Study of Drug Resistance and identification of Novel Drug Targets in *Mycobacterium Tuberculosis*

Under the Guidance of: - Dr. Jayashree Ramana

Group members:-

Saurabh Thakur (101504)

Adhiraj Singh Thakur (101506)

May 2014

Bachelor of Technology

In

Bioinformatics

DEPARTMENT OF BIOTECHNOLOGY AND BIOINFORMATICS

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY

WAKNAGHAT, SOLAN

CERTIFICATE

This is to certify that the work titled "**Study of Drug Resistance and Novel Drug Targets in *Mycobacterium tuberculosis***", submitted by Saurabh Thakur (101504) and Adhiraj Singh Thakur (101506) for the reward of Degree of Bachelor's Of Technology (Bioinformatics) from **JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT** has been carried out under my supervision. The work has not been submitted partially or wholly to any other university or institute for the award of this year or any other degree and diploma.

Signature of Supervisor

Name of Supervisor: Dr. Jayashree Ramana

Date:

Place: JUIT, Waknaghat

Acknowledgement

We take this opportunity to express our profound gratitude and deep regards to our guide, **Dr. Jayashree Ramana** for her exemplary guidance, monitoring and constant encouragement throughout the course of this project . The blessing, help and guidance given by her time to time shall carry us a long way in the journey of life on which we are about to embark.

We also take this opportunity to express a deep sense of gratitude to our Institution, college faculty and staff members for their cordial support, valuable information and guidance, which helped us in completing this task through various stages.

Name: Saurabh Thakur

Enrollment No. : 101504

Name: Adhiraj Thakur

Enrollment No. : 101506

Date: 15.05.2014

Place: JUIT, Waknaghat

Project Summary

Study of Drug Resistance and identification of Novel Drug Targets in Mycobacterium Tuberculosis (July 2013 - May 2014)

Brief description: An alarming emergence of multidrug-resistance and emerging total drug resistance in Mycobacterium tuberculosis and continuing high worldwide incidence of tuberculosis has invigorated the search for novel drug targets. Basically this project attempts to understand the mechanism behind the drug resistance in Mycobacterium tuberculosis and on the basis of this, finding out such novel drug targets that are crucial for Mycobacterium by using Computational Techniques.

This work accomplished two major objectives. First, the structure of novel drug targets (**FadH, Murl, Asps and Pks13)** was deduced computationally through homology modeling using the templates from bacterial homologues. Second, potent candidate inhibitors against these targets were identified through docking against already known inhibitors in other organisms or by using drug database.

For this, various **Bioinformatics software** and **tools** are used like **Discovery Studio, Gromacs, Easy Modeller, Zinc Database** etc.

TABLE OF CONTENTS

Introduction

Mycobacterium tuberculosis was discovered **by Dr. Robert Koch in 1882.** It is a pathogenic bacterial species in the family **Mycobacteriaceae** and the causative agent of most cases of tuberculosis (TB) [1]. **Tuberculosis** typically attacks the lungs, but can also affect other parts of the body. It is spread through the air when people who have an active TB infection cough, sneeze, or otherwise transmit respiratory fluids through the air.

There are two types of tuberculosis based on symptoms namely:

- **Pulmonary Tuberculosis:**
  If a tuberculosis infection does become active, it most commonly involves the lungs. Symptoms may include chest pain and a prolonged cough producing sputum. About 25% of people may not have any symptoms. Occasionally, people may cough up blood in small amounts, and in very rare cases, the infection may erode into the pulmonary artery, resulting in massive bleeding Tuberculosis may become a chronic illness and cause extensive scarring in the upper lobes of the lungs [13].
- **Extrapulmonary Tuberculosis:**
  In 15–20% of active cases, the infection spreads outside the lungs, causing other kinds of TB. These are collectively denoted as **extrapulmonary tuberculosis**. Extrapulmonary TB occurs more commonly in immunosuppressed persons and young children. In those with HIV, this occurs in more than 50% of cases [13].

The high mortality and morbidity rate [1] is largely attributed to the escalating resistance of the causal bacterium Mycobacterium tuberculosis (Mtb) to the currently available drugs, precipitating in the emergence of Multi-Drug Resistant (MDR) (resistant to isoniazid and rifampicin) and Extensively Drug Resistant (XDR) (resistant to isoniazid and rifampicin as well as fluoroquinolone and second line injectable agents) strains. This has intensified the search for novel drug targets in Mtb for chemotherapeutic intervention [2].

This very fact has been the motivation of our project, where we have not only tried to study the effect and mechanism of current line of drugs and the mechanism of resistance to them, but also tried to find a novel line of drugs by thorough computational studies and their validation.

Objective

This work accomplished three major objectives:

- First, to identify out possible drug targets.
- Second, to model the structure of possible drug targets.
- Third, to find out potent inhibitors for these drug targets either in existing organisms or in existing drug databases.

Important statistics

Global:

- Tuberculosis is second only to HIV/AIDS as the greatest killer worldwide due to a single infectious agent.
- In 2012, 8.6 million people fell ill with TB and 1.3 million died from TB.
- In 2012, an estimated 530 000 children became ill with TB and 74 000 HIV-negative children died of TB [3].

India:

- Tuberculosis is the biggest health issue that lies around India, but what makes is worse is the newly and recently discovered global phenomenon of **TDR-TB - Totally Drug-Resistant Tuberculosis** which is **most dangerous one** [4].
- An experiment was conducted at **Hinduja Hospital** in **Mumbai** in January, 2012 on four patients to test how accurate the "new category" of **TDR-TB** is. These patients were given all the **first-line drugs** and **second-line drugs** that usually are prescribed to treat TB, and as a result were resistant to all.
- Given below is the number of cases of Tuberculosis followed by the years , in India:

year(2005-2011)	Total TB cases in India
2005	1,294,550
2006	1,400,340
2007	1,474,605
2008	1,517,363
2009	1,533,309
2010	1,522,147
2011	1,515,872

In 2012, nearly 9 million people around the world became sick with TB disease. There were around 1.3 million TB-related deaths worldwide.

Thus these numbers prove that Tuberculosis has become a giant killer and is a threat to mankind that needs to be eliminated quickly.

Drug Resistance

Drug resistance is the reduction in effectiveness of a drug in curing a disease or condition. When the drug is not intended to kill or inhibit a pathogen, then the term is equivalent to dosage failure or drug tolerance. More commonly, the term is used in the context of resistance that pathogens have acquired, that is, resistance has evolved. When an organism is resistant to more than one drug, it is said to be multidrug-resistant [5].

Mechanisms

The four main mechanisms by which microorganisms exhibit resistance to antimicrobials are:

1. Drug inactivation or modification.
2. Alteration of target site
3. Alteration of metabolic pathway
4. Reduced drug accumulation by decreasing drug permeability or increasing active efflux (pumping out) of the drugs across the cell surface [5]

Current Drug line

1) Isoniazid (INH)

It is the organic compound that is the first line medication in prevention and treatment of MTB. INH is a prodrug (medication that administered to the body in inactive form) and must be activated by a bacterial catalase peroxidase enzyme which is encoded by KatG gene in MTB.

KatG gene couples the isonicotinic acyl with NADPH to form isonicotinic acyl NADPH complex. This complex binds tightly to the enoyl acyl carrier protein reductase known as InhA (this enzyme is a key enzyme of fatty acid synthesis), so blocking the action of fatty acid synthase. This process inhibits the synthesis of Mycolic acid which required for bacterial cell wall. The presence of mycolic acids gives MTB many characterstics that stop medical treatment. They lend the organisms increased resistance to chemical damage and dehydration [6].

Isoniazid reaches therapeutic concentrations in serum, cerebrospinal fluid, and within caseous granulomas. It is metabolized in the liver via acetylation. Two forms of the enzyme are responsible for acetylation, so some patients metabolize the drug more quickly than others. Hence, the half-life is bimodal, with peaks at one and three hours in the US population. The metabolites are excreted in the urine. Doses do not usually have to be adjusted in case of renal failure.

Mutation

Mutations leading to INH resistance have been identified in different gene targets like **KatG** and **InhA**. Amino acid replacements in the NADPH binding site of InhA result in INH resistance by preventing inhibition of mycolic acid biosynthesis.

A mutation in the KatG gene causes the enzyme catalase peroxidase unable to convert INH to its biologically active form. Ser315Thr substitution in KatG gene and in inhA gene InhA promoter mutations are more frequently seen and are present at positions -24(G-T), -16(A-G), or -8(T-G/A) and -15(C-T) approximately 70–80% of INH resistance in clinical isolates of M. tuberculosis can be attributed to mutations in the katG and inhA genes [6].

2) Rifampicin(RIF)

This drug binds to the beta subunit of RNA polymerase which is encoded by rpoB gene and to disrupt translation and RNA synthesis (elongation). It binds to the RNA polymerase at a site adjacent to the RNA polymerase active center and blocks the RNA synthesis. Crystal structure

data and biochemical data indicate that rifampicin binds to RNA polymerase at a site adjacent to the RNA polymerase active center and blocks RNA synthesis by physically preventing extension of RNA products beyond a length of 2-3 nucleotides [6].

Mutation

Most common mutation occurs in codons Ser531Leu, His526Tyr and Asp516Val. These changes occur in more than 70% of RIF's resistant isolates, also nearly 90% of RIF resistant strains are also INH resistant. Due to these mutations beta subunit protein now has different amino acid and so different confirmation and therefore RIF can no longer bind to the beta subunit of RNA polymerase [6].

3) Ethambutol (EMB)

It is a bacteriostatic agent that is active for growing bacilli and no effect on non replicating bacilli. Since cell wall have mycolic acids and also have arabinoglactan which is a biopolymer consist of Arabinose and galactose monosacchrides.

So mycolic acid attach to 5' hydroxyl group of Arabinose residues and form complex known as mycolyl-arabinoglactan peptidoglactan which is important in cell wall as it does not allow drugs to enter in the bacteria. So EMB interfere with the synthesis of Arabinoglactan by inhibit Arabinosl transferase (involved in polymerization of Arabinoglactan).

So disruption of Arabinoglactan synthesis inhibits the formation of this complex and lead to increased permeability of cell wall. So mainly EMB bind to Arabinosyl transferase enzyme (encoded by embB gene) to inhibit it.

Mutation

Five mutations in (embCAB) codon 306 [(ATG-GTG), (ATG-CTG), (ATG-ATA), (ATG-ATC) and (ATG-ATT)] which result in three different amino acid substitutions (Val, Leu and Ile) in EMB-resistant isolates. These five mutations are associated with 70–90% of all EMB resistant isolates. Missense mutations were identified in three additional codons: Phe285leu, Phe330Val and Thr630Ile in EMB resistant isolates.

