Asian Pub. Corp.

www.asianpubs.org

ARTICLE

Rational Lead Optimization Based on the Modeled Structure of Cysteine Protease of *Leishmania donovani*

Madhumita Dandopath Patra✉

Asian Journal of Organic & Medicinal Chemistry

Volume: 4 Year: 2019
Issue: 4 Month: October–December
pp: 256–266
DOI: https://doi.org/10.14233/ajomc.2019.AJOMC-P239

Received: 18 October 2019
Accepted: 15 December 2019
Published: 31 December 2019

ABSTRACT

In present study, the molecular modeling techniques were applied to generate a refined model of a cysteine protease of *Leishmania donovani* using the crystal structure of a homologous protease and used for lead optimization. The structures of a series of complexes of the protease with the designed inhibitors were predicted using a novel docking technique comprising of repeated cycles of molecular dynamics and energy minimization. Calculation of the free energies of binding of the model with the designed inhibitors suggested that three compounds can form stable complexes with dissociation constants in the nanomolar range (0.038-1.41 nM). Search in the human genome revealed that a number of proteases of the cathepsin family had high homology with the parasite protease with amino acid identity around 45 %. The X-ray structures of all these were available in the protein data bank. The structures of the complexes of the selected inhibitors with a few homologous human proteases of known 3-D structures were also predicted using the same technique of optimization. The electrostatic potentials around the binding sites of the proteases were highly negative, which served as a clue for the introduction of positively charged groups in the designed inhibitors for higher affinity. The comparison of interaction energies and hydrogen bonding patterns among these complexes and similar complexes with homologous human proteases allowed us to short-listed three molecules as effective antileishmanial cysteine protease inhibitors.

KEYWORDS

Inhibitor design, Comparative modeling, Docking, Molecular dynamics, Cysteine protease, *Leishmania donovani*.

INTRODUCTION

Cysteine proteases have been identified as promising targets for the development of antiparasitic chemotherapy. An attractive aspect of these enzymes is their widespread importance in both protozoan and helminth parasites of domestic animals and humans [1]. Cysteine proteases belong to an important class of enzymes, which catalyze the degradative processing of peptides and proteins. They are ubiquitous in nature and play vital roles in numerous pathophysiological processes including arthritis, osteoporosis, Alzheimer's disease, cancer cell invasion and apoptosis [2-4]. A possible strategy for combating parasitic infections is to inhibit cysteine proteases that are crucial to parasite metabolism and reproductive functions [5].

Author affiliations:

Department of Chemistry, Acharya Prafulla Chandra College, New Barrackpore, Kolkata-700131, India

✉To whom correspondence to be addressed:

E-mail: madhumita_ucst@yahoo.co.in

Available online at: http://ajomc.asianpubs.org

Omara-Opyene and Gedamu [6] cloned two cysteine proteases, Ldccys1 (Accession No. AAC38832.2) and Ldccys2 (Accession No. AAC38833.2), from the cDNA libraries prepared, respectively from total amastigote and promastigote RNA preparations. The gene of one of these cysteine proteases, Ldccys2, is expressed both in the promastigote and amastigote stages of *Leishmania donovani* (LD). The over-expressed protein product from this cloned gene was biologically active and could be inhibited by cysteine protease inhibitors [6].

In the present study, we modeled the structure of the protease product of the Ldccys2 gene of LD based on the core structure of homologous cysteine protease from *Trypanosoma cruzi* (cruzain) whose structure was solved by X-ray crystallography at 2.35 Å resolution (PDB ID: 1AIM) containing benzoyl-tyrosine-alanine-fluoromethylketone bound to its active site [7]. The modeled structure was used to design a number of inhibitor molecules. Calculation of electrostatic potentials around the active sites of the protease was used as a guide in the design of inhibitors. We docked the designed molecules into the binding site of modeled cysteine protease using a novel docking method of repeated molecular dynamics and energy minimization. We also studied the nature of interaction of these inhibitors with the homologous human proteases.

EXPERIMENTAL

The initial structure of the cysteine protease from the Ldccys2 (Accession No. AAC38833.2) of LD was predicted by knowledge based homology modeling using our in-house software package of ANALYN and MODELYN (Version PC-1.0 Indian Copyright No 9/98) [8]. From the BLAST search in the PDB database taking the Ldccys2 sequence as query we obtained two crystal structures of cruzain (1AIM & 2AIM) bound to different inhibitors as a significant hit (with 59 % identities, 74 % positive score and expect value of 3×10^{-75}). The next BLAST hit (1BY8) had only 36 % identities, 53 % positive score and expect value of 7×10^{-47}. So the first two hits were much better template for comparative modeling. Two structures of cruzain (1AIM & 2AIM) were superposed giving RMSD of about 0.224 Å taking C_α of 204 residues out of 212. So the first BLAST hit, the X-ray crystallographically determined structure of *Trypanosoma cruzi* (cruzain) (PDB ID: 1AIM) was selected as starting scaffold for comparative modeling.

The modeled structure was refined using the Insight II 2005 of Accelrys (San Diego, CA) equipped with DISCOVER as energy minimization and molecular dynamics module. Structural optimization involved energy minimization (100 steps each of steepest descent and conjugate gradient methods) using cff91 force-field followed by dynamics simulations. A typical dynamics run consisted of 100000 steps of one femtosecond after 1000 steps of equilibration with a conformational sampling of 1 in 100 steps at 300 K. At the end of the dynamics simulation, the conformation with lowest potential energy was picked for the next cycle of refinement using the ANALYSIS module of Insight II. The combination of minimization and dynamics were applied to separately identified segments (affected loop regions) of the model where insertions/deletions were done during comparative model generation separately. The structure of each segment was analyzed to check the parameters like deviations from standard bond lengths and bond angles, deviation from backbone dihedral angles from Ramachandran's plots, clashscores, energy, *etc*. These steps were repeated until structural parameters were within the permissible limits of deviations.

Initial structures of the enzyme-inhibitor complexes were obtained by the superposition of the modeled cysteine protease structure with the experimental structure of cruzain inhibited by benzoyl-tyrosine-alanine-fluoromethylketone (1AIM) followed by optimization with repeated energy minimization and dynamics simulations. During energy minimization and molecular dynamics of the enzyme-inhibitor complexes, atoms of the target molecule, which were more than 10 Å away from the inhibitor, were fixed by applying position constraints. Structures of other inhibitors were generated using the BUILDER module of Insight II followed by optimization with repeated energy minimization and molecular dynamics. The structures of the complexes with the designed inhibitors were obtained by initial placement of the molecules by superposing the common segments with the skeleton structure followed by repeated energy minimization and molecular dynamics.

In order to investigate the influence of water on the inhibitor binding, water molecules were added as a sphere of radius 18 Å having its center at an atom roughly at the center of the inhibitor molecule so as to surround it completely using the Assembly/Soak option of Insight II. In the aqueous environment, structure optimization of the inhibitor was done using energy minimization and molecular dynamics simulation in presence and absence of the protein molecule. From the values of the free energies of complex formation of inhibitors in water and water-protein environments, we calculated the absolute interaction energies following the linear interaction energy approximation method of Aqvist *et al*. [9] using the relation:

$$\Delta G_{bind} = \alpha\Delta\left\langle V_{l\text{-}s}^{el}\right\rangle + \beta\Delta\left\langle V_{l\text{-}s}^{vdw}\right\rangle$$

where ΔG_{bind} is the absolute binding free energy, D stands for differences in the electrical ($V_{l\text{-}s}^{el}$) and van der Waals ($V_{l\text{-}s}^{vdw}$) components of the free energies of the ligand solvent (l-s) systems *i.e.* in pure water and protein containing water environments. The weight factors of the electrical and van der Waals contributions were taken, respectively as 0.5 (α) and 0.16 (β) as proposed by Aqvist *et al*. [9] and used by earlier workers [10,11]. Association constant (K_a) was calculated using the thermodynamic relation $\Delta G_{bind} = -RT \ln K_a$ where R is the ideal gas constant and T is the absolute temperature; dissociation constant K_d was calculated by taking the inverse of K_a.

