Virtual screening and molecular docking of Anti-Antileishmanial for selected pharmacophore for visceral Leishmaniasis

SindhuPrava Rana, R Sivaperumal*

Department of Biotechnology, Sri Satya Sai University of Technology & Medical Sciences, Sehore-466001

ABSTRACT

Objective: DNA amplification of Cysteine protease of Leishmania donovani and study the interaction of cysteine protease inhibitors, antileishmanial compounds with cysteine protease receptor in various computational programs.

Materials and methods: Cysteine protease DNA of Leishmania donovani was amplified by PCR. The sequence of cysteine protease has been modeled and docked with suitable inhibitors by using various servers and computational tools. The model was designed, compared and validated by DOPE and Verify 3D scores. The model and the compound interaction were studied by LibDock and other programs.

Results: Cysteine protease DNA of Leishmania donovani was successfully amplified by PCR. The structural modeling was done to achieve effective enzyme inhibition, inhibitors block the binding sites of that protein. Homology modeling of cysteine protease has been done and docked with suitable inhibitors by using various servers and computational tools. The model was designed, compared and validated by DOPE and Verify 3D scores by using DSv3.5. Licochalcone-a alone showed 37 LibDock conformations with 6 different poses, which were suitably docked at the site 1 with hydrogen bond formation. The study would help to design the novel drugs in respect of resistant one for the treatment of harmful visceral Leishmaniasis.

Conclusion: The molecular interaction of vinyl sulfones, hydrazide derivatives, antileishmanial drugs molecules and carbohydrazide derivatives have exhibited ideal molecular interaction with cathepsin B, a cysteine protease of L. donovani, amino acids such as Cys29, His88 and Asn208 has been found to be active residues. Licochalcone-a and hydrazide derivative may become future antileishmanial compounds, which needs to be tested in vitro and vivo.

Keywords: Cysteine Protease, Vinyl Sulfone, Hydrazide, Antileishmanial Drugs, Licochalcone, Visceral Leishmaniasis.

INTRODUCTION

Visceral Leishmaniasis is the most dreaded and destructive amongst all the various forms of leishmaniasis. VL is also known as Kala- Azar, Black Sickness, Black Fever, Burdwan fever, Dumdum fever or Sarkari Bimari etc.3-4. VL is the most severe form of disease and if left untreated, is usually fatal. The parasite is responsible for a spectrum of clinical syndromes, which can, in most extreme cases, move from an asymptomatic infection to a fatal form (symptomatic) of VL. It is well characterized by prolonged fever, splenomegaly, hepatomegaly, substantial weight loss, progressive anemia, pancytopenia, and hypergammaglobulinemia (mainly IgG from polydonal B cell activation) and is complicated by secondary opportunistic infections. The parasite invades and multiplies within macrophages (free mononuclear phagocytic cells) and affects the reticulo-endothelial system including spleen, liver, bone marrow, and lymphoid tissue 5-10. The outcome of fully developed VL is death, usually said to be due to concomitant infection resulting from the weakened immunological state of the patient.

VL is typically caused by L. donovani complex, which includes three species: L. donovani donovani, L. d. infantum, and L. d. chagasi. L. donovani is the causative 12 in the Indian subcontinent and East Africa. L. infantum causes VL in the Mediterranean basin and L. chagasi is responsible for the disease in Central and South America11-14. VL is emerging as an important opportunistic infection among people with HIV-1 infection. In fact, the parasite may be a cofactor in the pathogenesis of HIV infection in human. There are more than 21 morphologically indistinguishable species of Leishmania that infect humans. Conventionally, they are classified and named mainly according to their geographical distribution and clinical characteristics of the disease where they affected15-18. The Post Isla-azar Dermal Leishmaniasis (PKDL) is a type of non ulcerative cutaneous lesion. After recovery from L.
donovani infection, VL patients may develop a chronic form of cutaneous leishmaniasis i.e., PKDL which is developed in about 10% of kala-azar patients generally one or two years after the completion of SAG treatment and requires a long and expensive treatment. PKDL lesions develop 1-13 months post antimony treatment in Sudan, and 1-3 years post antimony treatment in India.19-20.