4) Pyrazinamide (PZA)

It is a prodrug meaning, first it is inactive and then it converts into active form from Pyrazinamide to Pyrazinoic acid (POA) with the help of enzyme Pyrazinamidase which is encoded by pncA.

POA inhibit the fatty acid synthase which is required for synthesis of fatty acids. Under acidic condition, POA slowly convert into protonated acid which is diffuse easily back in the cell and then accumulate. But POA accumulate inside the baccilus at acidic pH not at neutral pH. This accumulation of POA in the cytoplasm results in the lowering of intracellular pH to a level that inactivates vital fatty acid synthase [6].

Mutation

Mutation in the pncA gene which encode pyrazinamidase enzyme is responsible for most pyrazinamidase resistance in MTB strain. pncA mutations are highly diverse and scattered along the gene despite this, there is some degree of clustering at three regions of the PncA, 3–17, 61–85 and 132–142. Most PZA-resistant M. tuberculosis strains (72–97%) have mutations in pncA [6].

Novel Potential Drug Targets

Initially, BLAST was performed between Human genome and Mycobacterium genome so as to find similar proteins and domains. Basically BLAST is done for finding the homologs.

Along with results of BLAST and literature reading, the following interesting potential drug targets were obtained: **2, 4-Dienoyl CoA reductase** which is encoded by **Fadh** gene in Mycobacterium, **Glutamate racemase** which is encoded by **Murl** gene**, Aspartyl synthetase** which is encoded by **Asps** gene and **polyketide synthase** which is encoded by **Pks13** gene in Mycobacterium [7].

1) 2, 4-Dienoyl CoA reductase

This enzyme can be a potential drug target as this enzyme is required for a very important process that is **beta oxidation** of **polyunsaturated fatty acids**.

Beta-oxidation (in human it occurs in mitochondria) is the process by which fatty acid molecules are broken down to Acetyl coenzyme A. Acetyl coenzyme A or acetyl-CoA is an important molecule in metabolism, used in many biochemical reactions. Its main function is to convey the

carbon atoms within the acetyl group to the citric acid cycle to be oxidized for energy production generate acetyl-coA [8].

So evidently, this enzyme **(2, 4-Dienoyl CoA reductase)** is very important in this process and if we can inhibit this enzyme, we may be able to disrupt this process.

2) Glutamate racemase

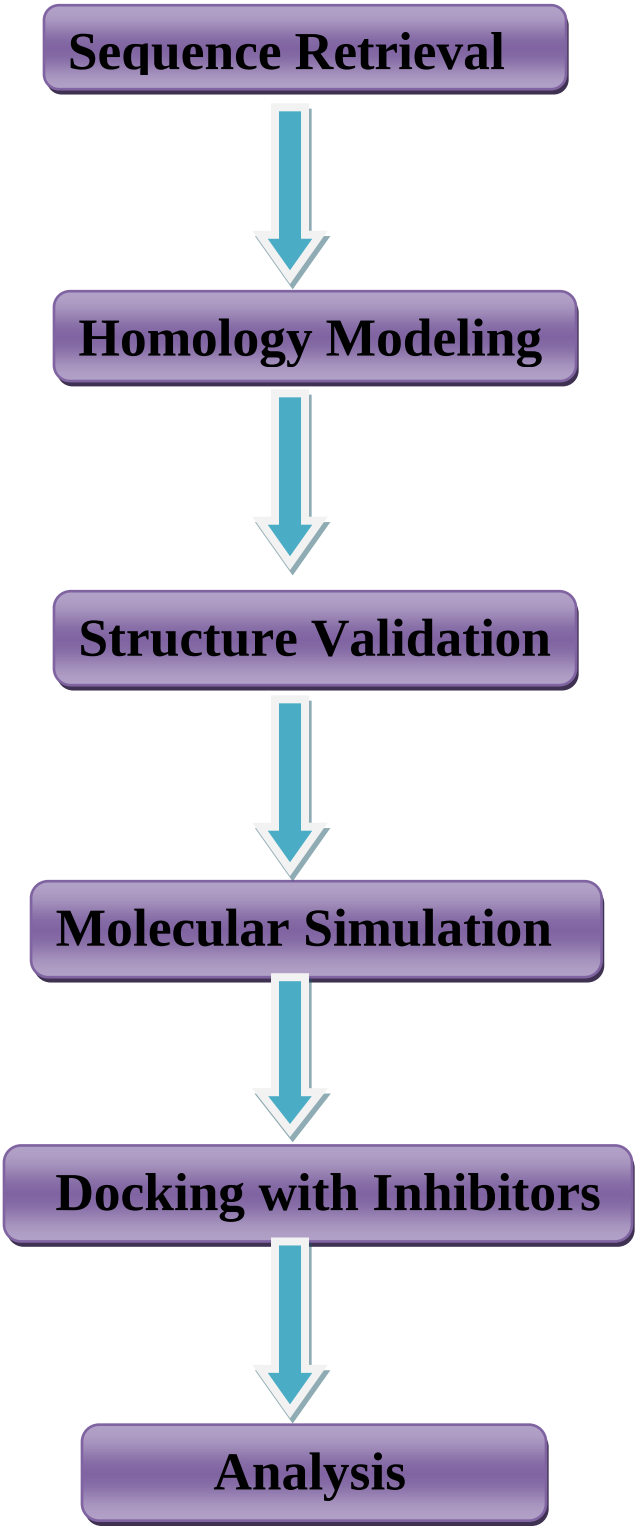
Glutamate racemase (MurI) serves two distinct metabolic functions: primarily, it is a critical enzyme in **cell wall biosynthesis**, but also plays a role in **gyrase inhibition** [9].

The bacterial cell wall is composed of **peptidoglycan** which provides rigidity to the bacterial cell wall. The synthesis of the peptidoglycan polymer involves the action of **Glutamate racemase**. This enzyme participates in glutamate metabolism that is essential for cell wall biosynthesis in bacteria. Glutamate racemase performs the additional function of gyrase inhibition, preventing gyrase from binding to DNA.

3) Polyketide synthase

This enzyme basically involve in the mycolic acid biosynthesis in bacterial cell wall. So this protein is very important component of cell wall (particularly fatty acid biosynthesis) [10].

Methodology



1) Sequence Retrieval

The amino acid sequences for four different genes were retrieved from NCBI Protein database (**http://www.ncbi.nlm.nih.gov/protein/**) known as query sequence.

1) Homology modeling

Homology modeling refers to constructing a model of the "target" protein from its amino acid sequence and an experimental three-dimensional structure of a related homologous protein (the "template"). Homology modeling relies on the identification of one or more known protein structures likely to resemble the structure of the query sequence and on the production of an alignment that maps residues in the query sequence to residues in the template sequence [11].

To create a homology model for the selected targets, BLAST search was first performed using the query sequence and the PDB database (**http://www.pdb.org/**) and then we found the best template on the basis of BLAST score and then did the homology modeling and generate 20 models by using software name as **Easy Modeler 4.0.**

2) Structure Validation

These 20 models were then validated on the basis of **RC plot, Verify-3D** and **Errat score** which were done on online server name as **Saves server.**

RC plot basically checks the stereo chemical quality of a protein structure by analyzing residue-by-residue geometry and overall structure geometry.

Verify-3D determines the compatibility of an atomic model (3D) with its own amino acid sequence (1D).

Errat score analyzes the statistics of non-bonded interactions between different atom types.

So on the basis of **plot, Verify-3D** and **Errat score,** out of 20 models best model was selected.

3) Molecular Simulation

Molecular simulation was performed to find out the most stable conformation of protein structure.

Minimization

Energy minimization is a procedure that attempts to minimize the potential energy of the system to the lowest possible point [12].

Equilibration

For most systems, this process will involve allowing some solvent to relax around the solute of interest. In many cases, the solvent is water, but it can also be a lipid bilayer, chloroform, etc [13] and in our case the solvent was water.

Dynamics

The dynamics stage is the stage of interest for determining thermodynamic averages or sampling new configurations. The stage used for these applications is sometimes known as production dynamics [14].

After selection of best model, Simulation was done by using software name as **Gromacs.** Simulation was done by using parameter temperature as 300 Kelvin and at ph 7.0 and every 1 femto second, new conformation of that protein structure was generated and most stable conformation was selected.

4) Docking and Analysis

After simulation, the structures of inhibitors were retrieved from **NCBI PubChem Bioassay database** and from **Zinc database**. These inhibitors were docked to the homology model of protein targets using **CDOCKER** program in **Discovery Studio 3.5** and then select the best inhibitors based on cdocker energy (minimum cdocker energy means best inhibitor and which indicates more favorable binding).

Results

A) Glutamate Racemase

1) Sequence Analysis

Sequence of this protein was retrieved from NCBI as accession id NP_215854.1 and did the sequence alignment by using Protein BLAST against Protein database and finds the best homolog (**3HFR substrate free and 2JFP substrate bound with D-glu**) of this protein. Homolog was selected on the basis of its query coverage and its identity.

Figure 1 displays the sequence alignment of MtMurI (NP_215854) with MurI homologs from different bacterial species. It is seen that MtMurI shares high degree of similarity with its homologs as both query coverage (94%) and identity (42%) is good.

