Chapter 3

NAD$^+$-dependent DNA ligase:
Strategies for inhibitor design and optimization
Chapter 3

Strategies for Inhibitor Optimization

3.1 Introduction

In the previous chapter, we reported the results of our efforts in utilizing the crystal structure of the adenylation domain of MtuLigA (Srivastava, et al., 2005b) and virtual screening strategies to identify novel classes of compounds (Srivastava, et al., 2007; Srivastava, et al., 2005a; Srivastava, et al., 2005b) capable of inhibiting the bacterial enzyme specifically. The inhibitors are furthermore capable of distinguishing between the human DNA ligase I and mostly act by competititively binding to the NAD\(^+\) binding site. NAD\(^+\)-dependent enzymes exhibit multidomain architectures with large interdomain movements during the enzyme action. Conceptually it should be possible to block some of the essential structural changes or block the functions of the individual domains too to result in novel inhibitors with different modes of LigA inhibition.

In the present Chapter, the ongoing efforts to identify new inhibitors of the enzyme; the challenges faced, and proposed strategies to improve both the potency and the specificity of a designed inhibitor for MtuLigA are summarized. We also report our efforts in optimizing the efficacy of the inhibitors reported in the earlier Chapter. In the quest to identify better MtuLigA inhibitors, there are three issues at stake, viz., the identified inhibitors should, in the first instance, be able to distinguish between NAD\(^+\) and ATP-dependent ligases. The present crop of inhibitors mostly block the binding of the cofactor, and thereby the enzyme activity. In the second instance, the inhibitor should possess high affinity (specificity) for a particular pathogen/bacterial species, in contrast to being a general anti-bacterial compound; given that the residues in the LigA binding sites are well conserved (Doherty & Suh, 2000); in fact, five of the six conserved sequence motifs form part of the site (Fig 3.1). In the third instance, identification of diverse scaffolds with better drug –likeliness and with alternate modes of enzyme inhibition was considered.

Given the conserved nature of the cofactor-binding site, most of the inhibitors are expected to exhibit some degree of general antibacterial activity too. Therefore, better inhibitor development was also focused on improving the specificity of the compounds for MtuLigA. Three such strategies are discussed below; one approach deals with utilizing the active site structural water molecules for inhibitor design, while the other
Chapter 3 Strategies for Inhibitor Optimization

approach involves the development of inhibitors that can bind to other regions of the molecule, in this case the BRCT domain, and block subsequent catalytic steps. Third approach involves fragment-based lead identification where inhibitors with diverse scaffolds can conceptually be built from a comparatively smaller set of compound fragments.

Fig 3.1 Conserved sequence motifs in DNA ligases. The adenylation domain contains five of six conserved sequence motifs in NAD\(^+\) ligases. The alignment includes LigA encoded by \textit{M. tuberculosis} (Pdb code: 1ZAU), \textit{T. filiformis} (1V9P), \textit{B. stearothermophilus} (1B04), \textit{E. faecalis} (1TAE) and human DNA ligase I (1X9N). The number of amino acid residues separating the motifs is indicated. The active site lysine is indicated by an arrow.

3.2 Experimental section

3.2.1 Modeling of the BRCT domain

The BRCT domain of LigA was disordered in the available crystal structures of the full-length \textit{T. filiformis} LigA and in the DNA complex with \textit{E. coli} LigA. The homology model for the BRCT domain of \textit{MtuLigA} was therefore generated using Modeller 6 v2 (Marti-Renom, et al., 2000) where the available NMR structure of the archaeal \textit{T. thermophilus} BRCT domain alone (PDB: 1L7B) was used as the template. The geometry of the model was optimized by subjecting to 100 iterations of minimization using the Tripos Force Field of the Sybyl7.1 software package (M/s. Tripos Inc.). The stereo-chemical quality of the model was verified using Procheck (Laskowski, et al., 1996) and structure was validated using ‘Protein Model Check’ utility implemented in the WhatIf server (Vriend, 1990). 87 % residues of the model are in most favored regions, 10.1% in additionally allowed regions and 2.9% in generously allowed region. The model showed no abnormality when checked for all angles (torsion, omega, Chi-1 and Chi-2),
bond lengths, unusual backbone conformation and rotamers. The Z-score, a parameter suggesting how well the backbone conformations of all residues are corresponding to the known allowed areas in the Ramachandran plot, was -2.4 and is within expected ranges for well-refined structures. Prior to the virtual screening experiments, hydrogen atoms were added to the model. The model of human pol λ BRCT domain was also generated as reported previously by the group (Takeuchi, et al., 2006). The three-dimensional structures of the BRCT domains in human XRCC1 (Pdb code: 1CDZ) has been used as template to prepare this model.