MATERIALS AND METHODS

Amplification of Leishmania donovani With Gene Specific Primers

PCR is both a thermodynamic and an enzymatic process. Successful gradient PCR requires amplification and detection under optimal conditions and each reaction component can affect the result. The annealing step is critical for high PCR specificity. When primers anneal to the template with high specificity, this leads to high yields of specific PCR products and increases the sensitivity of the amplification reaction. However, due to the high primer concentration in the reaction, primers will also hybridize to non-complementary sequences with mismatches. If the primers anneal to the template sequence with low specificity, amplification of nonspecific PCR products and primer–dimers may occur. Competition in the amplification reaction between these artifacts and the desired PCR product may reduce the yield of the specific product, thereby reducing the sensitivity and linear range of the real-time reaction.

A pair of primers was designed based on Leishmania donovani cysteine protease (Gen-Bank accession number (AF309627). Sequence: forward primer 5’-ATGCCGAGCTGCGGCGCTCTGCGCT-3’ and reverse primer 5’-CTACGTGATCAGCGGTTACGATGTCG-3’. The reaction was performed in 30 μL of the solution containing 24 μg gDNA, 12μl each of forward and reverse primers, 4 mM MgCl2, 6μl dNTPs, 12μl PCR buffer, 4μlTaq DNA polymerase, and H2O up to 50 μL. PCR amplification was carried out within 94 °C for 3 min, followed by 35 cycles of denaturation at 94 °C for 1.0 min, annealing at 48 °C for 1.0 min and extension at 72 °C for 1.30 min, and after final extension at 72 °C for 10 min in a thermal cycler (ABI).

1). For each 120μl PCR, the following reaction components were used. These volumes were multiplied by the number of PCR reactions x 5 to make up a single master mix add 24μl to each tube, which can then be dispensed in 120μl amounts into numbered, thin-walled PCR tubes:

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Volume</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>H2O</td>
<td>50μl</td>
<td></td>
</tr>
<tr>
<td>10x reaction buffer</td>
<td>12μl</td>
<td>[10μM primer stocks = 1μM final conc.]</td>
</tr>
<tr>
<td>PCR primer 1</td>
<td>12μl</td>
<td></td>
</tr>
<tr>
<td>PCR primer 2</td>
<td>12μl</td>
<td>[10mM stock]</td>
</tr>
<tr>
<td>dNTP mix</td>
<td>6μl</td>
<td></td>
</tr>
<tr>
<td>Taq polymerase</td>
<td>4μl</td>
<td>[5 units/μl stock]</td>
</tr>
<tr>
<td>Template Total</td>
<td>24μl</td>
<td></td>
</tr>
<tr>
<td></td>
<td>120μl</td>
<td></td>
</tr>
</tbody>
</table>

2) 6μl of the appropriate template DNA was added to each tube. Negative PCR control was set up which consists of reaction components and no added template DNA. If you are optimizing the PCR, using new template DNA or using a new set of primers, it is advisable to also set up positive PCR control by including a reaction that contains genomics DNA which has been amplified reliably in previous PCRs.

3) Place the tubes into the thermal cycler and close the lid, ensuring that the heated plate inside the lid is in contact with the tops of the tubes, by rotating the screw on the top of the lid until tight.

4) Cycling conditions:

The annealing temperature should be determined for each different set of primers, but primers are generally designed to have optimal annealing temperatures of 48, 50, 52, 54, 56,58,60°C (Tm values >60°C).