Chain A, Crystal Structure Of Glutamate Racemase From Listeria Monocytogenes

Sequence ID: pdb|3HFR|A **Length:** 269 **Number of Matches:** 1

See 5 more title(s)

Range 1: 2 to 256 GenPept Graphics Next Match Previous Match

Score	Expect	Method	Identities	Positives	Gaps
200 bits(508)	5e-62	Compositional matrix adjust.	108/256(42%)	148/256(57%)	2/256(0%)

```
Query  2    NSPLAPVGVFDSGVGGLTVARAIIDQLPDEDIVYVGDTGNGPYGPLTIPEIRAHALAIGD  61
            N+    +G  DSGVGGLTV R ++ QLP E + Y+GDT   PYGP    E+         +
Sbjct  2    NAXKQAIGFIDSGVGGLTVVREVLKQLPHEQVYYLGDTARCPYGPRDKEEVAKFTWEXTN  61

Query  62   DLVGRGVKALVIACNSASSACLRDARERYQVPVVEVILPAVRRAVAATRNGRIGVIGTRA  121
             LV RG+K LVIACN+A++A L D RE+  +PV+ VI P  R A+ ATRN +IGV+GT
Sbjct  62   FLVDRGIKXLVIACNTATAAALYDIREKLDIPVIGVIQPGSRAALKATRNNKIGVLGTLG  121

Query  122  TITSHAYQDAFAA-ARDTEITAVACPRFVDFVERGVTSGRQVLGLAQGYLEPLQRAEVDT  180
            T+ S AY  A     R  E+ ++ACP+FV  VE G          +    L PL+  ++DT
Sbjct  122  TVESXAYPTALKGLNRRVEVDSLACPKFVSVVESGEYKSAIAKKVVAESLLPLKSTKIDT  181

Query  181  LVLGCTHYPLLSGLIQLAMGENVTLVSSAEETAKEVVRVLTEIDLLRPHDAPPATRIFEA  240
            ++LGCTHYPLL  +I+   G+ V +++S EETA EV  +L   +LL   D     R F
Sbjct  182  VILGCTHYPLLKPIIENFXGDGVAVINSGEETASEVSALLDYHNLLDATDEEIEHRFF-T  240

Query  241  TGDPEAFTKLAARFLG  256
            TG  + F  +A  +L
Sbjct  241  TGSTQIFKDIAKDWLN  256
```

Fig. 1 Sequence alignment of MtFadH with its Homolog

2) Homology Modeling and Structure Validation

Homology models were obtained for MtMurI using the templates in the substrate-free (Figure 2.1) as well as substrate-bound (Figure 2.2) form. The best model was selected on the basis of its **RMSD (root mean square deviation) value, verify-3D score, errat score** and **RC plot**.

Figure 3.1 and 3.2 shows the two best models superimposed with their homologs and their RMSD value were also good that is for the free and bound models with their respective templates stood low at 0.90 and 0.85 Å respectively.

Figure 4.1 shows the RC plot, verify-3D score and Errat score for the model which was modeled by using substrate bound template and not any single residue lie in unfavoured reagion.

Figure 4.2 shows the RC plot, verify-3D score and Errat score for the model which was modeled by using substrate free template and for this structure also not any residue lie in unfavoured reagion.

The active site of MtMurI (Figure 5) consists of the residues Cys75, His187, Asp12, Ser13, Thr11 Pro43, Tyr44, Val149, Gly45, Ser77, Asn7 and Thr186 involved in hydrogen bonding interactions with D-Glu.

As RMSD value for both structure were low, not a single residue lie in unfavoured reagion and other scores were also good, so this depicts that our selected models were good. Now these structures can undergo simulation and further docking process.

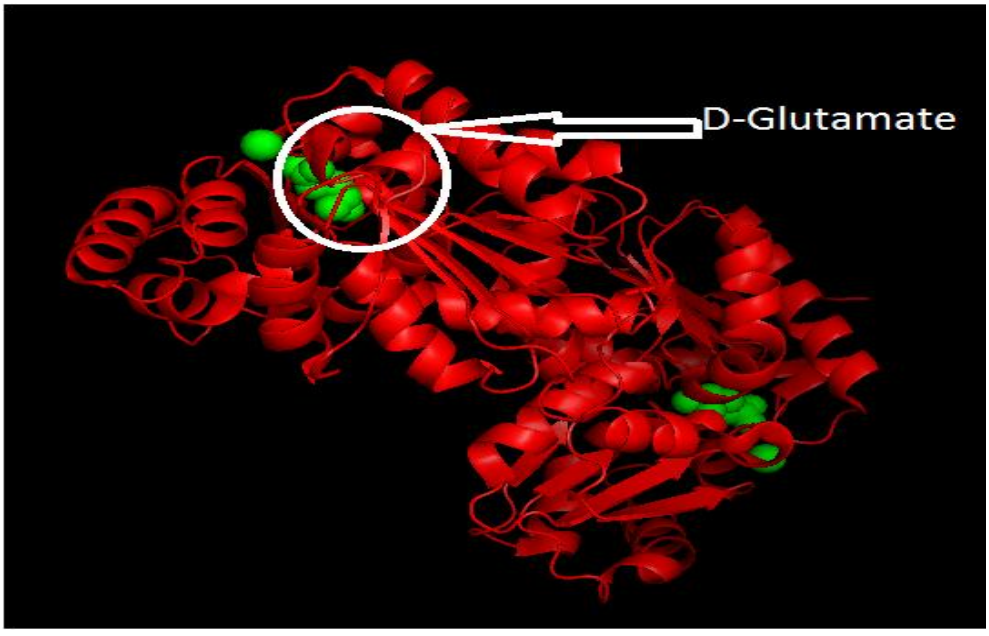


Figure 2.1 the structure modeled MurI with 2JFP template (substrate bound)

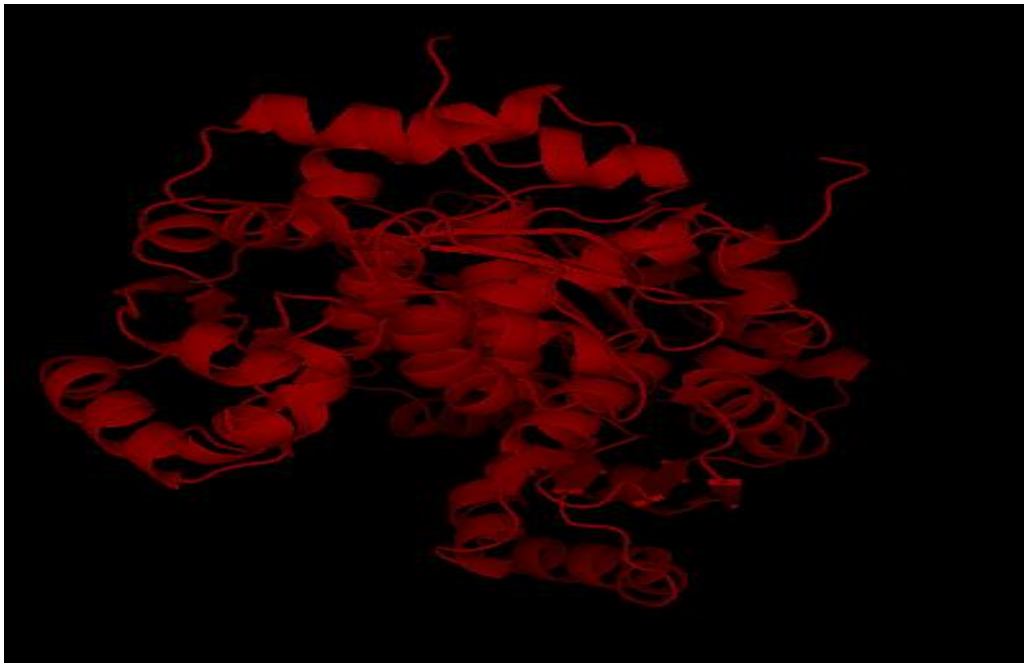


Figure 2.2 the structure modeled MurI with 3HFR template (substrate free)

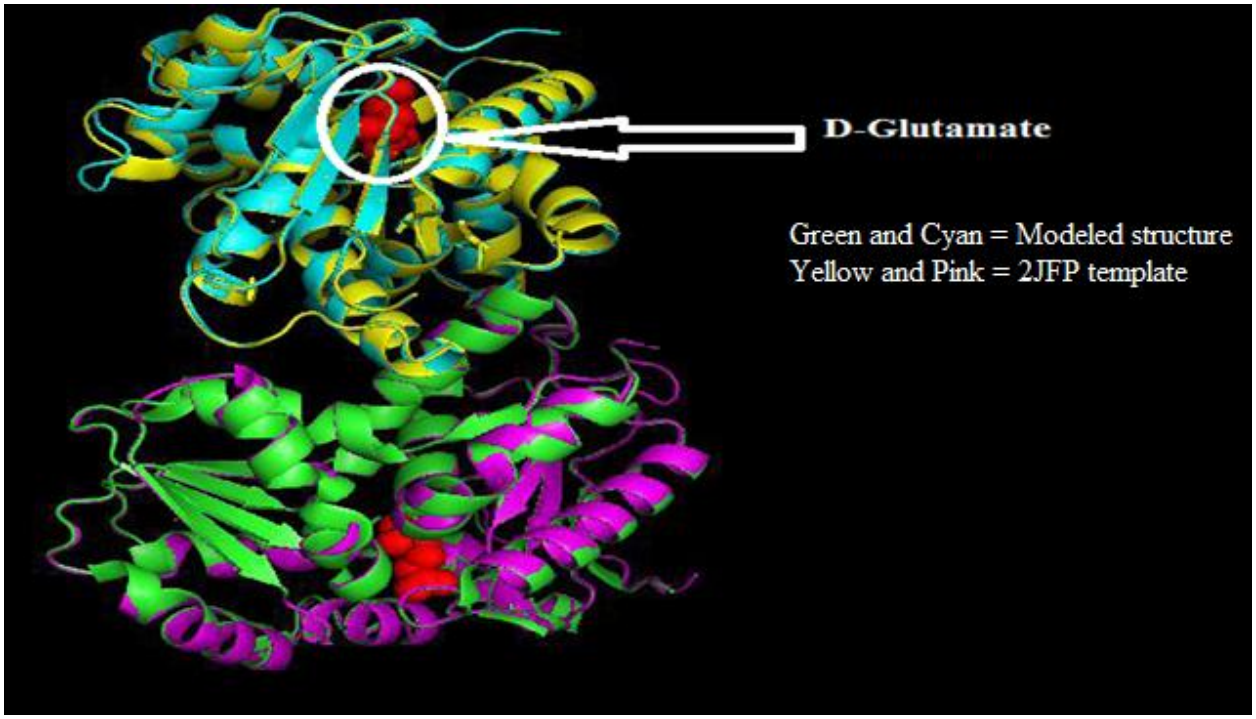


Fig 3.1 Superimposition of the models for substrate-bound MtMurI with its templates 2JFP.