3.2.2 Virtual Screening Protocols

Ligand Databases used

Two in-house ligand libraries were used; first one CDRI Database which is described in Chapter 2 and the second FragDB (fragment database): a virtual fragment library which is recently made available by our group to the Open Source Drug Discovery Initiative (OSDD) Network, CSIR (http://www.csir.res.in) and can be accessed at: http://sysborgtb.osdd.net/bin/view/OpenLabNotebook/FragDBVirtualFragmentLibrary. All the fragments were built and optimized using the Builder module of Insight II (M/s. Accelrys Inc.). This library consists of about 3400 drug-like fragments and is described in detail in the result section of this chapter.

The commercially available Ludi/CAP 2002 database (M/s. Accelrys Inc.) was also used for the virtual screening experiments. CAP (Chemicals Available for Purchase) database offered by Accelrys Inc. is a fragment database consisting of approximately 75,000 commercially available compounds and can be used directly with LUDI (Bohm, 1992) - a de novo design tool.

Docking using the AutoDock program

The detailed procedure for automated docking with AutoDock has been described in Chapter 2.

De novo design with LUDI

The Ludi module, interfaced with Insight II package (M/s. Accelrys Inc., 2000), was used to identify fragments and design ligands that would potentially overlap with the active
Strategies for Inhibitor Optimization

Chapter 3

site water. Ludi is based on the fragment approach, whereby it saturates small fragments into the clefts of the target sites in such a way that hydrogen bonds can be formed with the enzyme, and hydrophobic pockets are filled with hydrophobic groups. These fragments might be linked together by using Ludi in “link mode”. When structure of receptor is known, the Ludi module can be used in for receptor-based design (in ‘Targeted mode’) and when only ligands are known, the Ludi module can be used for analog-based design (in ‘Analogue mode’).

Generating novel ligands de novo: The binding pockets within the modeled BRCT domain were defined by the conserved glycine residues where the active site cavity was identified using SiteID search module interfaced with Insight II (M/s. Accelrys Inc., 2005) and in case of adenylation domain the active site was defined by conserved water cluster. The Ludi parameters were reviewed, and modified, and the Ludi library was also specified. The program was executed in the ‘Targeted Mode’. The ‘Center of Search’ parameter with a radius sphere of 5 Å was specified, based on two conserved glycine residues in the BRCT domain and also water clusters within the adenylation domain (Pdb: 1ZAU). The docked conformations of the identified hits were analyzed, and Ludi scores tabulated within the modeling and simulation environment of Insight II. Autodock v3 (Morris, et al., 1998) was also used, in several instances, for examining the poses of potential inhibitors with the enzyme.

Docking with FlexX
The detailed procedure for automated docking with FlexX has been described in Chapter 2.

Water cluster analysis
Crystal structures of the adenylation domain of LigA from B. subtilis, E. faecalis, T. filiformis and MtuLigA were used for structural superposition and water cluster analysis. These models include the open and bound forms of NAD⁺-dependent DNA ligase encoded by E. faecalis (PDB code: 1TA8 and 1TAE), T. filiformis (PDB code: 1V9P) and B. stearothermophilus (PDB code: 1B04). The chosen structures were superimposed structurally using the Homology module of InsightII (M/s. Accelrys Inc., 2000) to transform the water coordinates into the reference frame of the MtuLigA model. Initially,
the distance matrix between all the water molecules present in the binding site of the enzyme was calculated and clustering was done by placing the two closest elements together into a cluster; if the maximal distance is less than 2.4 Å then the elements were grouped into a cluster. This process was repeated until no further elements could be clustered without exceeding the selected maximum distance. This procedure resulted in clusters with a maximum inter-water distance of 2.4 Å. The centroids of each cluster were then used to position water molecules for calculations.

3.2.3 Enzyme assays

All assays have been carried out using procedures previously described in Chapter 2.