ANALYN was used for the homology analysis of pre-aligned sequences of the target and scaffold cysteine proteases. It was run on IBM-compatible PC. MODELYN was used for automated prediction of the target structure and its structural analysis after refinement; it was run on both on IBM-compatible PC in the windows environment and on FUEL workstation of Silicon Graphics, Inc. in the UNIX environment. ABGEN [12] was used for structural superposition of the protein molecules at selected locations and was run on FUEL workstation of Silicon Graphics, Inc. in the UNIX environment. Insight II was run on FUEL workstation and Altrix 350 server of Silicon Graphics, Inc. in the IRIX environment. CLUSTALW [13] was

run through the Internet for multiple alignment of the cysteine protease sequences. The electrostatic potential surfaces of the cysteine proteases were determined by MOLMOL [14]. PROCHECK [15] was used for checking the quality of the backbone conformation (Ramachandran's plot) and side-chain planarity of the planar groups, namely, phenylalanine, tyrosine, tryptophan, histidine, arginine, glutamine, asparagines, glutamic acid and aspartic acid by measuring RMS distances of planar atoms from the best fitted plane; residues having RMS distances greater than 0.03 Å for rings and 0.02 Å for other groups were marked as outliers. Both MOLMOL and PROCHECK were run on FUEL in the UNIX operating system.

MOLPROBITY [16] was used for all-atom contact analysis in term of clashscores (number of atoms having atom pair overlaps ≥ 0.4 Å out of 1000 atoms) and for calculation of rotamer outliers. MOLPROBITY, being a general-purpose web service offering quality validation for three-dimensional (3D) structures of proteins, nucleic acids and complexes, was used through the internet. Hydrogen bonding patterns of the modeled and X-ray structure was obtained by adding hydrogen followed by optimization of the enzyme-inhibitor complex by energy minimization and molecular dynamics. The binding affinities of the enzyme-inhibitor complexes were obtained from the DOCKING module of Insight II. Protein BLAST [17] was used through the Internet for finding homologous sequences.

RESULTS AND DISCUSSION

The three-dimensional structure of the cysteine protease from LD was predicted by comparative modeling based on the homologous protease from the parasite *Trypanosoma cruzi* (cruzain), whose structure was determined by X-ray crystallography [7]. Sequences of these two parasite proteins had 59 % identical and 74 % similar amino acids. Standard structural parameters like, all atom clashscores, rotamer outliers, deviations from side-chain planarity, bond lengths and bond angles were calculated and compared with those of the X-ray structure of the template protease (Table-1); these values for the predicted model were good. After initial refinement the quality of the backbone conformations was checked by calculating the ϕ and ψ dihedral angles and drawing Ramachandran's plots for the modeled structure using PROCHECK which showed that more than 95 % of the ϕ-ψ plots were in the core region and less than 3 % were in the disallowed region. On further refinement of the main chain conformations, ϕ-ψ plots of all the non-glycine residues of the modeled structure were brought within the core and allowed regions only.

The optimal sequence alignment of the target and template proteins is shown in Fig. 1. The identical amino acid residues are marked by '*' and the residues involved in catalytic triad formation are shown as bold faced letters. Optimal superposition of the predicted structure of the LD protease and the starting scaffold, cruzain, showed an RMSD with respect to Cα atoms of 0.318 Å over a region of 100 amino acids. The catalytic triad (Asp-140, Cys-25 and His-162) of cruzain mapped very close to the catalytic tried of the LD protease (Asp-140, Cys-25 and His-162). Docking of the inhibitor benzoyl-tyrosine-alanine-fluoromethylketone on to the active site of the modeled LD protease gave the structure of the complex which showed comparable interaction energy with that of cruzain; calculated ΔG_{bind} and K_d values being -6.46 and 0.21 μM for cruzain and -5.69 and 0.76 μM for the LD protease complex (Table-2), respectively. In case of cruzain, Glu-208 (residue numbers are according to Fig. 1) played an important role in the inhibitor binding but in the LD protease this glutamic acid was not conserved; however, interaction with other residues compensated

TABLE-1
STANDARD STRUCTURAL PARAMETERS OF CRYSTAL STRUCTURE AND MODELED STRUCTURE

Proteases from the species	All atom clashcore (per 1000 atom)	Rotamer outliers (%)	Planarity outliers (%)	R.M.S. deviation in bonds (Å)	R.M.S. deviation in angle (°)
Trypanosoma cruzi (cruzain) [X-ray structure]	5.58	3.01	2.9	0.016	3.32
Leishmania donovani [modeled structure]	5.38	3.49	2.8	0.016	2.88

```
Trypanosoma cruzi       APAAVDWRARGAVTAVKDQGQCGSCWAFSAIGNVECQWFLAGHPLTNLSEQMLVSCDKTD
Leishmania donovani     GVMSVDWREKGVVTPVKNQGMCGSCWAFATTGNIEGQWALKNHSLVSLSEQVLVSCDNID
                        .  :**** :*.**.**:** *******:: **:* ** * .*.*..****:*****: *

Trypanosoma cruzi       SGCSGGLMNNAFEWIVQENNGAVYTEDSYPYASGEGISPPCTTSGHTVGATITGHVELPQ
Leishmania donovani     DGCNGGLMQQAMQWIINDHNGTVPTEDSYPYTSAGGTRPPCHDNG-TVGAKIAGYMSLPH
                        .**.****::*::**::::**:* *******:*. *  ***  .* ****.*:*::.**:

Trypanosoma cruzi       DEAQIAAWLAVNGPVAVAVDASSWMTYTGGVMTSCVSEALDHGVLLVGYNDSAAVPYWII
Leishmania donovani     DEEEIAAYVGKNGPVAVAVDATTWQLYFGGVVTLCFGLSLNHGVLVVGFNRQAKPPYWIV
                        ** :***::. **********::*  * ***:* *.. :*:****:**:* .*  ****:

Trypanosoma cruzi       KNSWTTQWGEEGYIRIAKGSNQCLVKEEASSAVVG
Leishmania donovani     KNSWGSSWGEKGYIRLAMGSNQCLLKNYAVTATID
                        **** :.***:****:* ******:*: * :*.:.
```

Fig. 1. Multiple alignments of the sequences of modeled protein and the starting scaffold (1AIM). Amino acid residues involved in catalytic tried formation are marked in bold. The conserved set of AA in all the sequences are identified by a '*' and semi-conserved AAs are marked with ':' or '.'

TABLE-2
EMPIRICAL FREE ENERGIES, THEIR DIFFERENCE IN WATER AND WATER-PROTEIN ENVIRONMENTS AND CORRESPONDING ΔG_{bind} AND K_d VALUES FOR THE COMPLEX FORMATION BETWEEN PROTEINS (CRYSTAL STRUCTURES AND THE MODELED STRUCTURES) AND INHIBITOR (BENZOYL-TYROSIN-ALANINE-FLUROMETHYL KETONE) IN THE AQUEOUS SOLUTION

Compound	Free energy (kcal/mol)			Difference (kcal/mol)		ΔG_{bind} (kcal/mol)	K_d (μM)
	Vdw	Electrical	Total	Vdw	Electrical		
Trypanosoma cruzi (cruzain)(TI)	-69.85	-48.89	-118.74	-8.16	-10.30	-6.46	0.21
Inhibitor (I)*	-61.69	-38.59	-100.28	–	–	–	–
Leishmania donovani (LI)	-74.21	-45.97	-120.18	-12.52	-7.38	-5.69	0.76
FI	-59.58	-36.21	-95.79	+2.11	+2.38	+1.53	–
KI	-68.44	-42.89	-111.33	-6.75	-4.3	-3.23	4.57 mM
VI	-68.82	-39.10	-107.92	-7.13	-0.51	-1.39	97 mM
SI	-68.34	-38.83	-107.17	-6.65	-0.24	-1.18	139 mM

*Value corresponding to the interaction energy in presence of water molecules only as needed for the calculation of K_d value using linear interaction energy approximation method of Aqvist *et al*. [9].

for the binding strength [18]. Hydrogen bonding at Gln-19 is conserved and a hydrogen bond formed with the backbone oxygen of Asp-161 in cruzain while similar hydrogen bonding is formed with Gly-66 in the LD protease. These results justify the use of the modeled structure for the design of the effective inhibitor molecule against the LD protease. Recently, comparative modeled structures of the GPCR (G protein-coupled receptor) family of receptors have been used in rational design and MD simulations were used for docking active molecules in the binding site [19].