Fig 3.2 Superimposition of the models for substrate-free MtMurI with its templates 3HFR

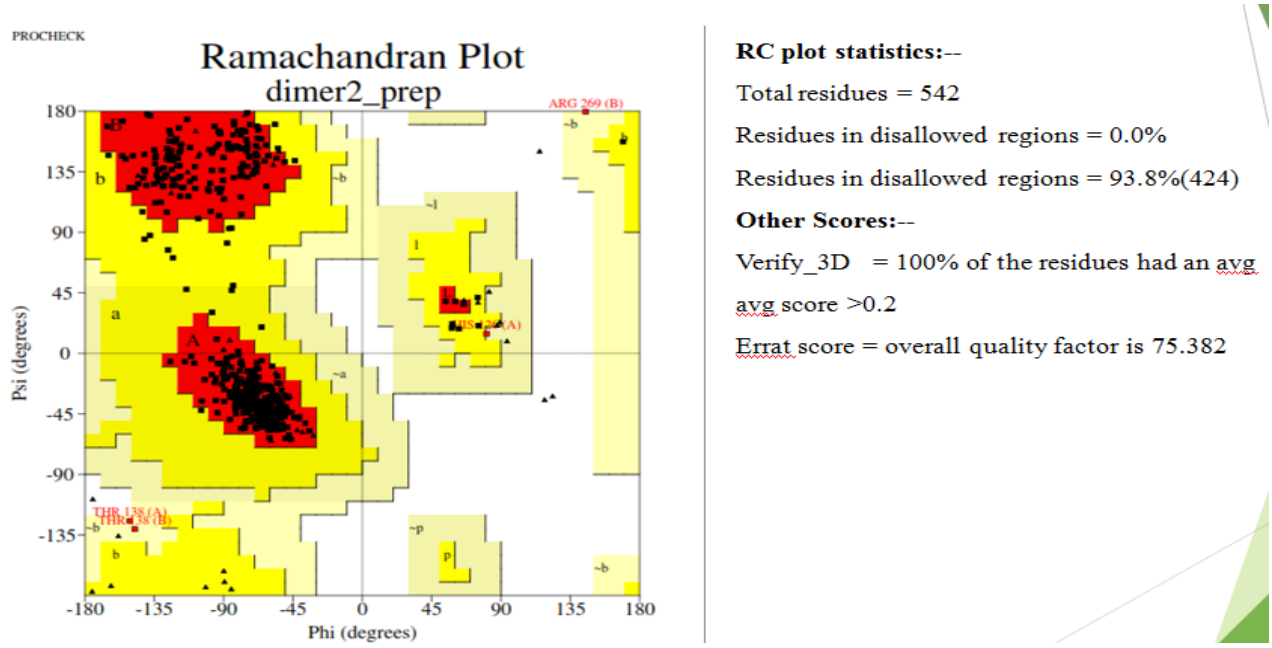


Fig 4.1 Ramachandran plots for model generated for substrate-bound (2JFP) of the enzyme MtMurI and some other scores

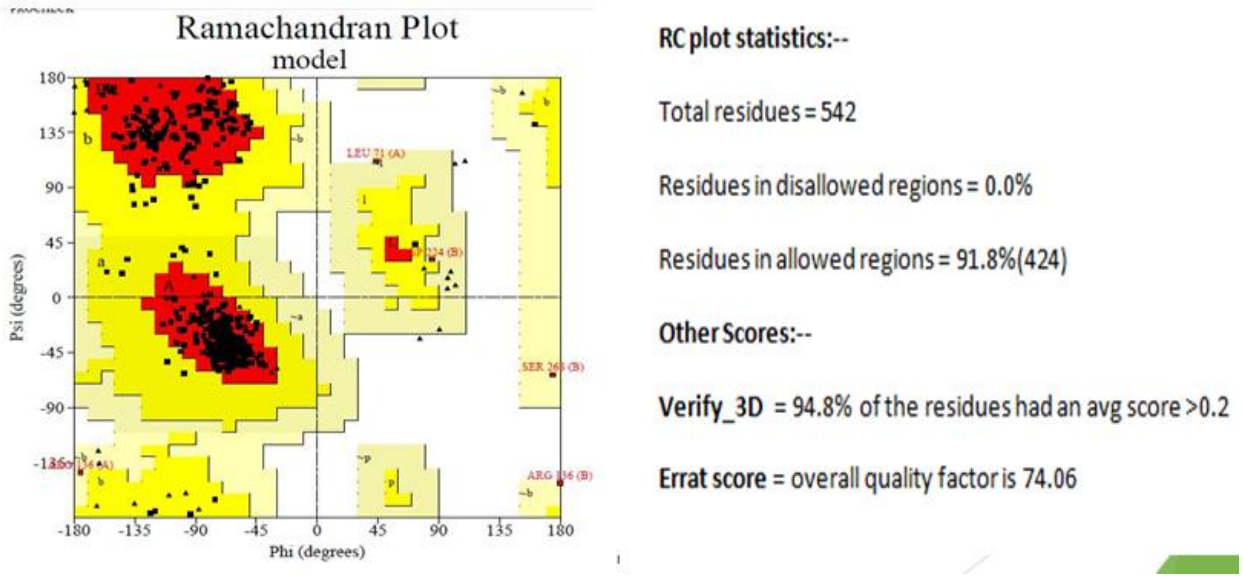


Fig 4.2 Ramachandran plots for model generated for substrate-free (3HFR) of the enzyme MtMurI and some other scores

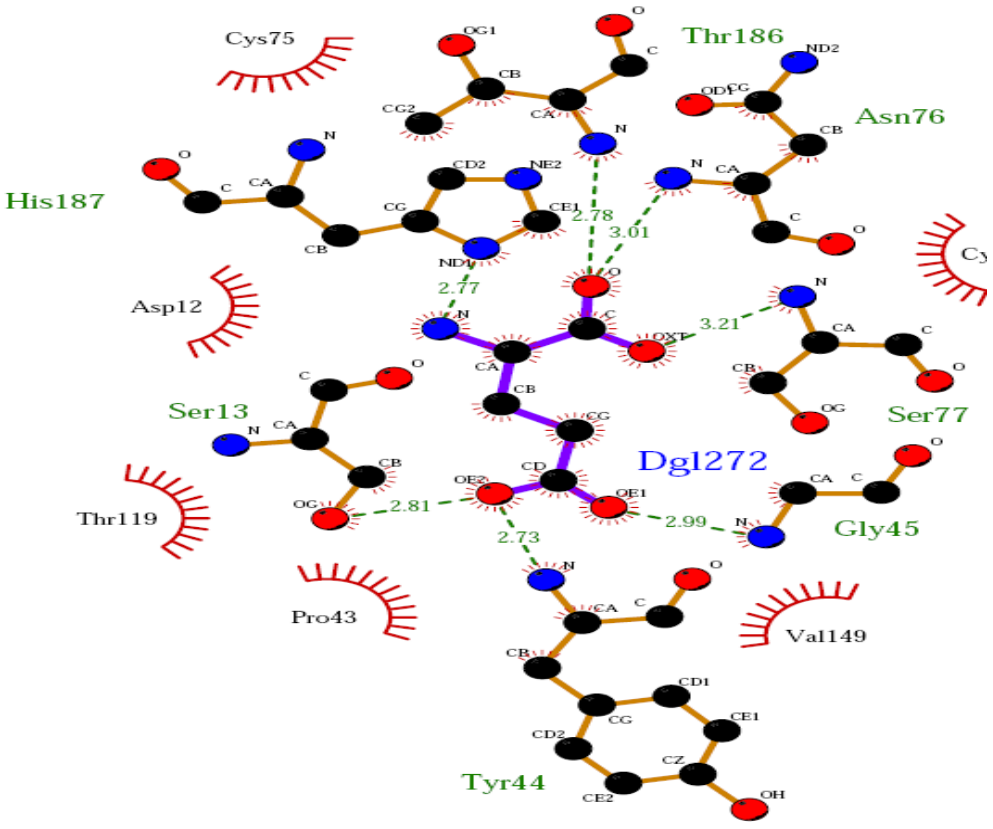


Fig. 5 Interactions between MtMurI and D-Glu

3) Molecular Simulation

After validation of modeled structures and selecting the best one, simulation was done on them and most stable conformation (Fig 6) was retrieved.

There are three typical stages that happen in molecular dynamics simulation:

Minimization

Energy minimization is a procedure that attempts to minimize the potential energy of the system to the lowest possible point [12].

Equilibration

For most systems, this process will involve allowing some solvent to relax around the solute of interest. In many cases, the solvent is water, but it can also be a lipid bilayer, chloroform, etc [13] and in our case the solvent was water.

Dynamics

The dynamics stage is the stage of interest for determining thermodynamic averages or sampling new configurations. The stage used for these applications is sometimes known as production dynamics [14].

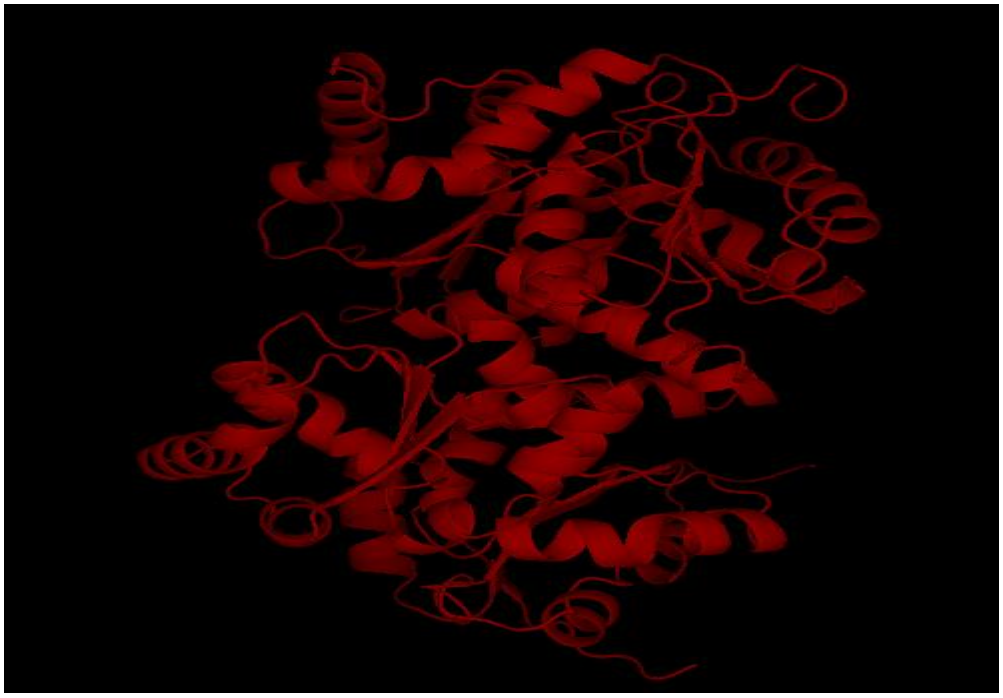


Fig 6. Most stable conformation of MurI

4) Docking and Analysis

The structures of already known inhibitors for MurI I in other organisms were retrieved from NCBI PubChem Bioassay database. 46 inhibitors were obtained from the Bioassay ID 208942. These represent competitive inhibitors analogues and were docked to the D-glutamate binding site using the substrate-bound homology model (after removing the D-glutamate).