Illustrations

Images were made using PyMol (http://www.pymol.org), Photoshop (Adobe Systems) and InsightII (M/s. Accelrys Inc.) and ChemDraw (http://www.chembridgesoft.com).
Chapter 3  Strategies for Inhibitor Optimization

3.3 Results

3.3.1 Optimization of Glycosylamines and Glycosylureides inhibitors

The inhibitors, which were identified from these compound classes, were of rather larger molecular weight. The first strategy was therefore to trim down these compounds within the allowed spectrum of synthetic feasibility. Particularly the side chain extensions like ethyl alcohol and \(-\text{OCH}_2\text{Ph}\) groups and the size of central hydrophobic linker \((\text{CH}_2)_n\) was reduced first while the parent backbone was conserved keeping in mind that it might be responsible for the biological activity of the molecule. These modified compounds were built and optimized by the Builder module of the InsightII. The designed compounds were subjected to a check for their binding affinities with the enzyme through docking approaches and then were prioritized for synthesis (Table 3A). The docking calculations were performed with Autodock v3 and the binding affinities shows that these smaller compounds exhibit less affinity \textit{in silico} compared to the parent compound. These were however synthesized because the smaller size actives if identified could be subjected to better optimization strategies. When the modified compounds of these classes were subjected to experimental testing, they showed only negligible inhibition of the enzyme. The plausible reason for this could be that the actual pharmacophore or the core structure responsible for the inhibition is yet to be identified correctly or the molecules adopt inactive conformations preferentially in the absence of the ‘trimmed’ molecular extensions. Co-crystallization of the enzyme with the compounds was not successful and hence a co-crystal structure was not available for use as the basis for optimization. Current approaches involve further modifications to identify the active pharmacophores in the respective compound classes.

3.3.2 Optimization of Indole Derivatives

The indole derivatives are relatively better amenable for optimization compared to the glycosylamines and glycosylureides because these are smaller in size, non-toxic and selectively inhibit \(\text{MtuLigA}\) with ~15 fold specificity. These compounds were optimized following a similar strategy as above where modifications to the known active structure (Table 3B) were quickly tested \textit{in silico} with molecular docking approaches before...
compound synthesis. The predicted compounds are currently being synthesized by Dr. K. Hajela's group, MPC Division, CDRI, Lucknow, India and will be subsequently assayed.

TABLE 3A. Optimization of the Glycosylamine scaffold: The parent scaffold (numbered 1) along with the positions where modifications were carried out (marked with square brackets) is depicted. The *in silico* optimized compounds are also shown (numbered 2-7). The docking energies were calculated using AUTODOCK v3.

<table>
<thead>
<tr>
<th>S.No.</th>
<th>Optimized Glycosylamines</th>
<th>Docking Energy (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><img src="image1.png" alt="Image" /></td>
<td>-15.8</td>
</tr>
<tr>
<td>2</td>
<td><img src="image2.png" alt="Image" /></td>
<td>-11.37</td>
</tr>
<tr>
<td>3</td>
<td><img src="image3.png" alt="Image" /></td>
<td>-10.29</td>
</tr>
</tbody>
</table>

1,3-Bis-(6-hydroxy-2,2-dimethyl-tetrahydro-furo[2,3-d][1,3]dioxol-5-ylmethyl)-urea

1-(6-Hydroxy-2,2-dimethyl-tetrahydro-furo[2,3-d][1,3]dioxol-5-ylmethyl)-3-phenyl-urea
<table>
<thead>
<tr>
<th></th>
<th>Chemical Structure</th>
<th>Molecular Formula</th>
<th>pIC₅₀</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td><img src="image1" alt="Chemical Structure 4" /></td>
<td>1-(4-Fluoro-phenyl)-3-(6-hydroxy-2,2-dimethyl-tetrahydro-furo[2,3-d][1,3]dioxol-5-ylmethyl)-urea</td>
<td>-9.16</td>
</tr>
<tr>
<td>5</td>
<td><img src="image2" alt="Chemical Structure 5" /></td>
<td>1-(4-Fluoro-phenyl)-3-(6-hydroxy-2,2-dimethyl-tetrahydro-furo[2,3-d][1,3]dioxol-5-ylmethyl)-thiourea</td>
<td>-10.96</td>
</tr>
<tr>
<td>6</td>
<td><img src="image3" alt="Chemical Structure 6" /></td>
<td>1-(6-Hydroxy-2,2-dimethyl-tetrahydro-furo[2,3-d][1,3]dioxol-5-ylmethyl)-3-(4-methoxy-2-nitro-phenyl)-thiourea</td>
<td>-9.83</td>
</tr>
<tr>
<td>7</td>
<td><img src="image4" alt="Chemical Structure 7" /></td>
<td>4-[(6-Hydroxy-2,2-dimethyl-tetrahydro-furo[2,3-d][1,3]dioxol-5-ylmethyl)-(3-oxo-butyl)-amino]-butan-2-one</td>
<td>-9.59</td>
</tr>
</tbody>
</table>
Table 3B. Modifications predicted for Indole Derivatives class of compound (II) are shown as R1 and R2 side chains on the parent scaffold. The R1 side chain mimics the interaction of first group of water cluster (W1-W8) while R2 side chain mimics the interaction of second group of water cluster (W8-W13) (Fig 3.2). The water clusters are discussed in the next part of this chapter. The docking energies were calculated by Autodock v3.