<table>
<thead>
<tr>
<th>Stage</th>
<th>Temperature /°C</th>
<th>Time /min</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>94</td>
<td>3min</td>
</tr>
<tr>
<td>2</td>
<td>94</td>
<td>1min</td>
</tr>
<tr>
<td>3</td>
<td>64</td>
<td>1min</td>
</tr>
<tr>
<td>3</td>
<td>72</td>
<td>1.30min</td>
</tr>
<tr>
<td>3</td>
<td>repeat steps 2, 3 and 4 a further 34 times</td>
<td>72</td>
</tr>
<tr>
<td>3</td>
<td>4.0</td>
<td>Forever</td>
</tr>
</tbody>
</table>

5) After cycling, check that amplification was successful by running 30μl of each reaction on an agarose gel, with size standards.

Agarose Gel Electrophoresis

Gel electrophoresis is a widely used technique for the analysis of nucleic acids and proteins. Most every molecular biology research laboratory routinely uses agarose gel electrophoresis for the preparation and analysis of DNA. We will be using agarose gel electrophoresis to determine the presence and size of PCR products. Electrophoresis is a method of separating substances based on the rate of movement while under the influence of an electric field. Agarose is a polysaccharide purified from sea weed. The purpose is to determine the presence or absence of PCR products and quantify the size (length of the DNA molecule) of the product.

Materials needed for preparing 1.2% agarose gel: Agarose-0.6gm, TAE Buffer-50ml, Ethidium bromide-0.2μl, DNA ladder standard, Electrophoresis chamber, Power supply, Gel casting tray and combs, DNA stain, Staining tray, Gloves, Pipette and tips. Add Tris base to ~900 ml H2O. Add acetic acid and EDTA to solution and mix. Pour mixture into 1L graduated cylinder and add H2O to a total volume of 1L.

Preparing the Agarose Gel 1.2%

Measure 0.6g Agarose powder and add it to a 500 ml flask. Add 50ml 1xTBE Buffer to the flask. Melt the agarose in a microwave until the solution becomes clear. (if using a microwave, heat the solution for several short intervals - do not let the solution boil for long periods as it may boil out of the flask). Let the solution cool to about 50-55°C then add Ethidium bromide-0.2μl, swirling the flask occasionally to cool evenly. Seal the ends of the casting tray with two layers of tape. Place the combs in the gel casting tray. Pour the melted agarose solution into the casting tray and let cool until it is solid (it should appear milky white). Carefully pull out the combs and remove the tape. Place the gel in the electrophoresis chamber. Add enough TBE Buffer so that there is about 2-3 mm of buffer over the gel.

Loading the Gel

2μl of 6X conc. sample loading dye to each 8μl of PCR product along with this a controls and ladder was loaded. Carefully 10 μl of each sample/ladder mixture loaded into separate wells in the gel. The order of each sample loaded on the gel was recorded.
Running the Gel

Place the lid on the gel box, connecting the electrodes. The electrode wires to the power supply, making sure the positive (red) and negative (black) are correctly connected. (Remember — Black (negative) to Red (positive)). Turn on the power supply to about 100 volts. Maximum allowed voltage will vary depending on the size of the electrophoresis chamber. Check to make sure the current is running through the buffer by looking for bubbles forming on each electrode. Check to make sure that the current is running in the correct direction by observing the movement of the blue loading dye — this will take a couple of minutes (it will run in the same direction as the DNA). Let the power run until the blue dye approaches the end of the gel. Turn off the power. Disconnect the wires from the power supply. Remove the lid of the electrophoresis chamber. Using gloves, carefully remove the tray and gel.

Purification of PCR product and sequencing

The PCR product was purified using PCR product purification kit (Qiagen) to remove unused dNTPs, enzyme and salt. The ABI Prism BigDye Terminator v1.1 Cycle Sequencing ReadyReaction Kit (Applied Biosystems, Foster City, CA) was used for the sequencing of the PCR product. The sequencing reaction mixture contained 4μL of Big Dye premixture, 0.5 x buffers, 3.2 pmol of sequencing primer, and approximately 150ng of PCR product template in a total volume of 20μL. Sequencing PCR was carried out with the same forward primer. PCR amplification was carried out at conditions 96°C for 60 sec, followed by 25 cycles of denaturation at 96°C for 10sec, annealing at 50°C for 05sec and extension at 60°C for 4 min in a thermal cycler (ABI).