From electrostatic potential calculations it was observed that the surfaces of the binding site of both modeled LD protease (Fig. 2a) and cruzain were mostly negatively charged with a slight neutral region. Based on the nature of the electrostatic

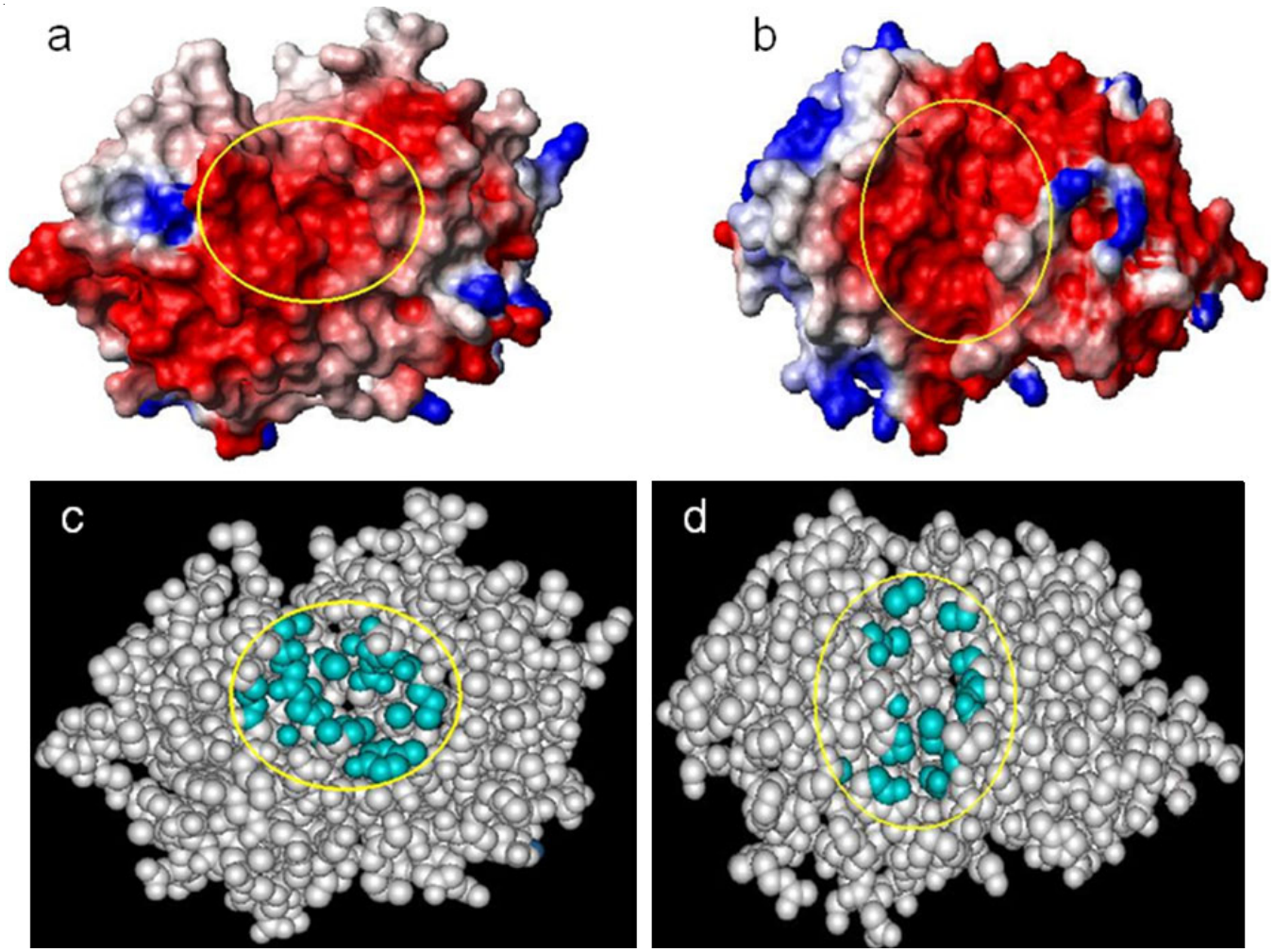


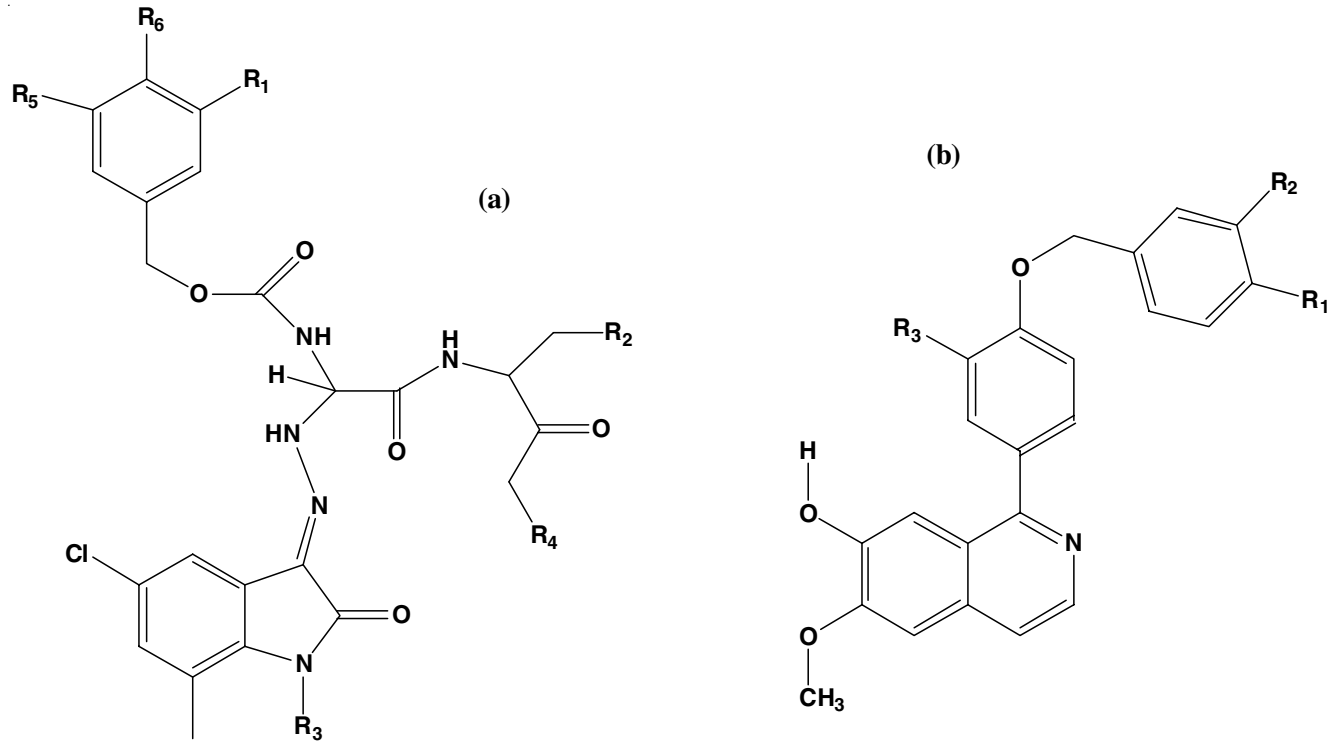
Fig. 2. Electrostatic potential surfaces of the modeled cysteine protease from LD (a) and homologous human protein cathepsin F (b). Blue colour represents positively charged environment, red for negatively charged and white for neutral hydrophobic surroundings. The inhibitor binding site is shown by yellow bands. Space-filling representation of the modeled cysteine protease from LD (c) and homologous human protein cathepsin F (d) showing the atoms in contact with the inhibitor (LI6, Table-3) in cyan inside the yellow band