<table>
<thead>
<tr>
<th>R1 SIDE CHAIN</th>
<th>Docking energy (kcal/mol)</th>
<th>Parent Molecule</th>
</tr>
</thead>
<tbody>
<tr>
<td>-NH₂ PRIMARY AMINE</td>
<td>-14.18</td>
<td></td>
</tr>
<tr>
<td>-OH HYDROXYL</td>
<td>-12.91</td>
<td></td>
</tr>
<tr>
<td>SULFONAMIDE</td>
<td>-15.68</td>
<td></td>
</tr>
<tr>
<td>AMIDINE</td>
<td>-17.20</td>
<td></td>
</tr>
<tr>
<td>AMIDE</td>
<td>-15.42</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>R2 SIDE CHAIN</th>
<th>Docking energy (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>-OH HYDROXYL</td>
<td>-15.34</td>
</tr>
<tr>
<td>-NH₂ PRIMARY AMINE</td>
<td>-14.33</td>
</tr>
</tbody>
</table>

Docking energy: -12.71
3.3.3 Water cluster analysis

The roles of conserved water molecules in crystal structures have been analyzed in many structurally homologous protein families including those involving thrombin, trypsin, MHC class-I, legume lectins and alanine racemase. Such structurally conserved water molecules, especially those in the active site, have been exploited to increase the inhibitory potency of designed compounds (Loris, et al., 1994; Mustata & Briggs, 2004; Ogata & Wodak, 2002; Sanschagrin & Kuhn, 1998). Often a polar atom like oxygen or nitrogen of the inhibitor/ligand can take the place of the displaced water oxygen and replace its interactions with the protein through direct or water-mediated interactions (Ravishankar, et al., 1997; Ravishankar, et al., 1999). This reportedly contributes up to 20-fold increase in affinity. The specificity of a known inhibitor, in this approach, can be enhanced by engineering polar atoms at appropriate locations in it to form hydrogen bonds with structural water in the binding site (Frederick, et al., 2006), or by displacing them from the binding site (Chen, et al., 1998).

We therefore analyzed the crystallographic waters in the NAD$^+$ binding site from available structures of the enzyme to examine the possibility of utilizing such an approach in our inhibitor design experiments. There are structures of the adenylation domain from *B. stearothermophilus*, *E. faecalis*, *T. filiformis*, and from *M. tuberculosis* in the protein databank (http://www.rcsb.org). In *E. faecalis*, structures are available with bound NAD$^+$ and also in an open conformation without NAD$^+$ in the active site (PDB codes 1TAE and 1TA8, respectively). In *B. stearothermophilus* the available structure (PDB code: 1B04) is that of the unliganded form while the *T. filiformis* structure has bound AMP (PDB code 1V9P). The *MtuLigA* structure (PDB code 1ZAU) is with bound AMP but no water molecules have been added because of the modest resolution. There are four chains in the asymmetric unit in the NAD$^+$ bound structure in *E. faecalis* while the *B. stearothermophilus* and *T. filiformis* structures contain two chains each. Accordingly the individual subunits from all structures were superposed as explained in experimental section. Table 3C summarizes the analysis while Figure 3.2 is a graphic representation of the network of bound water and distances between them.
Fig 3.2 Water clusters in the *MtuLigA* adenylation site (split-stereo view). Water molecules are represented as blue spheres while interacting atoms have standard colors. Polar interactions less than 3.6 Å with interacting residues and bound NAD⁺ (blue) or AMP (red) are shown and denoted by dotted lines as also interactions between the centroids of the respective water clusters. A SO₄²⁻ ion occurring in 1TAE is shown as a green sphere. Interacting residues and water clusters are numbered for clarity.
Table 3C. Water Cluster Analysis

Occupancy of water molecules in the NAD binding pocket of LigA. Available crystal structures were superposed as explained in experimental section. W1-W17 are water clusters.

