

# A ClpP protein model as tuberculosis target for screening marine compounds

Abhilasha Tiwari, Smita Gupta, Shipra Srivastava, Rajeev Srivastava, Anil Kumar Rawat\*

Biotechnology and Bioinformatics Division, BIOBRAINZ, 566/29 J, Jai Prakash Nagar, Alambagh, Lucknow 226005, U.P., India; Anil Kumar Rawat\* E-mail: akrbz@rediffmail.com; Phone No: 91- 0522 4042284; \*Corresponding author

Received September 20, 2009; accepted November 13, 2009; published March 31, 2010

## Abstract:

ATP-dependent Clp protease (ClpP) is a core unit of a major bacterial protease complex employing as a new attractive drug target for that isolates, which are resistant to antibiotics. *Mycobacterium tuberculosis*, a gram-positive bacterium, is one of the major causes of hospital acquired infections. ClpP in *Mycobacterium tuberculosis* is usually tightly regulated and strictly requires a member of the family of Clp-ATPase and often further accessory proteins for proteolytic activation. Inhibition of ClpP eliminates these safeguards and start proteolytic degradation. Such uncontrolled proteolysis leads to inhibition of bacterial cell division and eventually cell death. In order to inhibit Clp protease, at first three dimensional structure model of ClpP in *Mycobacterium tuberculosis* was determined by comparative homology modeling program MODELLER based on crystal structure of the proteolytic component of the caseinolytic Clp protease (ClpP) from *E. coli* as a template protein and has 55% sequence identity with ClpP protein. The computed model's energy was minimized and validated using PROCHECK to obtain a stable model structure and is submitted in Protein Model Database (PMDB-ID: PM0075741). Stable model was further used for virtual screening against marine derived bioactive compound database through molecular docking studies using AutoDock 3.05. The docked complexes were validated and enumerated based on the AutoDock Scoring function to pick out the best marine inhibitors based on docked Energy. Thus from the entire 186 Marine compounds which were Docked, we got best 5 of them with optimal docked Energy (Ara-A: -14.31 kcal/mol, Dysinosisin C: -14.90kcal/mol, Nagelamide A: -20.49 kcal/mol, Strobilin: -8.02 kcal/mol, Manoalide: -8.81 kcal/mol). Further the five best-docked complexes were analyzed through Python Molecular Viewer software for their interaction studies. Thus from the Complex scoring and binding ability its deciphered that these Marine compounds could be promising inhibitors for ClpP as Drug target yet pharmacological studies have to confirm it.

**Keywords:** Homology Modeling; Docking; ClpP; *M.tuberculosis*; AutoDock; Modeller; Procheck; Tuberculosis; AutoDockTools; Clp protease.

## Background:

Bacterial pathogens have evolved numerous defense mechanisms against antimicrobial agents, and resistance to old and new produced drugs are on the rise. The alarming increase of antibiotic-resistant bacterial pathogens points to the need for novel therapeutic approaches to combat infection. Tuberculosis (TB) is an infectious disease that is caused by the gram-positive bacteria known as *Mycobacterium tuberculosis*. It is the major causes of death from an infectious disease in the world resulting in an estimated 2 billion people have latent TB, while another 3 million people worldwide die of TB each year [1]. This alarming increase of bacterial pathogens points to the need for novel therapeutic approaches to combat infection. The ATP-dependent Clp protease of *Mycobacterium tuberculosis* CDC1551 consists of two subunits, the ClpP subunit which is capable of degrading peptides in energy-dependent manner and ClpX, which possesses ATPase activity and activates the proteolytic activity of ClpP in vitro. ClpP is a core unit of a major bacterial protease complex employing as a new drug target for that isolates, which are resistant to antibiotics. ClpP is usually tightly regulated and strictly requires a member of the family of Clp-ATPase and often further accessory proteins for proteolytic activation. Binding of inhibitor with ClpP eliminates these safeguards. The inhibitor-activated ClpP core is capable of proteolytic degradation in the absence of the regulatory Clp-ATPases. Such uncontrolled proteolysis leads to inhibition of bacterial cell division and eventually cell death [2]. ClpP is a key catalyst in proteolytic activation and a promising attractive drug target for novel antibiotics.