potential and the existing knowledge of cruzain inhibitors [2,3,20-25], we selected the skeletal structures for generating two sets of inhibitors against the LD protease as shown in **Scheme-I**. The skeletal structure of the first set of inhibitors was designed by combining the basic features of a few known cysteine protease inhibitors, namely, E-64c [24], dipeptidyl α'β'-epoxy ketones [2], dipeptidyl fluoromethyl ketone [7,21], the natural product isatin-based N-substituted thiosemicarbazone derivatives [23], thiosemicarbazone [19,22,24] and vinyl sulphone derivatives [3,25]. The skeletal structure of second set of inhibitors was designed from the existing *Plasmodium falciparum* cysteine protease, falcipain-2, inhibitor, which is a 1-benzyloxyphenyl-6,7-tri-substituted dihydroisoquinoline derivative [26].

A very important aspect of drug design against any pathogenic target is to make the drug more specific towards the parasite compared to any human protein, which may have very similar structure and function of the parasite protein. Therefore, we searched for homologous human proteases using BLAST taking the sequence of the LD protease as a query in the protein data bank (PDB) to look for human proteases of known structures. Thus, it was found that many proteases from cathepsin group have sequence homology with the LD protease were identified. The proteins of known structures namely, cathepsin F [27] (identity 43 %), cathepsin K [28] (identity 45 %), cathepsin V [29] (identity 43 %) and cathepsin S [30] (identity 43 %) were selected to compare the relative interaction energy of the designed inhibitors with respect to the LD protease. This structural similarity of the LD protease is expected because cruzain and cathepsin family of proteases bind and hydrolyze the peptide bond in substrates by placing either Phe or Arg in the S2 specificity pocket [18]. The electrostatic potential around the active site was also negative (Fig. 2b) in the cathepsin family. Among the designed inhibitors of each set, the promising molecules were docked on the active site of these human proteases using the optimization procedure similar to that with the LD protease. We also searched for cathepsin sequences in the human genomic database to check if any cathepsin, whose structure is not yet determined, have higher sequence homology than those of known 3-D structures. Surprisingly, we found only one protease of cathepsin F group with 30 % amino acid identity. This may be due to the fact that many members of the cathepsin family of human origin have been studied by X-ray crystallography.

We modified the primary skeletal structure of the first set of inhibitors by introducing different groups for R_1-R_6 (**Scheme-Ia** and Table-3) to obtain the stable and favourable enzyme-inhibitor complexes with the LD protease. Interaction energies of the enzyme-inhibitor complexes showing significant affinity are presented in Table-3 along with their hydrogen bonding patterns. LI1 is the core structure and R_1 is a hydrogen atom. In LI2, we have introduced NH_3^+ as R_1 into the benzene ring to increase the electrostatic interaction with the protease and also to impose positive charge on the inhibitor required to fit into the negatively charged binding pocket of the cysteine protease. This resulted in a dramatic increase in the interaction energy with empirical free energy of the complex formation at -285.95 kcal/mol as compared to the core structure (-53.62 kcal/mol). This change also introduced an additional hydrogen bond (Table-3). The major gain in the interaction energy comes from electrostatic interaction with about 40 fold increase. This is because LI1 is neutral and LI2 is positively charged which experiences strong electrostatic interaction towards a highly negative binding pocket (Fig. 2a). It may be noted that in case of LI2 there is about 2 folds decrease in the van der Waals component compared to LI1.

In LI3 an additional OH is introduced at R_2, which resulted in a small increase in van der Waals interaction. Although the



Scheme-I

TABLE-3
EMPIRICAL INTERACTION ENERGIES OF SOME CYSTEINE PROTEASE-INHIBITOR COMPLEXES, DERIVED FROM THE STRUCTURE SHOWN IN **Scheme-Ia**, ALONG WITH THEIR HYDROGEN BONDING PARTNERS

Compound	R1	R2	R3	R4	R5	R6	Interaction energy (kcal/mol)	Residues involved in H-bonding
							VdW Electrical Total	
LI1							-47.11 -6.51 -53.62	Gly-66:HN-1:O Gly-66:O-1:HG111
LI2	NH_3^+						-27.32 -258.64 -285.95	Asn-64:O-1:HG6 Gly-66:HN-1:NG11 Gly-66:O-1:HG111
LI3	NH_3^+	OH					-34.91 -255.28 -290.19	Asn-161:O-1:HOG91 Asn-161:HD22-1:FG831
LI4	NH_3^+	OH	CH_3				-35.93 -246.17 -282.11	Asn-64:O-1:HG6
LI5	NH_3^+	OH	CH_2Ph				-31.61 -253.19 -284.81	Asn-161:HD21-1:OG91
LI6	NH_3^+	OH	CH_2CH_3	CH_3			-41.51 -261.77 -303.29	Asn-64:O-1:H3 Asn-64:O-1:H8
LI7	NH_3^+	OH	CH_2CH_3	CH_3	CH_3		-42.92 -260.31 -303.23	Asn-64:O-1:H3 Asn-64:O-1:H8 Asp-60:OD1-1:H14
LI8	NH_3^+	OH	$(CH_2)_2CH_3$	CH_3			-43.49 -254.96 -298.96	Asn-64:O-1:H3 Asp-60:OD1-1:H15
LI9	NH_3^+	OH	CH_2CH_3	CH_3		CH_3	-43.82 -261.03 -304.85	Asn-64:O-1:H3 Asn-64:O-1:H8
LI10	NH_3^+	OH	CH_2CH_3	CH_3		C_2H_5	-43.11 -261.65 -304.75	Asn-64:O-1:HG6 Asn-64:O-1:HG14 Gly-66:HN-1:OD3
LI11	NH_3^+	OH	$COCH_2CH_3$	CH_3			-47.87 -262.27 -310.15	Asn-64:O-1:H3 Asn-64:O-1:H8 Asp-60:OD1-1:H15 Asn-161:HD22-:OG10
LI12	NH_3^+	OH	$CO(CH_2)_2CH_3$	CH_3			-48.35 -262.15 -310.51	Asn-64:O-1:H3 Asn-64:O-1:H8 Asp-60:OD1-1:H15 Asn-161:HD22-:OG10
FI6	NH_3^+	OH	CH_2CH_3	CH_3			-35.19 -197.09 -232.29	Met-64:O-1:H3 Met-64:O-1:H21 Asn-70:OD1-1:H13
KI6	NH_3^+	OH	CH_2CH_3	CH_3			-31.69 -183.19 -214.89	Asp-61:OD1-1:H15 Asp-61:OD2-1:H14 Gly-64:O-1:H3
SI6	NH_3^+	OH	CH_2CH_3	CH_3			-45.81 -83.82 -129.64	Gly-62:O-1:H14 Asn-67:O-1:H3 Phe-70:N-1:H15 Lys-64:HZ1-1:O20
VI6	NH_3^+	OH	CH_2CH_3	CH_3			-47.65 -44.95 -92.59	Gln-63:OE1-1:H13 Gly-61:O-1:H14 Asn-66:O-1:H3
FI11	NH_3^+	OH	$COCH_2CH_3$	CH_3			-36.26 -195.15 -231.41	Met-64:O-1:H3 Leu-67:N-1:H13
KI11	NH_3^+	OH	$COCH_2CH_3$	CH_3			-39.11 -194.97 -234.08	Glu-59:O-1:H13 Asp-61:N-1:H15 Cys-63:O-1:H8 Gly-64:O-1:H1,H3
SI11	NH_3^+	OH	$COCH_2CH_3$	CH_3			-46.05 -59.52 -105.57	Gly-62:O-1:H14 Gly-23:O-1:H8
VI11	NH_3^+	OH	$COCH_2CH_3$	CH_3			-50.69 -80.62 -131.32	Gln-63:OE1-1:H14 Gly-61:O-1:H13 Asn-66:O-1:H3