<table>
<thead>
<tr>
<th>PDB code</th>
<th>ITAE-A</th>
<th>ITAE-B</th>
<th>ITAE-C</th>
<th>ITAE-D</th>
<th>ITA8-A</th>
<th>ITA8-B</th>
<th>1BO4-A</th>
<th>1BO4-B</th>
<th>1V9P-A</th>
<th>1V9P-B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution (Å)</td>
<td>2.8</td>
<td>2.8</td>
<td>2.8</td>
<td>2.8</td>
<td>1.8</td>
<td>2.7</td>
<td>2.7</td>
<td>2.9</td>
<td>2.9</td>
<td></td>
</tr>
<tr>
<td>Cofactor</td>
<td>NAD</td>
<td>NAD</td>
<td>NAD</td>
<td>NAD</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>AMP</td>
<td>AMP</td>
<td></td>
</tr>
<tr>
<td>W1</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>W2</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>W3</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>W4</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>W5</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>NP</td>
<td>NP</td>
<td></td>
</tr>
<tr>
<td>W6</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>NP</td>
<td>NP</td>
<td></td>
</tr>
<tr>
<td>W7</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>NP</td>
<td>NP</td>
<td></td>
</tr>
<tr>
<td>W8</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>W9</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>W10</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>W11</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>W12</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>W13</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>W14</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>W15</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>W16</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>W17</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>NP</td>
<td>NP</td>
<td></td>
</tr>
</tbody>
</table>

+ refers to the occurrence of a particular water while - refers to its absence; NP refers to those cases where the occurrence of a water molecule is not possible because of the binding of either AMP or NAD+ in the respective structures being superposed. The relatively lesser occupancy of the water molecules in the remaining structures compared to ITA8 is attributed to the modest resolutions at which the other structures were resolved.
Chapter 3 Strategies for Inhibitor Optimization

A total of 17 water clusters (W1-W17) were observed after superposition (Table 3C). These clusters by themselves fall into three networked groups. The first group consists of W1 to W7 (Fig 3.2). This group of clusters occurs at the top of the adenosine nucleotide moiety of bound AMP/NAD<sup>+</sup>. The second group of clusters consists of W8 to W13 while the third group consists of clusters between W14 and W17. The first two groups of water clusters are part of the structural water attached to residues in the AMP binding sub-domain while the third group of water clusters contains water molecules which are attached to residues from NMN recognizing sub-domain (Gajiwala & Pinko, 2004). These two sub-domains undergo large relative spatial movements during enzyme action.

It is interesting to note that W5, W6 and W7 have similar spatial dispositions as the N6, N1, and N7 atoms of the adenosine nucleoside moiety (Fig 3.2) and mimic their interactions, respectively. W16 has the same spatial disposition as O3P atom of AMP in the MtuLigA structure (PDB: 1ZAU). PO3 of NMN in EfaLigA (PDB: 1TA8) and W14 share the same spatial disposition. Thus structural water molecules mimic the interactions of the polar atoms of the co-factor in their absence. The spatial region/cavity occupied by W1-W7 consists of both conserved and non-conserved residues. This region can therefore be potentially exploited in rational approaches in improving inhibitor potency and specificity for a given bacterial species. Another such region which can also be potentially exploited is the one occupied by W8-W13.

The indole derivatives were optimized based on these conserved water cluster analysis where R1 and R2 side chains were designed to mimic first two groups of water cluster (Table 3B).