In this study, we targeted the ClpP and proposed the three-dimensional structure model, determined by comparative homology modeling. Crystal structure of the proteolytic component of the caseinolytic Clp protease (ClpP) from *E. coli* having 55% identity in amino acid sequence with ClpP was used as a template protein for homology modeling. Stereochemical property of the model was checked. Model was further used for virtual screening against Marine Compound Database (MCDB) [3,4]. Since inhibitor of ClpP can be a good candidate of the new antimicrobial drug, we studied interactions between the ClpP with Inhibitors

## Methodology:

### Model building

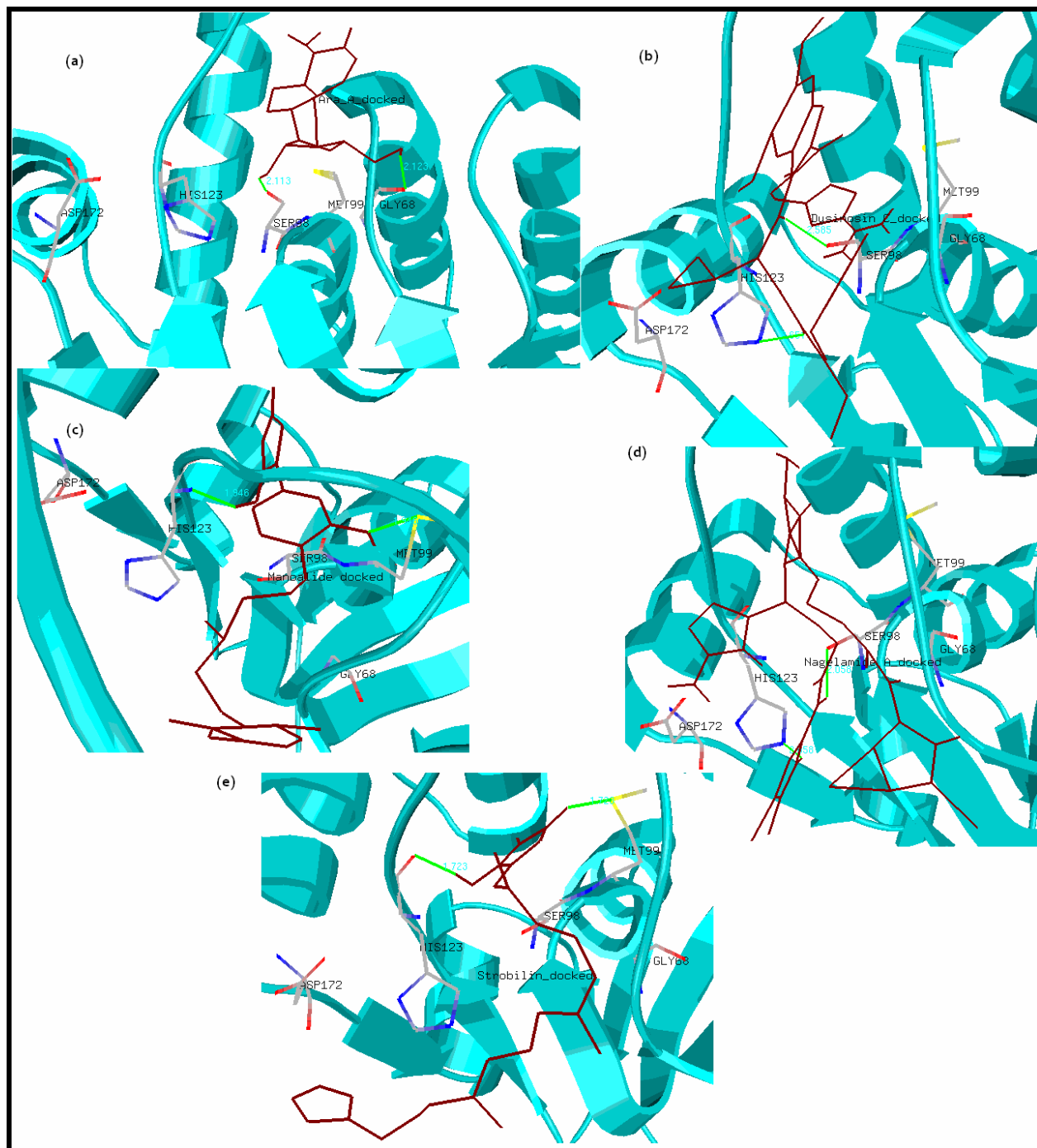
The single amino acid sequence of ATP-dependent Clp protease proteolytic subunit in *Mycobacterium tuberculosis* CDC1551 (Accession No: NP\_337022, Version: NP\_337022.1, GI: 15841985) was retrieved from protein sequence database sited at NCBI [5] and taken as target protein sequence. The modeling of 3D structure of target protein followed a stepwise procedure, starting with a template structure search from PDB (protein databank) [6], related to the target sequence using BLASTP [7]. From a number of hits, a potential template structure (PDB-ID: 1TYF), representing the crystal structure of the proteolytic component of the caseinolytic Clp protease (ClpP) from *E. coli* was taken as template for model building. The template and target sequences were aligned using the align2d script available in comparative protein modeling program MODELLER9v6 [8]. Based on the alignment, five comparative models of the target sequence were built by MODELLER, applying the default model building routine 'model' with fast refinement. In the first step of model building, distance and dihedral angle restraints on the target sequence were derived from its alignment with the template 3D-structure.