The product was processed, dried and resuspended in 19μl formamide and then loaded in ABI 3130xL genetic analyzer for sequencing following the manufacturer’s recommendations. The sequencing results were analyzed with Sequencer software under the condition of signal/noise > 98%.

RESULTS AND DISCUSSION

Sequence retrieval, alignment and template selection

The amino acid sequence of cysteine protease with NCBI id AAL09444 was retrieved from NCBI, the protein database of the National Center for Biotechnology Information (www.ncbi.nlm.nih.gov/). A BLASTP search was performed against the Protein Data Bank (PDB) to identify a suitable template for homology modeling. Generally, the quality of a template increases with its overall sequence identity and decreases with the increasing number and length of gaps in the alignment. The PDB structure of 1PPP chain A was selected as the best template based on the amino acid identities, the lengths of the gaps and the E-value.

Template Selection

Template selection for homology modeling, best one target templates are selected based on the Sequence identity to the target protein, resolution and completeness of the structure (figure 1 A B C). The template (PDB ID:1AimA) with 58.8% structural identity as shown in the following table which is the crystallographic structure of a cysteine protease (cathepsin L) in Trypanosoma cruzi showed highest similarity to most of the target proteases. The structure had sequence identity 58.8% similarity to target sequence coverage by template structure proteases relatively high resolution of 2.00Å. The template had fairly good sequence coverage of the target proteases; however it was incomplete in the N-terminal domain associated with an inactive peptidase or zymogen. Removal of the N-terminal inhibitor domain either by interaction with a second peptidase or by autocatalytic cleavage activates the zymogen.

<table>
<thead>
<tr>
<th>Target</th>
<th>Template</th>
<th>Model residues range</th>
<th>Sequence id</th>
<th>E-value</th>
<th>QMEAN4-Z score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accession No</td>
<td>Sequence lengths</td>
<td>PDB ID</td>
<td>Lengths</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AAL09444.1</td>
<td>PDB ID:1AimA</td>
<td>126-341</td>
<td>15824693</td>
<td>9.99e-94</td>
<td>-2.08</td>
</tr>
</tbody>
</table>

The template structural identity as shown in the above target template value was predicted from Swiss Automated Model programs superfamily sequence search program.

Figure 1: Predicted homology models of cysteine protease from Leishmania donovani varieties; A is a The structure of cysteine protease enzyme, B Receptor-ligand molecule, C. Groove binder, The protein model that was carried out using homology modeling SWISS model software
Template-Target Alignment

The 3D structures of *Leishmania donovani* cysteine proteases were built and refined using Swiss model and structure were further validate check its structure quality. Based on this assessment of the structure loop refinements were performed to improve the model quality. The summary indicates the model QMEAN4 Z-score (-2.08) energy scores calculated by Swiss model and some of the scores of model validation that are discussed in greater detail ahead.

The identity best template (PDB-1aimA) was good quality having (216) residues found to have residues in most favored region and allowed regions 98.9%. The Swiss model indicated the templates as having acceptable stereochemistry; the secondary structures classifications two type SCOP alpha helix (magenta) and beta strand (yellow) and turn residues (blue) in remaining residues random (white) part of this PDB chain in the (domain red) SCOP tail (violet): domain residues (1-382) are colored by quality to enhance visualization of errors [31]. Blue represents highly scored residues and red poorly scored residues that are likely to be erroneous. This infers that the structures are of enough good quality to be used for modeling. The sub site region of these structures occurs between the left and right lobe and is mostly blue indicating high scoring residues in this region. The conserved domain structure is also of good quality with an exception of its N-terminal, where there is a 1-turn α-helix that is colored red. This C-terminal helix region however doesn't directly overlay the active site and may not be employed in the study of protease inhibition. Given the CATH domain id (1aimA00) FunFams papain like domain (Red 3, 1, 1). CP domain (Green 3, 90, 70, 10) higher (>95%) lower level domain (<35%) for the templates respectively. Modeling allowed us to realize locally stressed regions of template (that is local regions of the template that had high energy values). Swiss automated model was used to assess the stereochemistry of the templates and the results are shown in (Figure 2: A, B).