Virtual screening approaches using Ludi with the commercial compound library, the CAP database, identified several fragments predicted to displace and mimic the interactions of active site water. All the three groups were explored for potential binding ligands. Some of the identified fragments are 1, 2, 3, 4-tetrahydroquinoline; 1-ethylpiperidine and N-methylformanilide. These can displace conserved water interactions as shown in Figure 3.3B. These compounds have been procured and are now being tested for enzymatic inhibition. One of the compound 1,2,3,4-tetrahydroquinoline has indeed shown in vivo inhibition with a MIC value of 100 µg/ml and in vitro inhibition assays are still to follow.
Fig 3.3 (A) Conserved water clusters in the active site of LigA. Adenosine monophosphate (AMP) is shown in stick form and the water clusters that form networks amongst them are indicated by dashed circles. Some interacting residues are shown for clarity. (B) Some fragments identified by virtual screening from a commercial library that are predicted to be able to displace water clusters. These fragments will now be ‘stitched’ together and evaluated. (C) The A, B and C structures occupies the three water clusters from top to bottom as shown in (A). The distance between Compound A and B where these can be potentially linked is 6.2 and 6.5 Å. The potential linker candidate of same size range is also shown.
Once the enzyme assays confirms the inhibition, the compound fragments can be subsequently linked by standard linking approaches. For another set of fragments shown in Figure 3.3C, 3-methylamino-cyclohex-2-enone (A), 2-(1H-indol-3-yl)-1-methyl-ethylamine (B), 6-Methyl-1-phenyl-2,3-dihydro-1H-pyrrolo[2,3-b]quinolin-5-ylamine (C); propenyl-benzene could be the possible linker. However, the synthesis of the compounds awaits the outcome of the preliminary experimental results with the fragments.

3.3.4 BRCT domain and inhibitor development

LigA is a highly modular enzyme with the BRCT domain occurring at the C-terminus of the enzyme (Fig 2.2). More work is necessary to actually elucidate the role of this domain in detail. Its deletion resulted in threefold reduction in activity of *E. coli* LigA (Wilkinson, *et al.*, 2001), whereas the *T. filiformis* LigA retains activity even after its deletion (Jeon, *et al.*, 2004). Even in the enzyme found in viral sources, the same situation exists, as this domain is absent in the active enzymes from *A. moorei* (Sriskanda, *et al.*, 2001) and *M. manguinipes* (Lu, *et al.*, 2004), while the LigA from *mimivirus* possesses a BRCT domain, essential for its activity (Benarroch & Shuman, 2006).

Our group had earlier identified that the BRCT domain of *MtuLigA* is essential for activity (Srivastava, *et al.*, 2007), in contrast to some other characterized LigA, including those *T. filiformis*. Another group has also shown through mutational analysis of BRCT domain, that it is essential for effective nick sealing in *E. coli* (Wang, *et al.*, 2008). Thus, inhibitors designed to prevent BRCT domain interactions with the adenylation site or DNA would be conceptually novel and block other steps in the enzyme’s action. Previously, mutations in a homologous enzyme (Feng, *et al.*, 2004) had identified two glycine residues as being important for its activity. We accordingly modeled the *MtuLigA* BRCT domain and used the structure in virtual screening experiments to identify potential binders. Two potential binding pockets were identified based on two conserved glycine residues in archeal *T. filiformis* LigA (referred as Site I for G34 and Site II for G57). Compounds designed to bind to a pocket near the corresponding glycine
residues in the model (Fig 3.4) should expectedly be candidates for further testing in \textit{in vitro} assays. These compounds identified from the CAP database belong primarily to azine and azonane type of scaffolds (Table 3D). These compounds have been procured and are being evaluated in an ongoing program. In another report (Takeuchi, \textit{et al.}, 2006), specific inhibitors of the human pol $\lambda$ BRCT domain were identified. Our modeling and docking simulations involving a curcumin derivative, monoacetylcurcumin (Fig 3.4C), suggest that the same compound should bind to different regions of the BRCT domain in the respective proteins.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{fig3-4}
\caption{(A) Sequence alignment of BRCT domains of \textit{MtuLigA} and \textit{Thermus thermophilus} LigA (PDB code: 1L7B). Two conserved and essential glycine residues are circled and indicated. (B) Cartoon representation of the modeled \textit{MtuLigA} BRCT domain. The essential glycine residues were used as the center-of-search during \textit{in silico} screening calculations. Some potential inhibitors in both regions are also depicted. (C) Potential binding sites at the modeled \textit{MtuLigA} BRCT domain (black ribbon) identified through binding site analysis (InsightII, M/s. Accelrys, San Diego, CA, USA). The sites are indicated by grey crosses. The modeled BRCT domain of human pol $\lambda$ (grey) is also shown to highlight possible differences in the structures of the BRCT domain. These differences are to be exploited in the design of species specific inhibitors. Docked monoacetylcurcumin is shown in ball-and-stick representation.}
\end{figure}
Table 3D. The compounds identified by *de novo* design tool Ludi as potential BRCT domain inhibitors are shown for both the sites I and II. The corresponding Ludi scores along with the structure are also shown for each of the compounds.