introduction of a polar group is not expected to increase in van der Waals interaction but the alteration in dipole-dipole interactions might have changed the orientation a little to increase van der Waals contact. In LI4 and LI5, CH_3 and CH_2Ph are introduced separately at R_3 with the expectation of increasing hydrophobic interactions but it is not very effective as indicated in the interaction energies (Table-3), which may be due to steric constraints. In LI6, CH_2CH_3 and CH_3 groups are simultaneously introduced at R_3 and R_4, respectively which increase the van der Waals interactions and give better interaction energy parameters (Fig. 2c). In LI7, addition of another CH_3 group at R_5 does not give any improved binding. In LI8, one CH_3 group is shifted to the R_3 group [$=(CH_2)_2CH_3$] from R_5 to elongate the alkyl group at R_3 which resulted in the decrease in affinity. On the other hand, in LI9, the CH_3 group at R_5 of LI7 is moved to R_6, which did not give any significant change in interaction energy. Replacement of CH_3 group at R_6 of LI9 as done in LI10 with a bulkier CH_2CH_3 group also gives the same result. So, it is concluded that the substitution at R_6 position is not very effective to obtain a better enzyme-inhibitor complex. The structure, LI11, is obtained by replacing the CH_2CH_3 at R_3 of LI6 with a polar group $COCH_2CH_3$ and it gives higher affinity. Further increase in the size of this group, as done in LI12 [$R_3 = CO(CH_2)_2CH_3$] did not result in any significant increase in interaction energy.

These results suggest that the compounds LI6 and LI11 are the effective inhibitors for the LD cysteine protease. The structures of these two compounds are very similar; they differ at the R_3 constituent, in LI6 the group is a non-polar ethyl group while in LI11 it is a polar group containing an additional carbonyl group. Only one residue Asn-64 is involved in hydrogen bonding with the inhibitor (LI6). Role of water molecule on the binding of compounds LI6 and LI11 were examined and also the values of absolute free energy of binding (ΔG_{bind}) were calculated using linear interaction energy approximation as described earlier [9] and presented in Table-4. It may be noted that the calculated ΔG_{bind} values for the complex of LI6 and LI11 are negative. Therefore, these compounds will form stable complexes with cysteine protease as these complexes have lower free energy in the protein-water environment compared to water. Values of ΔG_{bind} for LI6, LI11 correspond, respectively to dissociation constants (K_d) of 0.038 and 1.41 nM (Table-4).

We have docked the structures of LI6 and LI11 into the binding site of the homologous human cathepsins and calculated the interaction energies and hydrogen bonding patterns as listed in Table-3. The compound LI6 fits into the binding sites of the human cathepsins, FI6 (Fig. 2d), KI6, SI6 and VI6 with the total interaction energies of -232.29, -214.89, -129.64 and -92.59 kcal/mol, respectively, which are much less compared to the interaction energy of -303.29 kcal/mol for LD protease-LI6 complex. Similarly, homologous human cathepsins form complexes, FI11, KI11, SI11 and VI11, with the compound LI11 giving total interaction energy values of -231.41, -234.08, -105.57 and -131.32 kcal/mol, respectively, which are also less than those of the corresponding complexes with the LD protease (-310.15 kcal/mol). Binding of compounds, LI6 and LI11, with human cathepsin F and cathepsin K were also studied in water and protein-water environments, the results of which give the calculated ΔG_{bind} values for FI6, KI6, KI11 as positive (Table-4) indicating that the complex formation of these inhibitors with cathepsins in the aqueous medium is thermodynamically unfavourable. But the calculated ΔG_{bind} value for FI11 is negative giving dissociation constants (K_d) of 6.67 M, which is very high indicating practically no binding in millimolar concentrations of these molecules.

We also designed another set of inhibitor molecules based on another skeletal structure as shown in **Scheme-Ib**. We have modified three side chains *i.e.* R_1, R_2, R_3 groups to obtain inhibitors of diverse structures and docked them on to the active sites of LD cysteine protease to generate the structures of the complexes. Interaction energies of the enzyme-inhibitor complexes along with the hydrogen bonding patterns are given in Table-5. In LI14 an acetyl group at R_1 position is present, while increased the affinity only a little compared to the base molecule, LI13. But an additional insertion of a positively charged amino group at R_2 position (LI15) results in a dramatic increase in affinity (about 4 fold), which can be explained as an effect of neutral to positive charge transition in a highly negative binding site as discussed earlier. When the $OCOCH_3$ group of LI15 is replaced with a little bulkier group, $OCOCH_2CH_3$, at

TABLE-4
EMPIRICAL FREE ENERGIES, THEIR DIFFERENCE IN WATER AND WATER-PROTEIN ENVIRONMENTS AND CORRESPONDING ΔG_{bind} AND K_d VALUES FOR THE COMPLEX FORMATION BETWEEN PROTEINS (CRYSTAL STRUCTURES AND THE MODELED STRUCTURES) AND INHIBITORS IN THE AQUEOUS SOLUTION

Compound	Free energy (kcal/mol)			Difference (kcal/mol)		ΔG_{bind} (kcal/mol)	K_d (μM)
	Vdw	Electrical	Total	Vdw	Electrical		
LI6	-91.41	-251.09	-342.50	-30.80	-18.90	-14.38	0.038
I6 *	-60.61	-232.19	-292.81	–	–	–	–
LI11	-98.31	-244.39	-342.70	-33.91	-13.60	-12.23	1.41
I11 *	-64.40	-230.79	-295.19	–	–	–	–
LI21	-102.36	-358.55	-460.91	-18.43	-20.94	-13.42	0.195
I21 *	-83.93	-337.61	-421.54	–	–	–	–
FI6	-67.15	-205.49	-272.64	-6.54	+26.70	+12.30	–
FI11	-73.33	-230.24	-303.57	-8.93	-0.55	-1.16	>1000
FI21	-68.30	-313.46	-381.76	+15.63	+24.15	+14.58	–
KI6	-68.74	-203.86	-272.60	-8.13	+28.33	+12.87	–
KI11	-61.88	-204.59	-266.47	+2.52	+26.20	+13.50	–
KI21	-74.52	-306.66	-381.18	+9.41	+30.95	+16.99	–

*Values corresponding to the interaction energies in presence of water molecules only as needed for the calculation of K_d value using linear interaction energy approximation method of Aqvist *et al*. [9]

TABLE-5
EMPIRICAL INTERACTION ENERGIES OF SOME CYSTEINE PROTEASE-INHIBITOR COMPLEXES, DERIVED FROM THE STRUCTURE SHOWN IN **Scheme-Ib**, ALONG WITH THEIR HYDROGEN BONDING PARTNERS

