ribbon and the molecules occupy the location where DNA is present in crystal structure. The Docking of selected compounds in DNA binding site of hLigI is shown in Fig 8(B) HTS01682 and Fig 8(C) NRB00556; showing hydrogen bonds in black dashed lines. The protein residues are displayed in cyan colour.”

To explore the mechanism of inhibition in more detail the binding modes of the two most active compounds were analysed comprehensively. Figure 8(A) demonstrates the binding modes obtained from docking of the two biologically active molecules in the DNA binding region of the catalytic site of ligase I and forming hydrogen bond with DNA binding residues including Arg451, Gly453, Lys768, and Lys770. The overall pattern of binding was found to be similar for these compounds. A visual inspection of
the docked complexes shows that the selected molecules accommodated fairly well in the binding pocket and occupied the same position. Figure 8(B) shows the binding mode of HTS01682. A thorough inspection of the complex was done in order to identify those interactions which might play an important role in binding DNA to protein and in turn maybe important for ligation. By preventing contact of DNA from these residues, the ligation process may be inhibited. The binding mode of the compound HTS01682 was analysed in terms of residues of hLigI that are involved in the interaction with DNA. Side chain nitrogen of Arg449 forms two hydrogen bonds with CN group attached to pyrimidine moiety of this molecule. The side chain nitrogen of Arg451 also forms a hydrogen bond with HTS01682. Another important interaction was found between the backbone nitrogen of Gly453 and HTS01682. Side chain nitrogen of Lys770 of hLigI was found to form two hydrogen bonds with the pyrimidine moiety of HTS01682. All these residues make direct contact with DNA in the crystal structure of hLigI.

Figure 8(C) shows the binding mode of NRB00556. As can be seen from the figure, a hydrogen bond forms between the side chain of Arg449 and carboxyl oxygen attached to one of the benzene rings of NRB00556. The backbone nitrogen of Arg451 also makes hydrogen bonds with carboxyl groups attached to a second benzene ring of this compound. Lys770 also forms hydrogen bonds with one carboxyl group attached to the mentioned benzene ring. Gly453 is involved in forming a hydrogen bond with carboxyl group of another benzene moiety of the compound. Leu454 also forms a hydrogen bond with this carboxyl group. It also interacts with hLigI through hydrogen bonds with Arg768 and Gly448. As discussed earlier, these residues are seen to form hydrogen bonds with DNA in hLigI crystal structure.

"Table 4. Antiproliferative activity of the ligase inhibitors HTS01682 and NRB00556 was tested by SRB assay at 10µM concentration against an array of cancer cell lines. Both compounds showed maximum anti-proliferative activity against the colon cancer cell line DLD-1. The HTS01682 molecule clearly showed greater anti-proliferative activity of 75.4% as opposed to 57.6% anti-proliferative activity for NRB00556 at 10µM concentration. The HTS01682 molecule also showed an appreciable anti-proliferative activity of 40.8% against the prostate cancer cell line PC3. No anti-proliferative activity was observed for either compound against any of the other cell lines tested, hence indicating its selective nature."

<table>
<thead>
<tr>
<th>Compound s at 10µM conc.</th>
<th>% inhibition in prostate Cancer (PC3)</th>
<th>% inhibition in Renal Cancer (786-O)</th>
<th>% inhibition in Colon Cancer (DLD-1)</th>
<th>% inhibition in Colon Cancer (SW-620)</th>
<th>% inhibition in Colon Cancer (A172)</th>
<th>% inhibition in Breast Cancer (MCF-7)</th>
<th>% inhibition in Lung Cancer (A549)</th>
<th>% inhibition in Liver Cancer (PLC/PRF/5)</th>
<th>% inhibition in Normal fibroblast (NIH/3T3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HTS01682</td>
<td>40.8</td>
<td>26.5</td>
<td>75.4</td>
<td>19.7</td>
<td>-14.3</td>
<td>5.6</td>
<td>14.9</td>
<td>21.5</td>
<td>15.24</td>
</tr>
<tr>
<td>NRB00556</td>
<td>10.0</td>
<td>8.7</td>
<td>57.6</td>
<td>5.1</td>
<td>-1.7</td>
<td>-2.6</td>
<td>1.8</td>
<td>17.2</td>
<td>3.19</td>
</tr>
</tbody>
</table>

3.9. Anti-proliferative activity of ligase inhibitors

The effect of the ligase inhibitors HTS01682 and NRB00556 were tested against the growth and proliferation of various cancer cell lines such as PC3, 786-O, DLD-1, PLC/PRF/5, SW620, A172, MCF-7 and A549 by following the NIH recommended SRB assay at 10µM concentration. The results from the SRB assay are tabulated in Table 4. We observed that at the tested 10µM concentration, both compounds demonstrated maximum anti-proliferative activity against DLD-1 (colon cancer) cell line with IC_{50} values of 6.67µM and 8.78µM respectively. Colon cancer is a dreaded form of cancer affecting a large number of people above 40 years of age. Most colon cancers originate
from small, noncancerous (benign) tumors called adenomatous polyps that form on the inner walls of the large intestine. Some of these polyps may grow into malignant colon cancers over time if they are not removed during colonoscopy. Colon cancer cells will invade and damage healthy tissue that is near the tumor, causing many complications. Once diagnosed with colon cancer, survival depends on what stage of cancer the person is diagnosed with and whether the person responds to available treatment. The present treatment options are often non-specific or have toxic side effects. The identification of these inhibitors in this study thus has the potential to be developed as a molecularly targeted therapy for the treatment of patients with colon cancer.

4.0 Conclusions

In this study, pharmacophore modelling and virtual screening led to the identification of novel inhibitors of human DNA ligase I protein. We have applied a multi-step virtual screening protocol including 3D-pharmacophore search, molecular docking and calculations of binding energy to prioritize the virtual screening hits against hlgi. Some of the top scoring hits were subjected to ligation assays where two compounds HTS01682 and NRB00556 inhibited DNA ligation activity significantly in the 25-35µM range. Hence our results demonstrate the efficiency of pharmacophore based virtual screening for the identification of potential hLigI inhibitors that also demonstrate potential anticancer activity against a colon cancer cell line.

Acknowledgements

Work reported in this project is supported by grants from CSIR network projects GENESIS (BSC0121) and BioprosPR (BPR0006), DST-FAST track project (SB/FT/LS-163/2012), and DBT-RGYI project (BT/PR6421/GBD/27/436/2012). SK and DKS acknowledge DBT project SSP0025 and CSIR-JRF for their fellowships. This manuscript is a CSIR-CDRI Communication No. 8628

References


42. Güner, O. F. (Ed.). Pharmacophore perception, development, and use in drug design (Vol. 2). Internat’l University Line. **2000**.

43. UNITY, SYBYL7.1, **2004**, Tripos, Inc., St. Louis, MO.


54. Moitessier, N., Henry, C., Maigret, B., & Chapleur, Y. Combining pharmacophore search, automated docking, and molecular dynamics simulations