<table>
<thead>
<tr>
<th>Compounds for site I</th>
<th>Ludi Score</th>
<th>Compounds for site II</th>
<th>Ludi Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>7,7-Dimethyl-</td>
<td>468</td>
<td>2,4,6-Trimethyl-[1,3,5]triazinane</td>
<td>270</td>
</tr>
<tr>
<td>bicyclo[2.2.1]heptane-2,3-diamine</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Azocan-3-yl-methyl-</td>
<td>437</td>
<td>(Octahydro-quinolizin-1-yl)-methanol</td>
<td>247</td>
</tr>
<tr>
<td>amine</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[1,4,7]Triazonane</td>
<td>361</td>
<td>3,6-Diphenyl-1,2-dihydro-[1,2,4,5]tetrazine</td>
<td>246</td>
</tr>
</tbody>
</table>
3.3.5 Fragment-based lead identification

VS against a fragment library combined with X-ray crystallography offers a powerful alternative to traditional screening approaches for discovering new leads in drug discovery programs. A virtual fragment library of around ~3400 compounds was designed based on available drug-ring systems and functionalities often found in known drug molecules. The design principles follow those suggested in a recent report (Hartshorn, et al., 2005). In general, the fragments in these sets have low molecular weights between 100 and 250. The building blocks used are simple organic ring systems that occur in drug molecules (Bemis & Murcko, 1996; Bemis & Murcko, 1999) and other simple carbocyclic ring systems with reduced toxicity and increased synthesizability. The drug ring systems are shown in Figure 3.5.

![Drug Ring Systems](image)

Fig 3.5 (A). Drug Ring Systems (B). Simple carbocyclic and heterocyclic ring systems.

83
Chapter 3  Strategies for Inhibitor Optimization

The side chains used in development of virtual fragment library can be classified into primary and secondary side chains. The primary set of side chains are those that are observed frequently in drug molecules (Bemis & Murcko, 1996; Bemis & Murcko, 1999). While the lipophilic/secondary set of side chains are those that can be substituted as a secondary side chains. Most of these are lipophilic and are intended to pick up hydrophobic interactions in a protein binding site. The primary side chains are shown in Figure 3.6(A) and the secondary side chains for substituting side chains onto a framework for carbon atoms (C-substituents) and for nitrogen atoms (N-substituents) are shown in Figure 3.6(B).

The virtual library was generated by substituting each of the relevant side chains onto each of the ring systems. Each carbon atom was substituted by the side chains from the...
preferred group and by those from the secondary side chains. The nitrogen atoms were
only substituted from the group of N-side chains. Each ring system was only substituted
at one position at a time, with the exception of benzene and imidazole. The latter two
were disubstituted at all pairs of positions, with either two preferred side chains or one
preferred side chain and one secondary side chain.

Targeted sets: The virtual screening with AutoDock and FlexX was used to construct
target-specific sets of compounds for MtuLigA. The docking jobs were submitted using
the protocol described in experimental section. In first strategy, the fragments so
identified could be stitched together by suitable linker molecules with the help of a de
novo design tool like Ludi (M/s. Accelrys Inc.). The designer ligands so obtained can be
synthesized and taken up further for enzyme assays. In the second strategy, from this
target-specific set, the compounds are ‘cocktailed’ together, usually into groups of 4.

Cocktailling: The crystals can be typically soaked in this cocktail at room temperature for
1–24 h before being cryo-cooled using liquid nitrogen and subsequently transferred to
storage until required for data collection. The X-ray diffraction data so collected for
soaked crystals can be then analyzed for potential ligand binding. These resulting
protein/ligand complexes can be used further as the basis for the structure-guided
fragment optimization process using de novo design procedures.

The LigA specific set identified by Autodock includes combination of fragments with
ring systems like benzene, pyridine, benzamidine, benzylamines and sidechains like
acetamide, methoxy, sulfonamides, hydroxyl groups and hydroxymic acid. While the
LigA specific set identified by FlexX picked ring systems like indoles, tetrazoles,
pyridine, benzylamine, Benzyl-phenyl-amine, Phenylethyl-pyridine, phenyl benzamide
and sidechains of acetamide, pentazin, acetamidine. These compounds are shown in
Table 3E. The pyridines, indoles and hydroxymate based compounds have already been
identified as MtuLigA inhibitors. The hydroxymates as MtuLigA inhibitors described in
Chapter 2, were synthesized based on these predictions by Dr. R. P. Tripathi’s group,
MPC Division, CDRI, Lucknow, India.
Table 3E. The compound identified from FragDB (fragment database) as potential *MtuLigA* inhibitors are shown. The structures on the left were identified by FlexX and those on the right by Autodock docking programs. The respective docking scores are also shown.