Compound	R1	R2	R3	Interaction energy (kcal/mol) VdW Electrical Total	Residues involved in H-bonding
LI13				-29.69 -9.82 -39.52	Asn-161:OD1-1:HH
LI14	$OCOCH_3$			-36.58 -11.14 -47.73	Asn-161:HD22- 1:O111
LI15	$OCOCH_3$	NH_3^+		-38.73 -161.63 -200.34	Gly-65:N- 1:H92
LI16	$OCOCH_2CH_3$	NH_3^+		-32.56 -164.67 -197.23	Asn-64:O-1:H16
LI17	$OCOCH_2CH_3$	NH_3^+	NH_3^+	-38.92 -318.14 -357.07	Gly-20:O-1:H10 Met-21:O-1:H12 Cys-22:O-1:H11 Asn-161:O-1:H7
LI18	$OCOCH(CH_3)NH_3^+$	NH_3^+	NH_3^+	-37.99 -483.16 -521.15	Gly-20:O-1:H11 Met-21:O-1:H10 Gln-19:OE1-1:H12 Asn-161:O-1:H7 Asn-161:O-1:H25
LI19	$OCOCH(CH_2CH_3)NH_3^+$	NH_3^+	NH_3^+	-38.64 -484.97 -523.61	Gly-20:O-1:H11 Met-21:O-1:H10 Gln-19:OE1-1:H12 Asn-161:O-1:H7 Asn-64:O-1:H18
LI20	$OCOCH(CH_3)NHCOCH_2Ph$	NH_3^+	NH_3^+	-56.19 -338.15 -394.34	Gln-19:OE1-1:H12 Gly-20:O-1:H11 Asn-161:O-1:H7
LI21	$OCOCH(CH_3)NHCOCH_2Ph(3'CH_3)$	NH_3^+	NH_3^+	-60.76 -329.05 -389.52	Gly-20:O-1:H11 Met-21:O-1:H10 Gln-19:OE1-1:H12 Asn-161:O-1:H7
LI22	$OCOCH(CH_3)NHCOCH_2Ph[3',5'(CH_3)_2]$	NH_3^+	NH_3^+	-61.17 -329.66 -390.84	Gly-20:O-1:H11 Gln-19:OE1-1:H12 Asn-161:O-1:H7
FI19	$OCOCH(CH_2CH_3)NH_3^+$	NH_3^+	NH_3^+	-34.27 -407.98 -442.26	Gly-20:O-1:H11 Cys-22:O-1:H10 Gln-19:OE1-1:H12 Asp-160:OD1-1:H7 Asp-160:O-1:H27 Met-64:O-1:H19 Met-64:SD-1:H18
FI21	$OCOCH(CH_3)NHCOCH_2Ph(3'CH_3)$	NH_3^+	NH_3^+	-53.35 -278.81 -332.16	Gly-20:O-1:H12 Cys-22:O-1:H11 Gln-19:OE1-1:H10 Asp-160:OD1-1:H7 Asp-160:O-1:H24 Gln-140:HE21-1:OZ2
KI19	$OCOCH(CH_2CH_3)NH_3^+$	NH_3^+	NH_3^+	-2.67 -490.36 -493.03	Asp-61:OD1-1:H18 Asp-61:OD2-1:H28
KI21	$OCOCH(CH_3)NHCOCH_2Ph(3'CH_3)$	NH_3^+	NH_3^+	-27.16 -333.21 -360.37	Asp-61:OD1-1:H10 Asp-61:OD2-1:H18 Asp-61:O-1:H10 Tyr-67:HH-1:O25 Glu-93:OE2-1:H12
SI19	$OCOCH(CH_2CH_3)NH_3^+$	NH_3^+	NH_3^+	-41.83 -158.62 -200.46	Ser-21:O-1:H12 Gly-23:N-1:H11 Asn-67:O-1:H19 Asn-164:ND1-1:H7 Asn-164:ND1-1:OH2

SI21	$OCOCH(CH_3)NHCOCH_2Ph(3'CH_3)$	NH_3^+	NH_3^+	-58.26 -120.27 -178.54	Gly-20:O-1:H12 Cys-22:N-1:H11 Gly-23:N-1:H11 Asn-163:O-1:H24 Gln-19:OE1-1:H10 His-164:ND1-1:H7 His-164:ND1-1:OH2
VI19	$OCOCH(CH_2CH_3)NH_3^+$	NH_3^+	NH_3^+	-20.63 -249.65 -270.27	Asn-66:O-1:H20 Gln-21:OE1-1:H10 Asn-66:OD1-1:H19
VI21	$OCOCH(CH_3)NHCOCH_2Ph(3'CH_3)$	NH_3^+	NH_3^+	-56.03 -118.18 -174.21	Asn-66:O-1:H20

R_1 of LI16 there is a slight decrease in the affinity. Sequential insertions of charged NH_3^+ group at R_3 and R_1 positions as in LI17 (-357.07 kcal/mol) and LI18 (-521.15 kcal/mol), respectively led to big increases in the stability of the complexes while further modifications of the group at R_1 (from methyl to ethyl) there is a slight increase in the stability (LI19).

A neutral group containing peptide type linkage, $OCOCH(CH_3)NHCOCH_2Ph$, which imparts some constraint in the group has been used at R_1 position of LI20, which gives an improved van der Waals interaction than LI19 (-38.64 to -56.19 kcal/mol) although there is a big reduction in the strength of electrostatic attraction (-484.97 to -338.15 kcal/mol) and a loss of two hydrogen bonding. Replacement of this group with a little bulkier group, $OCOCH(CH_3)NHCOCH_2Ph(32CH_3)$, at R_1 of LI21, containing the same peptide unit as in LI20 improves the van der Waals contribution a little and adds one more hydrogen bond (Fig. 3a-b). Further increase in the bulk of the group, as in LI22, did not improve binding strength.

Two compounds from Table-5, LI19 and LI21, are used to check cross reactivity with the same set of human homologues. LI19 has the highest negative interaction energy of -523.61 kcal/mol and has high positive charge while LI21 has relatively higher van der Waals interactions. Total empirical interaction energies of LI19 in the complexes FI19, KI19, SI19 and VI19 are -442.26, -493.03, -200.46 and -270.27 kcal/mol, respectively. It is important to note that the binding affinities of LI19 with the human homologues FI19 and KI19 are very close to that with the modeled LD protease. This result indicates that this compound is not suitable as an inhibitor as it may cause side effects due to interference with the human enzymes. Interaction energies of LI21 with FI21, KI21, SI21 and VI21 are -332.16, -360.37, -178.54 and -174.21 kcal/mol, respectively; values are considerably less than those with the LD protease, hence it may be short-listed as an effective inhibitor. The value of ΔG_{bind} for LI21, as shown in Table-4, corresponds, respectively to dissociation constants (K_d) of 0.195 nM. Calculated ΔG_{bind} values for the complexes of FI21 and KI21 are positive; hence no complex formation is expected in the aqueous medium. We also docked the inhibitor of *Trypanosoma cruzi* (cruzain), benzoyl-tyrosine-alanine floromethyl ketone, into the binding site of the human cathepsins and calculated the ΔG_{bind} and K_d values to compare the relative affinity of the designed inhibitors with respect to cruzain inhibitor (Table-2).