The most competitive inhibitors were found on the basis of CDOCKER score (Table 1). Lesser will be the cdocker score more will be the binding affinity of inhibitors and inhibitors will be more potent.

Each these potent inhibitors were analyse by seeing their interaction with the binding site means which atom of ligand is binding with which residue of MurI protein (Figure 7).

Molecular ID	CDOCKER energy
11076721	-57.1651
11066697	-56.0058
10970765	-55.408
10960522	-54.2701
10925032	-54.0107
10915261	-53.367
11120288	-53.1177

Table 1 the top seven competitive inhibitors for MtMurI on the basis of cdocker energy

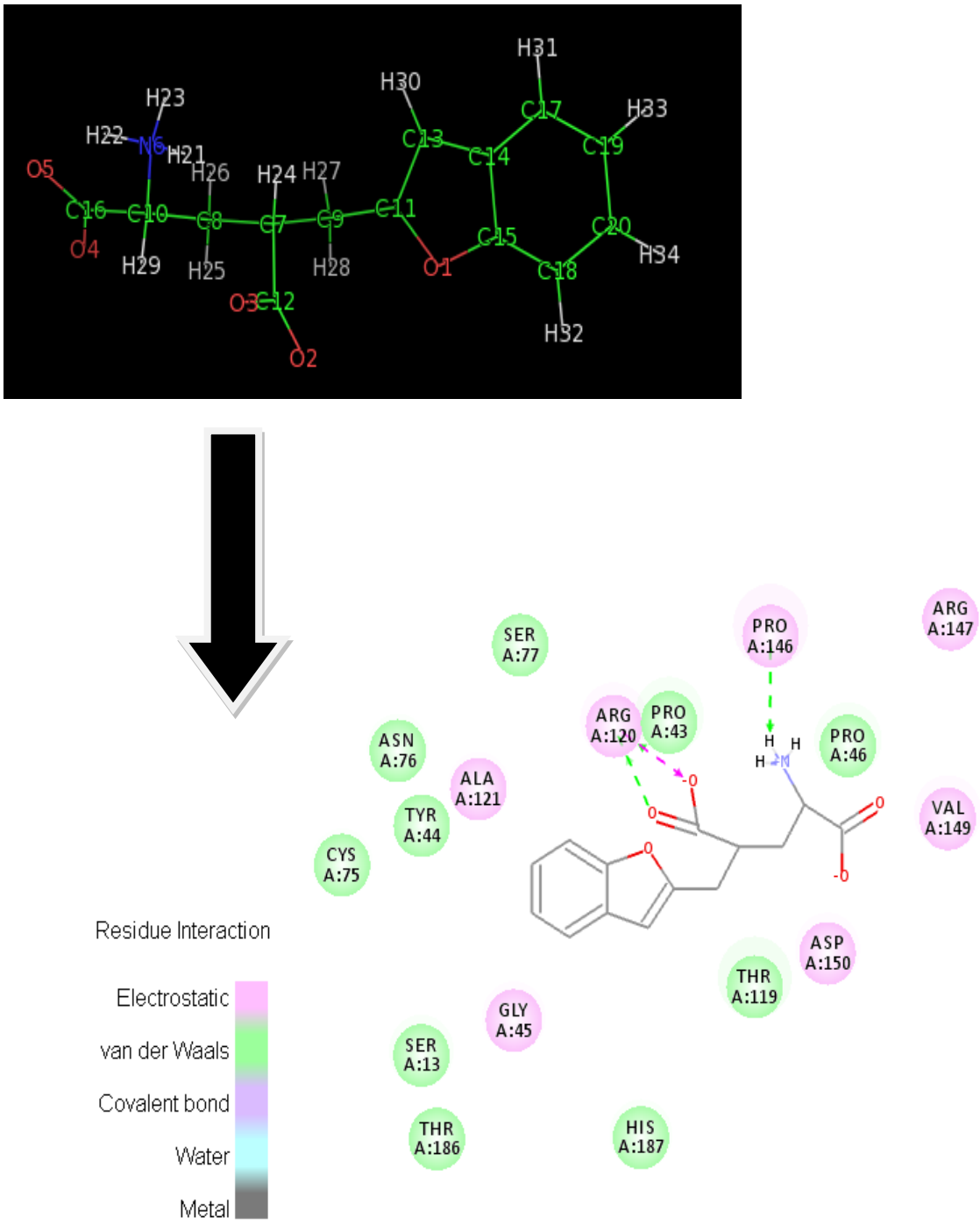


Figure 7.1 Interaction of 11076721 molecule with its binding site

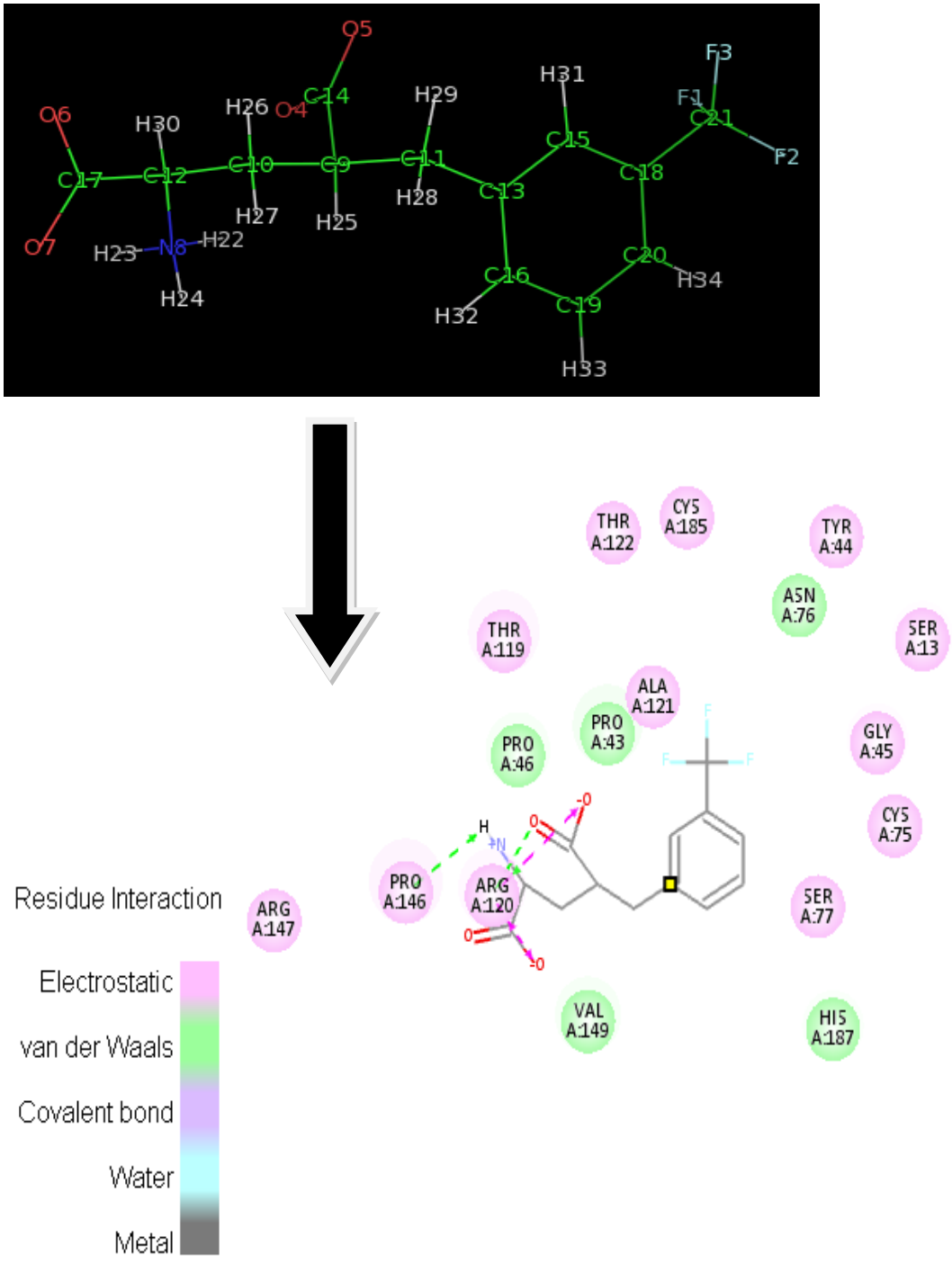