Figure 2: stereochemistry of the templates

Virtual Screening

Active site identification in the target protein is the starting point for virtual screening. Using was metaPocket 2.0 server. Identification ligand binding site in 3D structure of the cysteine protease. Three pockets were located in the target, and based on further analysis of amino acid residues involved in the active site. Become it having 13 residues in which Asn26, Gln237, Asp85, His86, His181 and His209 have been reported to be conserved in the active site of cysteine protease. Two components searching and scoring are involved in most of the docking algorithms. The vina scoring function amalgamates knowledge based potentials and empirical scoring functions, which extracts empirical information from both the conformational preferences of the receptor-ligand complexes and the experimental affinity measurements. The three-dimensional of all the anti-Leishmanial drugs and medicinal cysteine protease were screened from PubChem and Drug Bank databases were screened with inhibitor designs drugs and compounds and were selected for internal and comparative study. Using the ligdock program, schematic diagrams of protein-ligand interactions for top four receptor-ligand docked complexes were generated in 3D space. It represents the hydrogen and hydrophobic interactions between ligand and active site residues of the cysteine protease. The molecule was interacting with Ser309 Lys 307 and residue through hydrogen bond at a distance of 3.01 Å and 2.97 Å respectively while Ser154, Leu291, Val141, Pro140, Glu175, Val156, Lys142, His288, Phe153 & Gly151, 289, Thr139, Glu160 residues were involved in hydrophobic interactions.

Molecular docking studies

Docking studies were carried out by using the program discovery studio 2.5. Total docking score 8.338 ligand interactions. This program starts with a ligand molecule in an arbitrary conformation, orientation and position and finds favorable dockings in a protein binding site using both simulating annealing and genetic algorithms. The program discovery studio software Tools (DS v2.5), Molecular Viewer, was used to prepare the protein and the ligands. For the macromolecule the 3D crystal structure of CP (PDB ID: 1aimA), resolved at 2.0 Å, was chosen as the docking pattern from the Protein Data Bank. Earlier than docking studies, the macromolecular structure was modified to get more logical and precise outcomes. The A, and D chains were kept whereas the H, X, and Y chains were unmerged and deleted because the hexamer was symmetrical. There are two active sites in this macromolecule. The existed ligands in the crystal were all omitted so that other molecules could be docked and also crystallographic water molecules in the structure were eliminated. Polar hydrogen’s were added, and then
Kollman United Atom charges and atomic salvation parameters were assigned. The grid maps of docking studies were computed using the discovery studio included in the discovery studio 2.5.distribution.

Estimated free energy charge of binding = sum of final intermolecular energy and torsion free energy of ligand.

Figure 3: Amino acids exhibited interaction with ligand(4-Hydroxynonenal) are Glu175, Val156, Lys142, His288, Gly157, Thr139, Phe153, Gly289, Ser309, Cys150, Asn308, Val290, Ser154, Lys307, Leu291, Val141 and Pro140 of model protein of cystein protenase of Leishmania donovani in the docking score 8.338 GOLD. CID: 5283344 (4-Hydroxynonenal, or 4-hydroxy-2-nonenal or 4-HNE or HNE) is an α, β-unsaturated hydroxyalkenal that is produced by lipid peroxidation in cells.