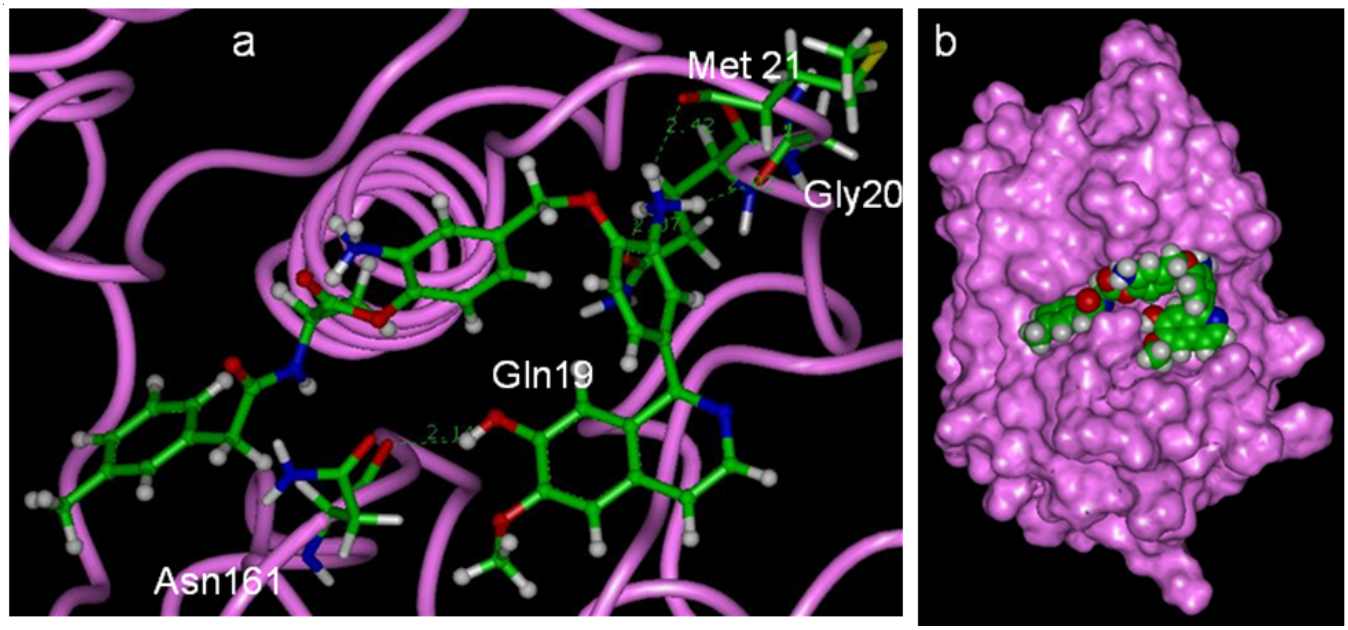


Fig. 3. (a) Mode of interaction of the modeled cysteine protease from LD and the optimized inhibitor (LI21, Table-5). Residues of the protease involved in hydrogen bonding with the inhibitor are indicated by stick representation. And the inhibitor is represented by ball and stick and the backbone of the modeled cysteine protease as pink ribbon; (b) The docked inhibitor (LI21, Table-5) is shown by space-filling representation in the active site in the Connolly surface environment (pink) of the modeled cysteine protease is indicated around the bound inhibitor

Conclusion

We have short-listed three compounds as effective inhibitor, LI6 and LI11 from Table-3 and LI21 from Table-5, which may be synthesized and tested for their antileishmanial activity. The striking feature of the LD protease binding site is its highly negative electrostatic environment of the active site (Fig. 2a), hence, positively charged group(s) on the effective inhibitor is essential for high affinity binding. However, the negatively charged environment of some close homologues of the protease (Fig. 2b) cause a big problem but we could identify some potential inhibitors with reasonable differences in the affinity. Charge-charge interactions are very much dependent on the electrolytic environment as the increase in dielectric constant can reduce the electrostatic contributions and van der Waals interactions may play a dominant role in binding. Hydrogen bonding interactions are not as sensitive to dielectric constant of the medium as the charge-charge and other dipole-dipole interactions. LI6 forms only two hydrogen bonds with the backbone oxygen atom of Asn-64 while LI11 forms two more hydrogen bonds with the side-chains of Asp-60 and Asn-161 while retaining the same bonds with Asn-64. LI21 forms two backbone hydrogen bonds with Gly-20 and Met-21 and two more with the side chains Gln-19 and Asn-161. Calculation of ΔG_{bind} values by the linear interaction energy approximation [9] in water and water-protein environments of the inhibitors suggests that inhibitors LI6, LI11, LI21 can form thermodynamically stable complexes with the cysteine protease from LD (Table-4) while the homologous human proteins such as cathepsin F and cathepsin K gave positive or slightly negative values indicating no complex formation or the formation of very weak complex with the inhibitors. The calculated dissociation constants (K_d) values for the inhibitors LI6, LI11, LI21 are 0.038 nM, 1.41 nM and 0.195 nM, respectively, which are expected to be high affinity inhibitors and their K_d values match well with the experimental values obtained for protease-inhibitor complexes as presented in the protein-ligand database [31]. It is expected that the three molecules, LI6, LI11 and LI21 will make good drug candidates with less interference with the human proteases.

ACKNOWLEDGEMENTS

This work was financially supported by the CSIR MMP Grant No CMM0017.

REFERENCES

1. J.H. McKerrow, Development of Cysteine Protease Inhibitors as Chemotherapy for Parasitic Diseases: Insights on Safety, Target Validation, and Mechanism of Action, *Int. J. Parasitol.*, **29**, 833 (1999); https://doi.org/10.1016/S0020-7519(99)00044-2.
2. W.R. Roush, F.V. González, J.H. McKerrow and E. Hansell, Design and Synthesis of Dipeptidyl α′,β′-Epoxy Ketones, Potent Irreversible Inhibitors of the Cysteine Protease Cruzain, *Bioorg. Med. Chem. Lett.*, **8**, 2809 (1998); https://doi.org/10.1016/S0960-894X(98)00494-6.
3. W.R. Roush, J. Cheng, B. Knapp-Reed, A. Alvarez-Hernandez, J.H. McKerrow, E. Hansell and J.C. Engel, Potent Second Generation Vinyl Sulfonamide Inhibitors of the Trypanosomal Cysteine Protease Cruzain, *Bioorg. Med. Chem. Lett.*, **11**, 2759 (2001); https://doi.org/10.1016/S0960-894X(01)00566-2.
4. P.M. Selzer, S. Pingel, I. Hsieh, B. Ugele, V.J. Chan, J.C. Engel, M. Bogyo, D.G. Russell, J.A. Sakanari and J.H. McKerrow, Cysteine Protease Inhibitors as Chemotherapy: Lessons from a Parasite Target, *Proc. Natl. Acad. Sci. USA*, **96**, 11015 (1999); https://doi.org/10.1073/pnas.96.20.11015.
5. P.V. Desai, A. Patny, Y. Sabnis, B. Tekwani, J. Gut, P. Rosenthal, A. Srivastava and M. Avery, Identification of Novel Parasitic Cysteine Protease Inhibitors Using Virtual Screening. 1. The ChemBridge Database, *J. Med. Chem.*, **47**, 6609 (2004); https://doi.org/10.1021/jm0493717.
6. A.L. Omara-Opyene and L. Gedamu, Molecular Cloning, Characterization and Overexpression of Two Distinct Cysteine Protease cDNAs from *Leishmania donovani chagasi*, *Mol. Biochem. Parasitol.*, **90**, 247 (1997); https://doi.org/10.1016/S0166-6851(97)00158-8.
7. M.E. McGrath, A.E. Eakin, J.C. Engel, J.H. McKerrow, C.S. Craik and R.J. Fletterick, The Crystal Structure of Cruzain: A Therapeutic Target for Chagas' Disease, *J. Mol. Biol.*, **247**, 251 (1995); https://doi.org/10.1006/jmbi.1994.0137.
8. C. Mandal, MODELYN – A Molecular Modelling Program Version PC-1.0, Indian Copyright No. 9/98 (1998).
9. J. Aqvist, C. Medina and J.-E. Samuelsson, A New Method for Predicting Binding Affinity in Computer-Aided Drug Design, *Protein Eng. Design Select.*, **7**, 385 (1994); https://doi.org/10.1093/protein/7.3.385.
10. J. Aqvist and S.L. Mowbray, Sugar Recognition by a Glucose/Galactose Receptor. Evaluation of Binding Energetics from Molecular Dynamics Simulations, *J. Biol. Chem.*, **270**, 9978 (1995); https://doi.org/10.1074/jbc.270.17.9978.
11. J. Hulten, N.M. Bonham, U. Nillroth, T. Hansson, G. Zuccarello, A. Bouzide, J. Åqvist, B. Classon, U.H. Danielson, A. Karlen, I. Kvarnstrom, B. Samuelsson and A. Hallberg, Cyclic HIV-1 Protease Inhibitors Derived from Mannitol: Synthesis, Inhibitory Potencies and Computational Predictions of Binding Affinities, *J. Med. Chem.*, **40**, 885 (1997); https://doi.org/10.1021/jm960728j.
12. C. Mandal, B.D. Kingery, J.M. Anchin, S. Subramaniam and D.S. Linthicum, ABGEN: A Knowledge-Based Automated Approach for Antibody Structure Modeling, *Nat. Biotechnol.*, **14**, 323 (1996); https://doi.org/10.1038/nbt0396-323.
13. J.D. Thompson, D.G. Higgins and T.J. Gibson, CLUSTAL W: Improving the Sensitivity of Progressive Multiple Sequence Alignment through Sequence Weighting, Position-specific Gap Penalties and Weight Matrix Choice, *Nucleic Acids Res.*, **22**, 4673 (1994); https://doi.org/10.1093/nar/22.22.4673.
14. R. Koradi, M. Billeter and K. Wuthrich, MOLMOL: A Program for Display and Analysis of Macromolecular Structures, *J. Mol. Graph.*, **14**, 51 (1996); https://doi.org/10.1016/0263-7855(96)00009-4.
15. R.A. Laskowski, M.W. MacArthur, D.S. Moss and J.M. Thornton, PROCHECK: A Program to Check the Stereochemical Quality of Protein Structures, *J. Appl. Cryst.*, **26**, 283 (1993); https://doi.org/10.1107/S0021889892009944.
16. I.W. Davis, L.W. Murray, J.S. Richardson and D.C. Richardson, MOLPROBITY: Structure Validation and All-atom Contact Analysis for Nucleic Acids and their Complexes, *Nucleic Acids Res.*, **32**, W615 (2004); https://doi.org/10.1093/nar/gkh398.
17. S.F. Altschul, T.L. Madden, A.A. Schäffer, J. Zhang, Z. Zhang, W. Miller and D.J. Lipman, Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs, *Nucleic Acids Res.*, **25**, 3389 (1997); https://doi.org/10.1093/nar/25.17.3389.
18. S.A. Gillmor, C.S. Craik and R.J. Fletterick, Structural Determinants of Specificity in the Cysteine Protease Cruzain, *Protein Sci.*, **6**, 1603 (1997); https://doi.org/10.1002/pro.5560060801.
19. S. Costanzi, B.V. Joshi, S. Maddileti, L. Mamedova, M.J. Gonzalez-Moa, V.E. Marquez, T.K. Harden and K.A. Jacobson, Human $P2Y_6$ Receptor: Molecular Modeling Leads to the Rational Design of a Novel Agonist Based on a Unique Conformational Preference, *J. Med. Chem.*, **48**, 8108 (2005); https://doi.org/10.1021/jm050911p.