Figure 7.2 Interaction of 11066697 molecule with its binding site

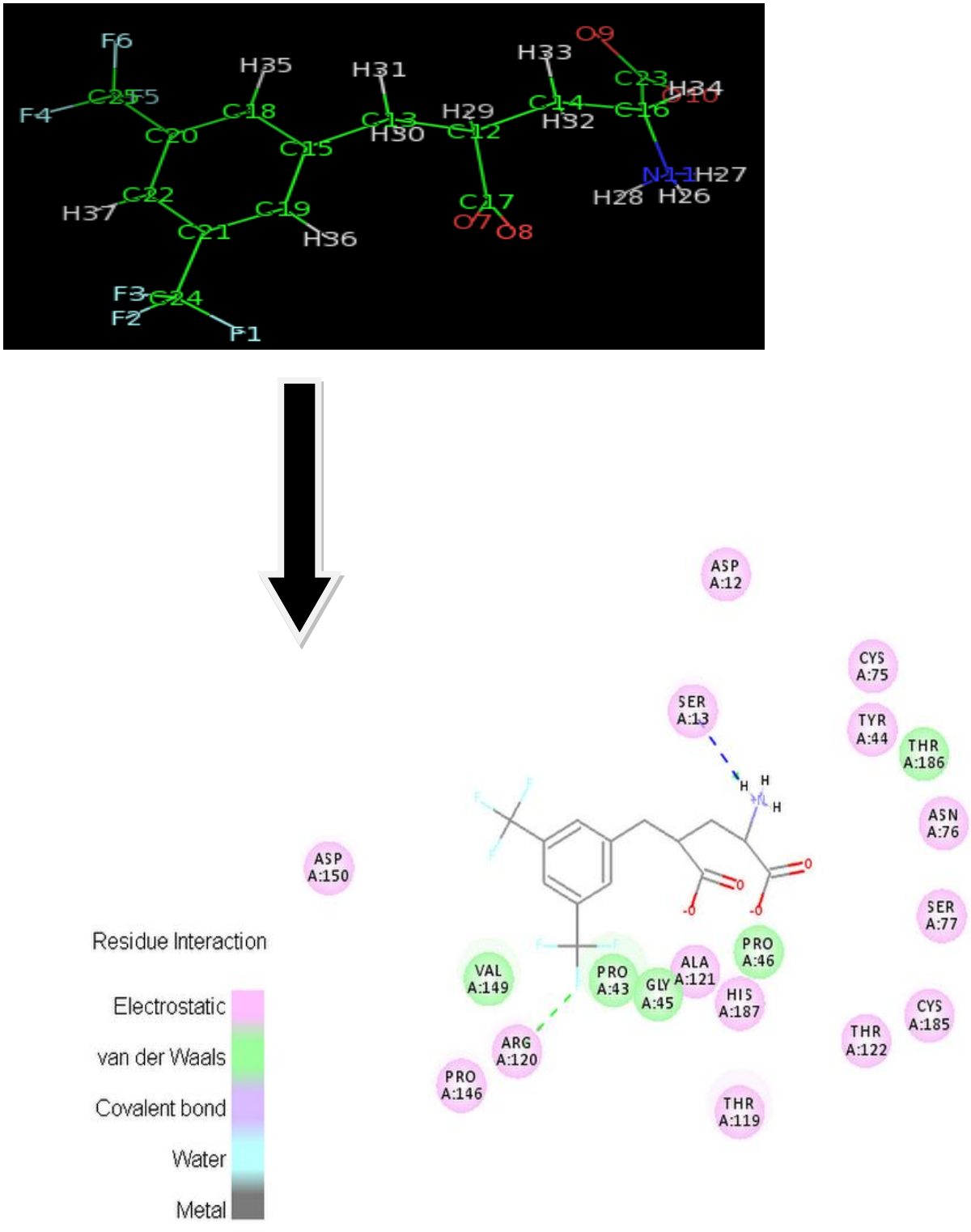


Figure 7.3 Interaction of 10970765 molecule with its binding site

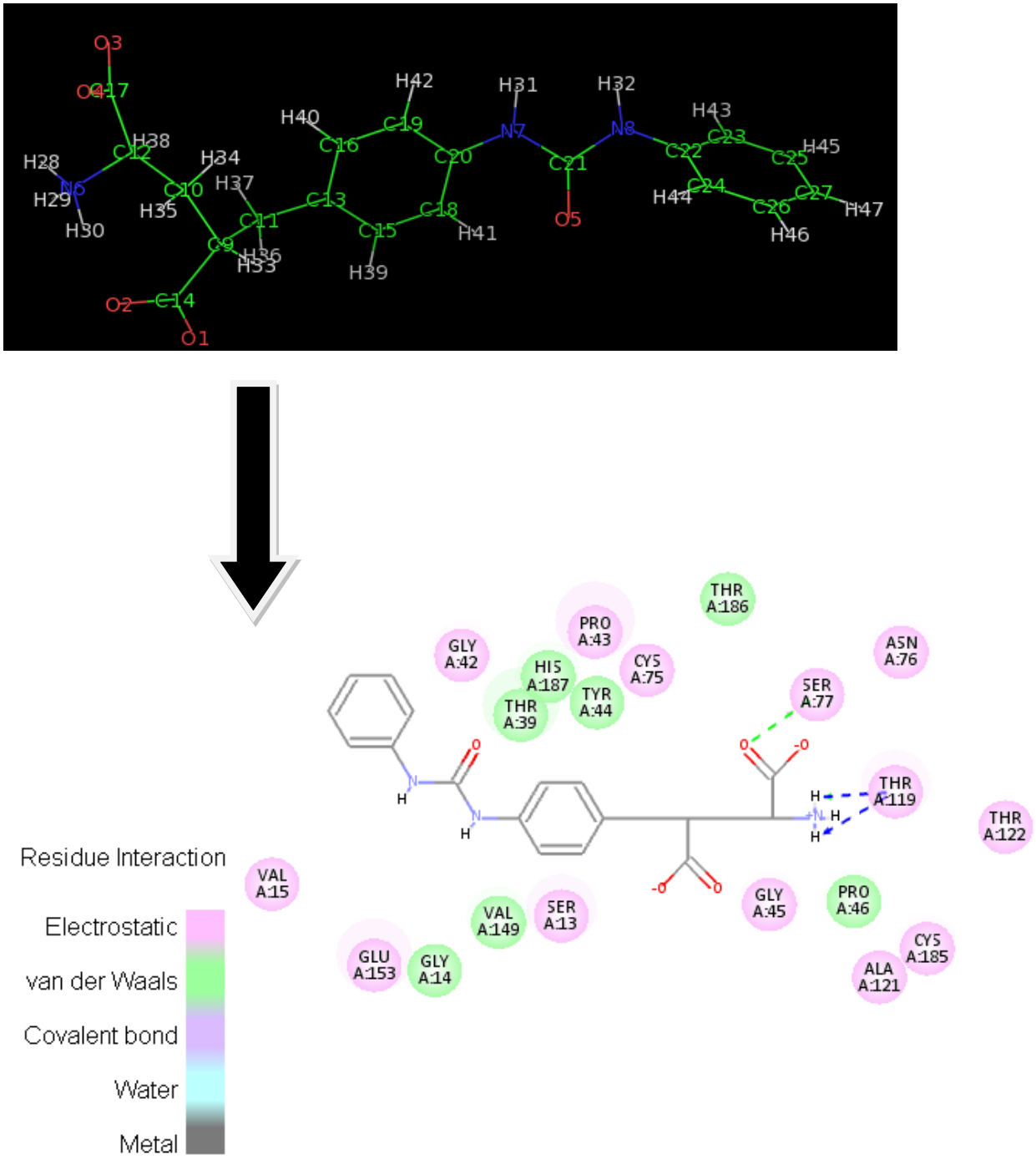


Figure 7.4 Interaction of 10960522 molecule with its binding site

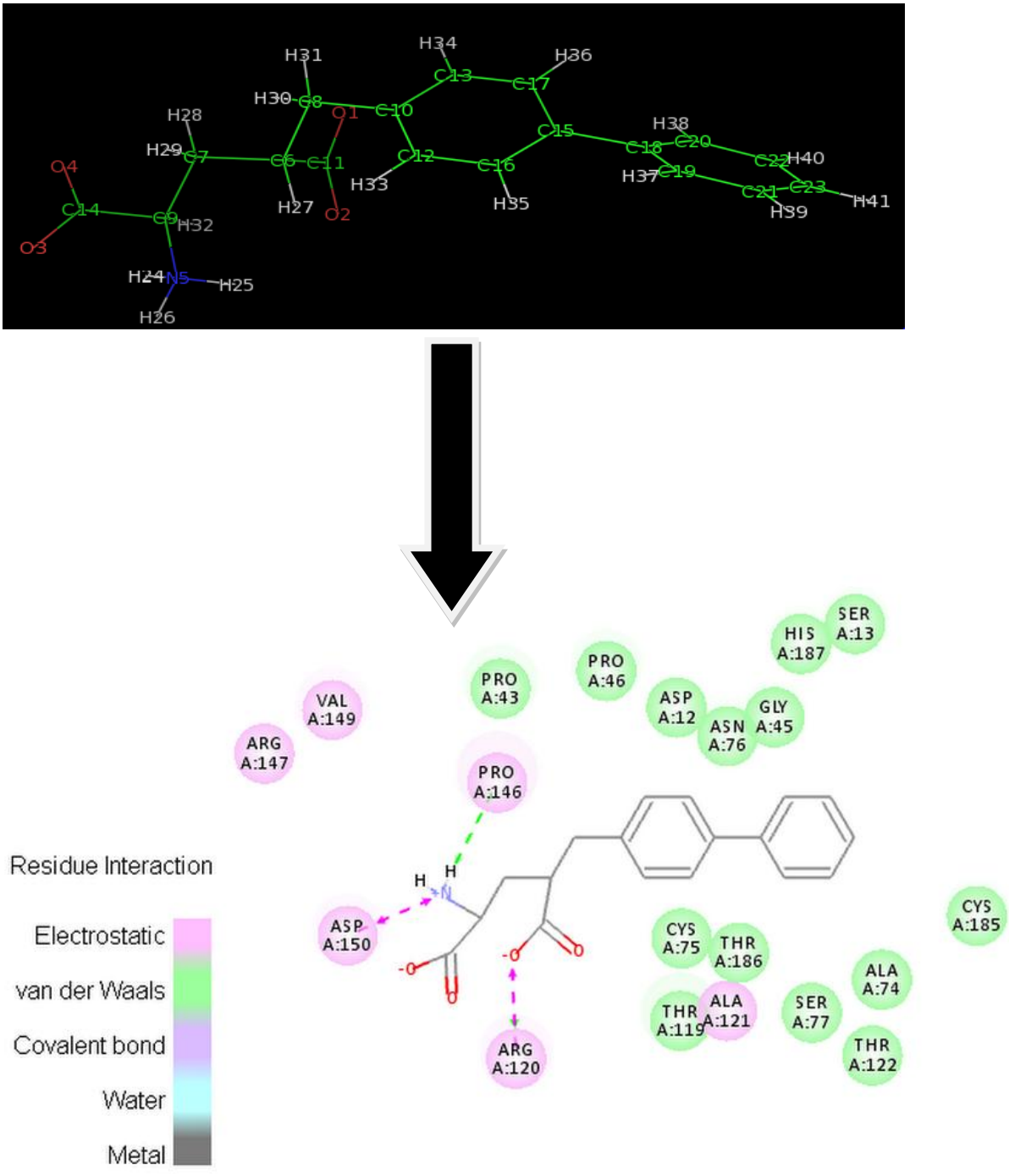


Figure 7.5 Interaction of 10925032 molecule with its binding site

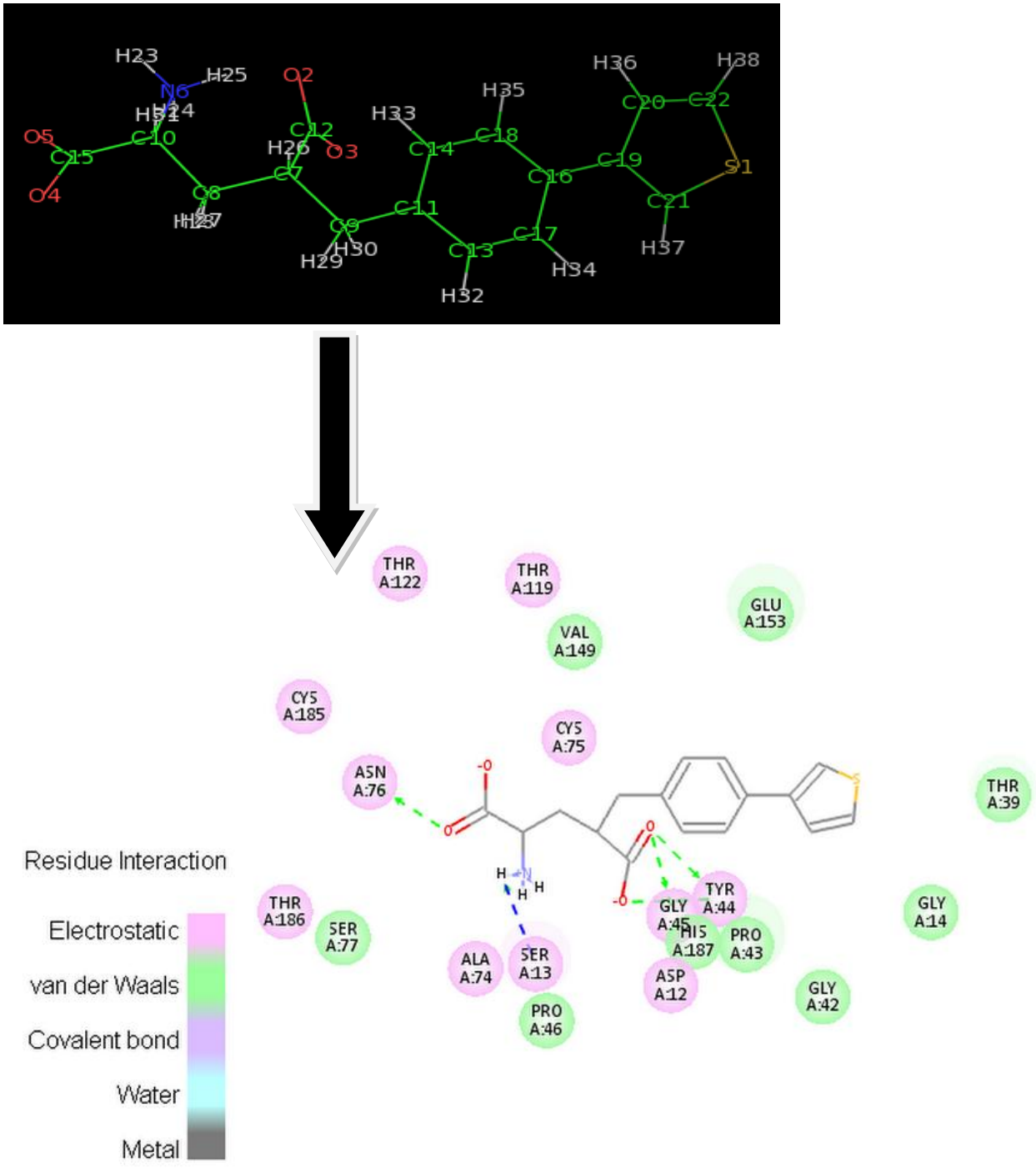


Figure 7.6 Interaction of 10915261 molecule with its binding site

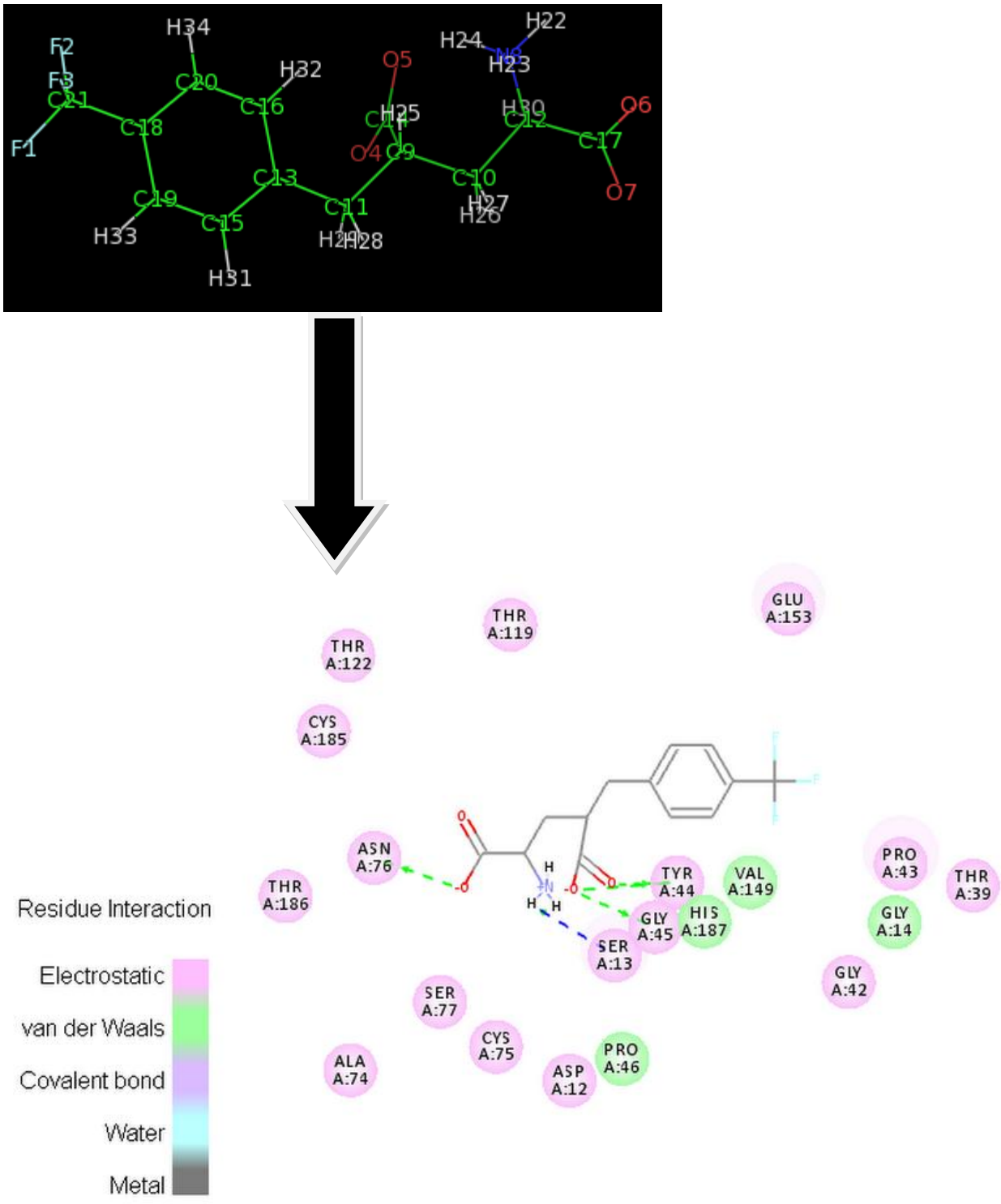


Figure 7.7 Interaction of 11120288 molecule with its binding site

B) 2, 4-Dienoyl CoA reductase

1) Sequence Analysis