### Evaluation of model

The best model can be selected by picking the model with the lowest value of the Modeller objective function and DOPE (Discrete Optimization Protein) score from a collection of models built by MODELLER. Modeller objective function and DOPE score are the statistical parameter for assessment of model using the standard Modeller Energy function. The Stereochemical qualities of the models are checked by PROCHECK [9].

### Active site analysis

After the final model was built, the possible Active site residues of ClpP in *Mycobacterium tuberculosis* was identified on the basis of CSA (Catalytic Site Atlas) [10] entry of PDB code of 1TYF contains a GLY-SER -MET -HIS-ASP Catalytic triad [11]. These residues of active site are conserved in various bacterial ClpP proteins. Thus the active site residues of ClpP in *Mycobacterium tuberculosis* have GLY68 - SER98 - MET99 - HIS123 -ASP172 Catalytic triad.



**Figure 1:** Docking model of (a) Ara-A inhibitor with active site residues of ClpP. (b) dysinosin C inhibitor with active site residues of ClpP. (c) Manoalide inhibitor with active site residues of ClpP. (d) Nagelamide inhibitor with active site residues of ClpP. (e) Strobilin inhibitor with active site residues of ClpP. Hydrogen bond between inhibitor and residue is represented by green line in each case.

**Ligand generation and optimization**

The 2D structures of ligands used in this study were downloaded from Marine Compound Database (MCDB) [3,4]. The downloaded ligands in the mol (MDL MOL) format were first converted to the hin (HyperChem) format using OpenBabel [12]. The 2D structures of ligands were converted to 3D structures using the HyperChem7.5 software [13]. All modeling procedures, including energy minimization and molecular dynamics, were also performed using the HyperChem. Energy calculations were carried out using the AMBER force field.

Optimized molecular structure using conjugate gradient method Polak-Ribiere until the maximum energy derivative was lower than 0.1kcal/A<sup>o</sup>mol in order to obtain a correct geometry. Further the optimize ligands in hin format were converted to the PDB (Protein Databank) format using OpenBabel.

**Virtual screening**

Virtual screening of the entire 186 marine compounds of MCDB [3,4] against ClpP protein structure, determined by comparative homology

modeling, were done using molecular docking program AutoDock 3.05 [14]. The Kollman charges and the solvation term were then added to the protein structure using the AutoDockTool [15]. A grid-box was generated that was large enough to cover the entire protein catalytic site and accommodate ligands to move freely. AutoDock3.05 and a Lamarckian Genetic Algorithm (LGA) were used for protein-fixed ligand-flexible docking calculations. Thirty search attempts (ga\_run parameter) were performed for each ligand. The maximum number of energy evaluations before the termination of LGA run was 250000 and the maximum number of generations of the LGA run before termination was 27000. Other docking parameters were set to the software's default values. After docking, the ligands were ranked according to their docked energy as implemented in the AutoDock 3.05 program.

#### Results and discussion:

Five models of ClpP were generated by MODELLER using crystal structure of the proteolytic component of the caseinolytic Clp protease (ClpP) from *E.coli* [PDB-ID: 1TYF] as a template protein for homology modeling. In order to select the best model, lowest value of Modeller objective function and DOPE score are used and further checked the structural validity by PROCHECK and is submitted in Protein Model Database (PMDb) [16]. 3D structure of all the 186 compound of MCDB [3, 4] were generated and subjected to energy minimization using conjugate gradient algorithm through HyperChem7.5 software [13]. Docking of these optimized compounds against modeled ClpP at the catalytic site residues were performed by AutoDock3.05 [14]. Out of 186 docked complexes, we got 5 best docked marine compounds having lowest docked energy and Root mean square deviation from a reference structure are shown in table 1 (see supplementary material) and rest of the data not shown in this work. Five best docked complexes were analyzed through Python Molecular Viewer [17] for their interaction study shown in Figure 1 (a, b, c, d, e). It is evident from this analysis that the best marine inhibitors are located in the center of the active site and is stabilized by hydrogen bonding interactions.

#### Conclusion:

ClpP is one of the most recent potent Drug target for tuberculosis. In this work, we have constructed a 3D model of ClpP, using the MODELLER software and obtained a refined model after energy minimization. The final refined model was further assessed by

PROCHECK program, and the results show that the model was stable and reliable. The stable model was further used for Virtual Docking of Marine Compounds. Docking results indicate that out of 186 Marine compounds, there were five inhibitory compounds for ClpP as target for tuberculosis. As it's well known, hydrogen bonding plays an important role for the structure and function of biological molecules, especially for inhibition in a complex. Thus our study confirms Ara-A, Dysiosin C, Manoalide, Nagelamide A and Strobilin are potential inhibitors for ClpP as target for tuberculosis forming a hydrogen bonding and with non-bonded interaction to act as a drug candidates yet Pharmacological study will yet confirm it to be promising.