20. C.R. Caffrey, M. Schanz, J. Nkemgu-Njinkeng, M. Brush, E. Hansell, F.E. Cohen, T.M. Flaherty, J.H. McKerrow and D. Steverding, Screening of Acyl Hydrazide Proteinase Inhibitors for Antiparasitic Activity against *Trypanosoma brucei*, *Int. J. Antimicrob. Agents*, **19**, 227 (2002); https://doi.org/10.1016/S0924-8579(01)00488-5.
21. I. Chiyanzu, E. Hansell, J. Gut, P.J. Rosenthal, J.H. McKerrow and K. Chibale, Synthesis and Evaluation of Isatins and Thiosemicarbazone Derivatives against Cruzain, Falcipain-2 and Rhodesain, *Bioorg. Med. Chem. Lett.*, **13**, 3527 (2003); https://doi.org/10.1016/S0960-894X(03)00756-X.
22. X. Du, C. Guo, E. Hansell, P.S. Doyle, C.R. Caffrey, T.P. Holler, J.H. McKerrow and F.E. Cohen, Synthesis and Structure-Activity Relationship Study of Potent Trypanocidal Thiosemicarbazone Inhibitors of the Trypanosomal Cysteine Protease Cruzain, *J. Med. Chem.*, **45**, 2695 (2002); https://doi.org/10.1021/jm010459j.
23. N. Fujii, J.P. Mallari, E.J. Hansell, Z. Mackey, P. Doyle, Y.M. Zhou, J. Gut, P.J. Rosenthal, J.H. McKerrow and R.K. Guy, Discovery of Potent Thiosemicarbazone Inhibitors of Rhodesain and Cruzain, *Bioorg. Med. Chem. Lett.*, **15**, 121 (2005); https://doi.org/10.1016/j.bmcl.2004.10.023.
24. D.C. Greenbaum, Z. Mackey, E. Hansell, P. Doyle, J. Gut, C.R. Caffrey, J. Lehrman, P.J. Rosenthal, J.H. McKerrow and K. Chibale, Synthesis and Structure-Activity Relationships of Parasiticidal Thiosemicarbazone Cysteine Protease Inhibitors against *Plasmodium falciparum*, *Trypanosoma brucei* and *Trypanosoma cruzi*, *J. Med. Chem.*, **47**, 3212 (2004); https://doi.org/10.1021/jm030549j.
25. K.A. Scheidt, W.R. Roush, J.H. McKerrow, P.M. Selzer, E. Hansell and P.J. Rosenthal, Structure-based Design, Synthesis and Evaluation of Conformationally Constrained Cysteine Protease Inhibitors, *Bioorg. Med. Chem.*, **6**, 2477 (1998); https://doi.org/10.1016/S0968-0896(98)80022-9.
26. S. Batra, Y.A. Sabnis, P.J. Rosenthal and M.A. Avery, Structure-Based Approach to Falcipain-2 Inhibitors: Synthesis and Biological Evaluation of 1,6,7-Trisubstituted Dihydroisoquinolines and Isoquinolines, *Bioorg. Med. Chem.*, **11**, 2293 (2003); https://doi.org/10.1016/S0968-0896(03)00117-2.
27. J.R. Somoza, J.T. Palmer and J.D. Ho, The Crystal Structure of Human Cathepsin F and Its Implications for the Development of Novel Immunomodulators, *J. Mol. Biol.*, **322**, 559 (2002); https://doi.org/10.1016/S0022-2836(02)00780-5.
28. E. Altmann, S.W. Cowan-Jacob and M. Missbach, Novel Purine Nitrile Derived Inhibitors of the Cysteine Protease Cathepsin K, *J. Med. Chem.*, **47**, 5833 (2004); https://doi.org/10.1021/jm0493111.
29. J.R. Somoza, H. Zhan, K.K. Bowman, L. Yu, K.D. Mortara, J.T. Palmer, J.M. Clark and M.E. Mcgrath, Crystal Structure of Human Cathepsin V, *Biochemistry*, **39**, 12543 (2000); https://doi.org/10.1021/bi000951p.
30. Y.D. Ward, D.S. Thomson, L.L. Frye, C.L. Cywin, T. Morwick, M.J. Emmanuel, R. Zindell, D. McNeil, Y. Bekkali, M. Hrapchak, M. DeTuri, K. Crane, D. White, S. Pav, Y. Wang, M.-H. Hao, C.A. Grygon, M.E. Labadia, D.M. Freeman, W. Davidson, J.L. Hopkins, M.L. Brown and D.M. Spero, Design and Synthesis of Dipeptide Nitriles as Reversible and Potent Cathepsin S Inhibitors, *J. Med. Chem.*, **45**, 5471 (2002); https://doi.org/10.1021/jm020209i.
31. D. Puvanendrampillai and J.B.O. Mitchell, Protein Ligand Database (PLD): Additional Understanding of the Nature and Specificity of Protein–Ligand Complexes, *Bioinformatics*, **19**, 1856 (2003); https://doi.org/10.1093/bioinformatics/btg243.