Sequence of this protein was retrieved from NCBI as accession id NP_215691 and did the sequence alignment by using Protein BLAST against Protein database and finds the best homolog (1PS9) of this protein. Homolog was selected on the basis of its query coverage and its identity.

Figure 8 displays the sequence alignment of FadH gene with its homolog (1PS9) from different species. It is seen that FadH shares high degree of similarity with its homolog as query coverage is almost 99% and also identity is good.

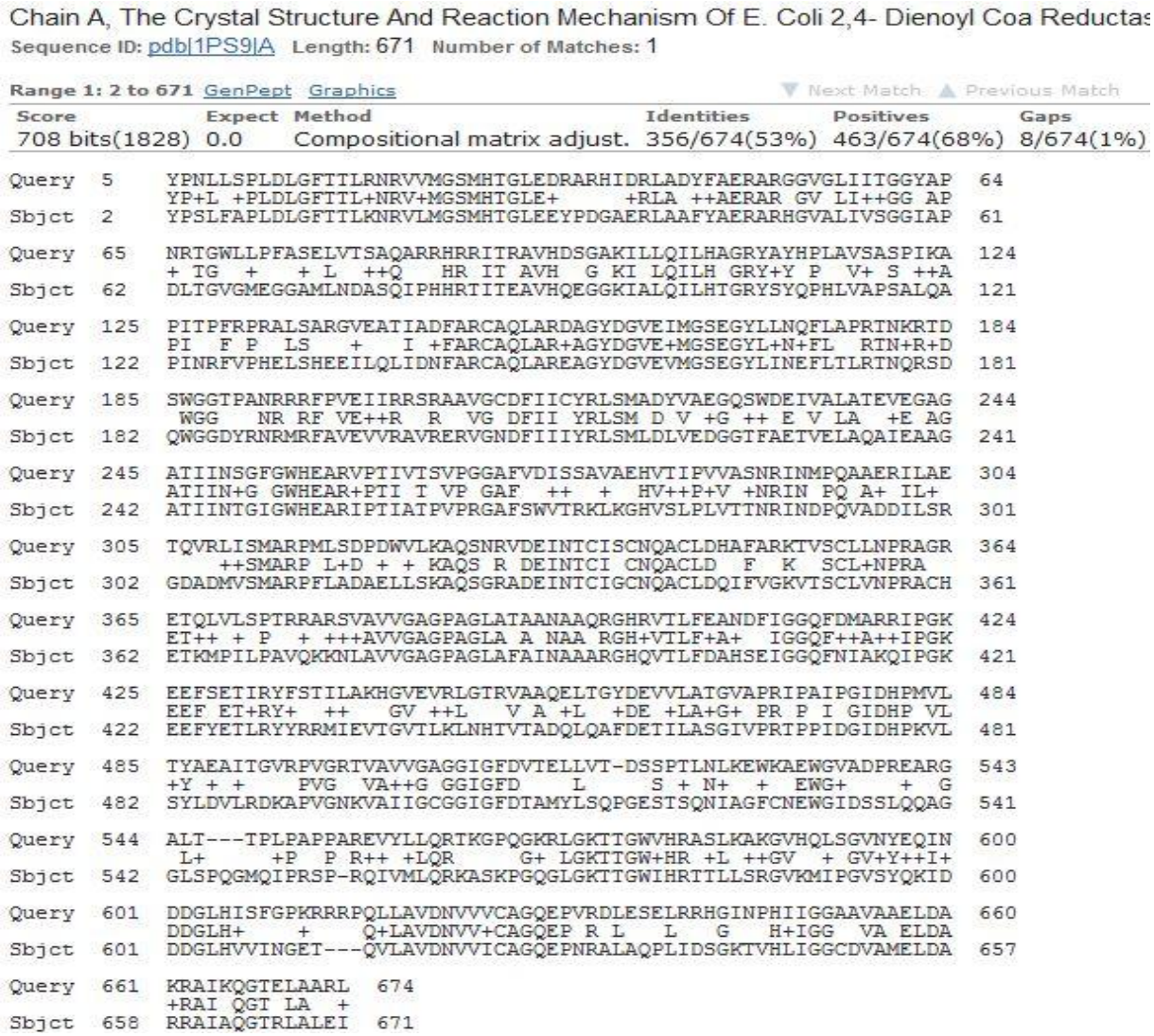


Fig. 8 Sequence alignment of MtFadH with its Homolog

2) Homology Modeling and Validation

Homology models were obtained for MtFadH using the template 1PS9. The best model was selected (Figure 9) on the basis of its **RMSD (root mean square deviation) value, verify-3D score, errat score** and **RC plot**.