Pharmacophore and 3D QSAR

The constructed model was tested for its predictive power with a test set of 29 molecules. Based upon a PLS factor of five, statistical significance of the model was achieved for useful structural insights. Pharmacophore is an important and unifying concept in rational drug design that embodies the notion that molecules are active at a particular enzyme or receptor because they possess a number of chemical features that favorably interact with the target and which possess geometry complementary to it. A pharmacophore hypothesis collects common features distributed in three-dimensional space representing groups in a molecule that participate in important interactions between drug and active site. Rapid development of combinatorial chemistry and high throughput screening methods in recent years has significantly increased a bulk of experimental structure-activity relationship (SAR) datasets. These developments have emphasized a need for reliable analytical methods for biological SAR data examination such as quantitative SAR (QSAR). QSAR, a CAMD (Computer Aided Molecular Design) technique has been traditionally perceived as a means of establishing correlations between trends in chemical structure modifications and respective changes of biological activity.

The computational analysis using both software and servers of the modeled protein structure enable us the biological properties. The cysteine protease of L. donovani, with the hypothetical protein modeled that was prepared and retrieved from Swiss model tool.

Modeled Structure prediction using homology modeling approaches:

The proper template identification after pBLAST, followed by modeled protein quality assessment enables to understand the suitable functions, and localization of the subcellular protein and their interaction with ligands. The most common and approximate prediction method is homology modeling that provides a proper idea about the protein. The homology model of cysteine protease is shown in figure 4. The figure labeled as sequence alpha (α), beta (β) and flexible loops (FL). The cysteine protease consists of seven alpha helixes, ten beta sheet, three η (small helixes) and few flexible loops. The secondary sequence was clearly mentioned the actual amino acids residues involved in making alpha helixes, beta sheet, flexible loops and small helixes (with three amino acids residues). All the models which were designed from the different servers were compared and validated by DOPE scores by using DSv3.5. The suitable template model that retrieved from the Swiss model has been taken for our study which has the lowest DOPE value of -43031.174938 as the best-modeled structure. The model structure was proved by Verify 3D that showed 83.06 % value. The protein verify by using DSv3.5, shows the 111.74 verify score which lies in between verify expected high score (139.723) and verify expected low score (62.8754). The model validation PROCHECK tool was used to determine Ramachandran plot that assures the quality model. The result of Ramachandran plot showed the amount of amino acid residues i.e. 85% favored regions, 13.1% additional allowed regions, 1.5% generously allowed regions and 0.4% disallowed regions favored represented as a reliable and good quality model. The reliability of the modeled protein was also checked by using ERRAT that showed 83.162 overall model quality.
CONCLUSION
As drug resistance problem persists in case of Leishmaniasis, Cathepsin like cysteine protease is an important target protein for the identification of novel lead compound of Leishmaniasis. Cathepsin like cysteine protease representing a major component of the proteolytic repertoire and play an important role in intracellular protein degradation. Cysteine protease also required for parasite replication and virulence. A three dimensional structure of Cathepsin L like cysteine protease (cpL) and Cathepsin B like cysteine protease (cpB) of Leishmania donovani was constructed through homology modeling using X-ray crystal structure of bovine cathepsin B (PDB ID: 1 QDQ) and recombinant rat cathepsin B (PDB ID: 1 CPJ) for the Cathepsin B like cysteine protease and X-ray crystal structure of cruzain (PDB ID: 1 AIM) for the Cathepsin L like cysteine protease. The homology modeling was done by using Modeller 9v5 and discovery studio software. The final model obtained by the molecular mechanics and dynamics method was assessed by PROCHECK & VERIFY 3-D graph, which showed that the final model is reliable. The model active site of Cathepsin B was used for the virtual high throughput screening of around hundred compounds (vinyl sulfones, hydrazides derivatives anti leishmanial drugs molecules, carbohydrate derivatives) for the potential cysteine protease inhibitors, using AutoDock v4.1, GOLDv2.1. Based on high simulation docking score and fitness value, different sixteen compounds were selected. Out of these 13 have better fitness score and hydrogen bond. These may act as a potential drug for Cathepsin like cysteine protease of Leishmania donovani in future.

REFERENCES