#### References:

- [1] World Health Organization. Global tuberculosis control report – surveillance, planning, finance. WHO report 2008
- [2] H.B. Oesterhelt *et al.*, *Nat Med.* 11(10):1082 (2005) [PMID: 16200071]
- [3] <http://www.progenebio.in/mcdb/index.htm>
- [4] P.A. Babu *et al.*, *Bioinformation*, 3(3): 142 (2008) [PMID: 19238254] <http://www.ncbi.nlm.nih.gov>
- [5] <http://www.rcsb.org/pdb/>
- [6] S.F. Altschul *et al.*, *Nucleic Acids Res.*, 25:3389 (1997) [PMID: 9254694]
- [7] A. Sali and T.L. Blundell, *J.Mol.Biol.*, 234:779 (1993) [PMID:8254673]
- [8] R.A. Laskowski *et al.*, *J. Appl. Cryst.*, 26:283 (1993)
- [9] C.T. Porter *et al.*, *Nucleic Acids Res.*, 32: D129 (2004) [PMID: 14681376]
- [10] J.Wang *et al.*, *Cell.* 1997 91(4): 447(1997) [PMID: 9390554]
- [11] <http://cdb.ics.uci.edu/cgi-bin/BabelWeb.psp>
- [12] HyperChem (TM) Release 7.5, Hypercube, Inc., 1115 NW 4th Street, Gainesville, Florida 32601, USA
- [13] D. S. Goodsell *et al.*, *J Mol Recognit.*, 9: 1 (1996) [PMID: 8723313]
- [14] G. M. Morris *et al.*, *J. Computational Chemistry*, 19: 1639 (1998)
- [15] T. Castrignanò *et al.*, *Nucleic Acids Res.*, 34:D306 (2006) [PMID: 16381873]
- [16] M.F. Sanner, *J. Mol. Graphics Mod.*, 17:57(1999) [PMID: 10660911]

Edited by P. Kanguane

Citation: Tiwari *et al.*, Bioinformation 4(9): 405-408 (2010)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.

**Supplementary material:**

**Table 1:** Docked marine compounds

S. No.	Marine Compound name	IUPAC name	Molecular formula and Mol. Wt.	Docked Energy (kcal/mol)	Ref RMS
1	Ara-A	(2R, 3S, 4S, 5R)-2-(6-minopurin-9-yl)-5-(hydroxymethyl) tetrahydrofuran-3, 4-diol	C <sub>10</sub> H <sub>13</sub> N <sub>5</sub> O <sub>4</sub> 267.2413	-14.31	16.03
2	Dysinoin C	[3-[[1-[(2S,3aR,5S,6S,7aS)-2-[2-[1-(diaminomethylene)-2,5-dihydropyrrol-1-ium-3-yl]ethylcarbamoyl]-5,6-dihydroxy-2,3,3a,4,5,6,7,7a-octahydroindole-1-carbonyl]-2-methyl-propyl]amino]-2-methoxy-3-oxo-propyl] sulfate	C <sub>25</sub> H <sub>42</sub> N <sub>6</sub> O <sub>10</sub> S 618.7002	-14.90	14.25
3	Manoalide	5-hydroxy-4-[(2R,6R)-6-hydroxy-5-[(E)-4-methyl-6-(2,6,6-trimethyl-1-cyclohexenyl)hex-3-enyl]-3,6-dihydro-2H-pyran-2-yl]-5H-furan-2-one	C <sub>25</sub> H <sub>36</sub> O <sub>5</sub> 416.5503	-8.81	15.05
4	Nagelamide A	N-[3-[2-amino-5-[(E)-3-[(4,5-dibromo-1H-pyrrole-2-carbonyl)amino]prop-1-enyl]-3H-imidazol-4-yl]-3-(2-amino-3H-imidazol-4-yl)propyl]-4,5-dibromo-1H-pyrrole-2-carboxamide	C <sub>22</sub> H <sub>22</sub> Br <sub>4</sub> N <sub>10</sub> O <sub>2</sub> 778.0919	-20.49	13.36
5	Strobilin	(5Z)-5-[(5E,9E)-13-(3-furyl)-2,6,10-trimethyl-trideca-5,9-dienylidene]-4-hydroxy-3-methyl-furan-2-one	C <sub>25</sub> H <sub>34</sub> O <sub>4</sub> 398.5351	-8.02	15.6