<table>
<thead>
<tr>
<th>Compounds</th>
<th>FlexX Score</th>
<th>Compounds</th>
<th>Docking energy (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Structure 1" /> 2222</td>
<td>-25.7</td>
<td><img src="image2.png" alt="Structure 2" /> 3075</td>
<td>-13.88</td>
</tr>
<tr>
<td><img src="image3.png" alt="Structure 3" /> 2227</td>
<td>-26.4</td>
<td><img src="image4.png" alt="Structure 4" /> 3016</td>
<td>-13.49</td>
</tr>
<tr>
<td><img src="image5.png" alt="Structure 5" /> 2235</td>
<td>-23.6</td>
<td><img src="image6.png" alt="Structure 6" /> 2111</td>
<td>-13.31</td>
</tr>
<tr>
<td>Chapter 3</td>
<td>Strategies for Inhibitor Optimization</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>--------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td><img src="image1.png" alt="Chemical Structure 3411" /></td>
<td><img src="image2.png" alt="Chemical Structure 1981" /></td>
<td>-25.86</td>
<td>-11.52</td>
</tr>
<tr>
<td><img src="image3.png" alt="Chemical Structure 3422" /></td>
<td><img src="image4.png" alt="Chemical Structure 3189" /></td>
<td>-25.56</td>
<td>-12.08</td>
</tr>
<tr>
<td><img src="image5.png" alt="Chemical Structure 2094" /></td>
<td><img src="image6.png" alt="Chemical Structure 3030" /></td>
<td>-20.12</td>
<td>-12.87</td>
</tr>
<tr>
<td><img src="image7.png" alt="Chemical Structure 3418" /></td>
<td><img src="image8.png" alt="Chemical Structure 3099" /></td>
<td>-25.25</td>
<td>-12.89</td>
</tr>
</tbody>
</table>
### 3.4 Discussion

LigA are important enzymes with good potential for development as novel drug targets. The existing inhibitors, several of which were identified by our group, have been shown to be capable of distinguishing between the human and pathogen enzymes; this being an important step in specific inhibitor development. Given the conserved nature of the NAD$^+$ binding site amongst enzymes from different bacterial species, these inhibitors also show general antibacterial activity. One approach to improve their specificity and affinity for an enzyme from a specific bacterial species is to utilize the spatial
dispositions of active-site water, where inhibitors designed to mimic the interactions of displaced water oxygen should offer better activity than that of the first generation.

Another approach is based on the fact there is possibility of developing inhibitors which can disrupt other steps of the mechanistic cycle by binding to regions other than the adenylation domain. Our group and very recently Shuman’s group have also identified that the BRCT domain of LigA is essential for enzyme activity as well as for bacterial viability in *M. tuberculosis* and *E. coli* respectively (Shrivastava, *et al.*, 2007; Wang, *et al.*, 2008). Compounds designed to bind to the domain to prevent the interactions of its key residues should have better specificity for a given pathogen because of the proposed finer variations in individual enzyme action. Towards this end, we have modeled the *MtuLigA* BRCT domain and demonstrated that there are regions that could potentially be exploited for novel inhibitor development. The modeling and docking results involving this BRCT domain also suggest structural differences that can also be exploited in rational inhibitor design.

The third approach of fragment-based lead discovery offers, in principle, a more efficient sampling of chemical space of the enzyme active site and thus facilitating the design of more potent inhibitors. Although fragments have weak potency (>100 μM) but are considered to be efficient binders relative to their size and may therefore represent suitable starting points for the evolution of good quality lead compounds with better specificity.

Some of the tested second-generation inhibitors like those of the glycosylamines and glycosylureides compound classes were not better than the parent compounds. Some other inhibitors, especially from the indole derivative class, are expected to show more promising results. Obviously inhibitor design and optimization is a long-range program and the diverse approaches listed here will hopefully result in compounds with better properties.
Chapter 3

Strategies for Inhibitor Optimization

3.5 References


Chapter 3

Strategies for Inhibitor Optimization


Sybyl 7.1 (2004) TRIPOS Inc, 1699 South Hanley Road, St. Louis, Missouri 63144, USA


91