The best model was superimposed with its homolog 1PS9 (Figure 10) and their RMSD value were also good which was 0.39 angstrom.

Strsucture was also validated using RC plot, verify-3D score and Errat score (Figure 11) in which very few residues lie in unfavoured reagion and other scores were also good.

Now this structure can be used for further simulation and docking process.

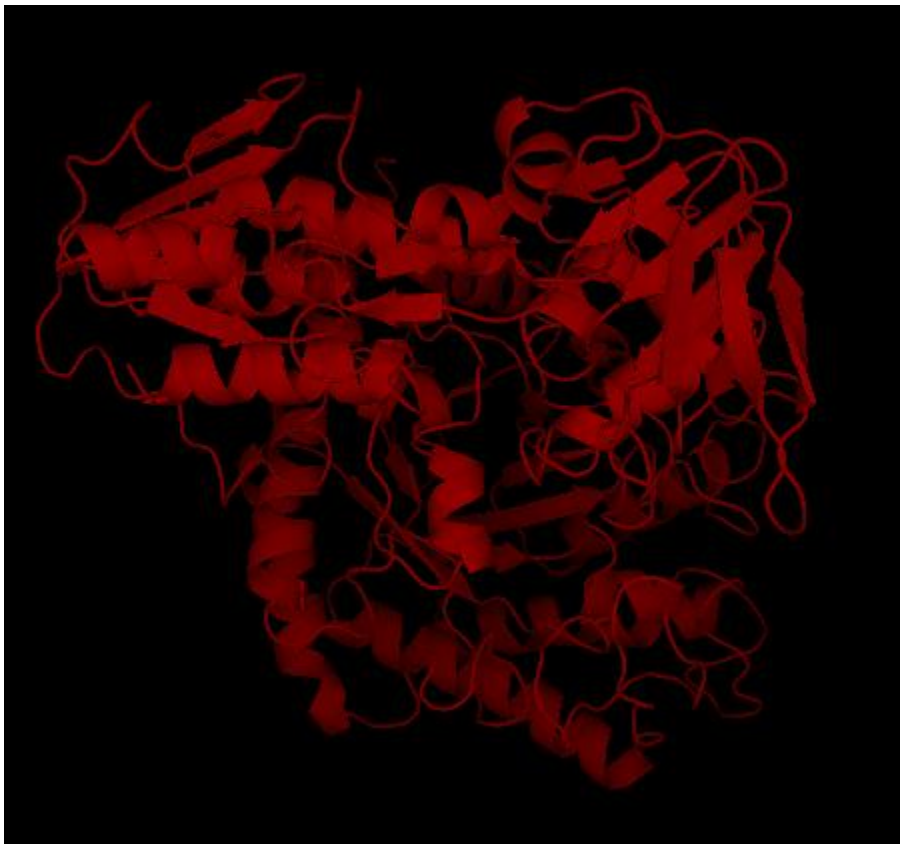


Fig. 9 Best Modeled structure of MtFadH protein

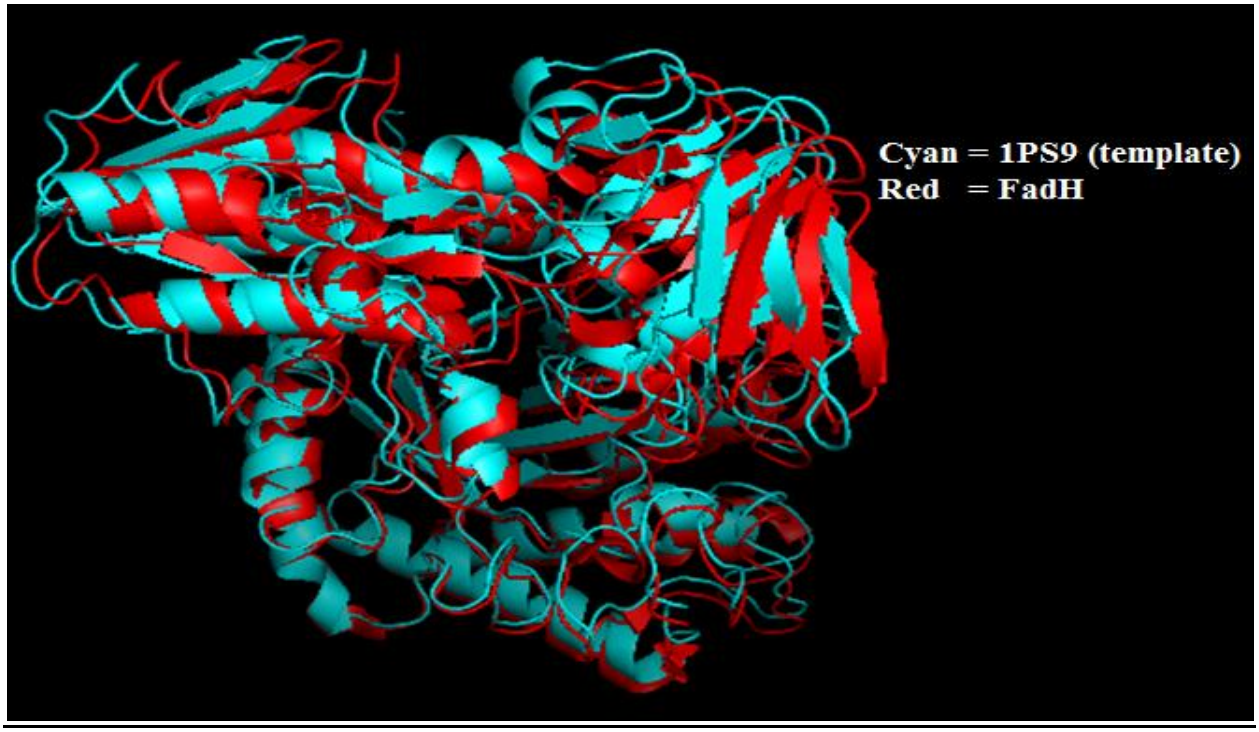


Fig. 10 Superimposition of MtFadH with 1PS9 (template)

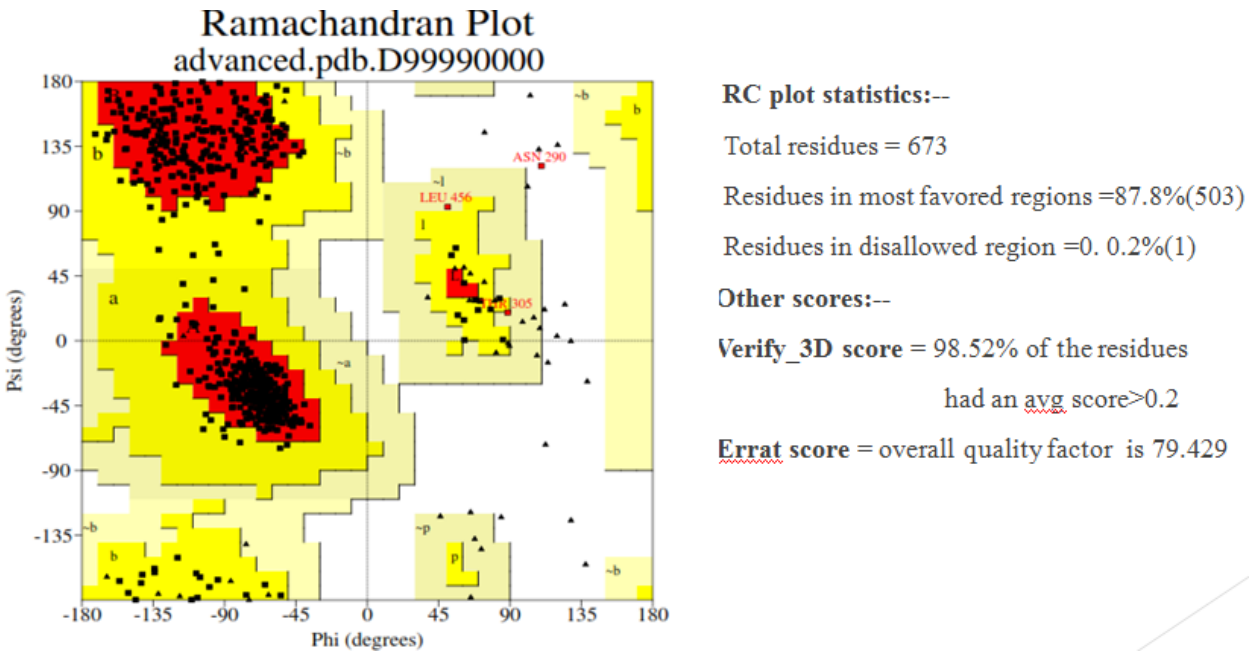


Fig. 11 Validation scores of best modeled MtFadH protein

3) Molecular Simulation

After validation of modeled structures and selecting the best one, simulation was done on it and most stable conformation (Fig 12) was retrieve.

There are three typical stages that happen in molecular dynamics simulation:

Minimization

Energy minimization is a procedure that attempts to minimize the potential energy of the system to the lowest possible point.

Equilibration

For most systems, this process will involve allowing some solvent to relax around the solute of interest. In many cases, the solvent is water, but it can also be a lipid bilayer, chloroform, etc and in our case the solvent was water.

Dynamics

The dynamics stage is the stage of interest for determining thermodynamic averages or sampling new configurations. The stage used for these applications is sometimes known as production dynamics.

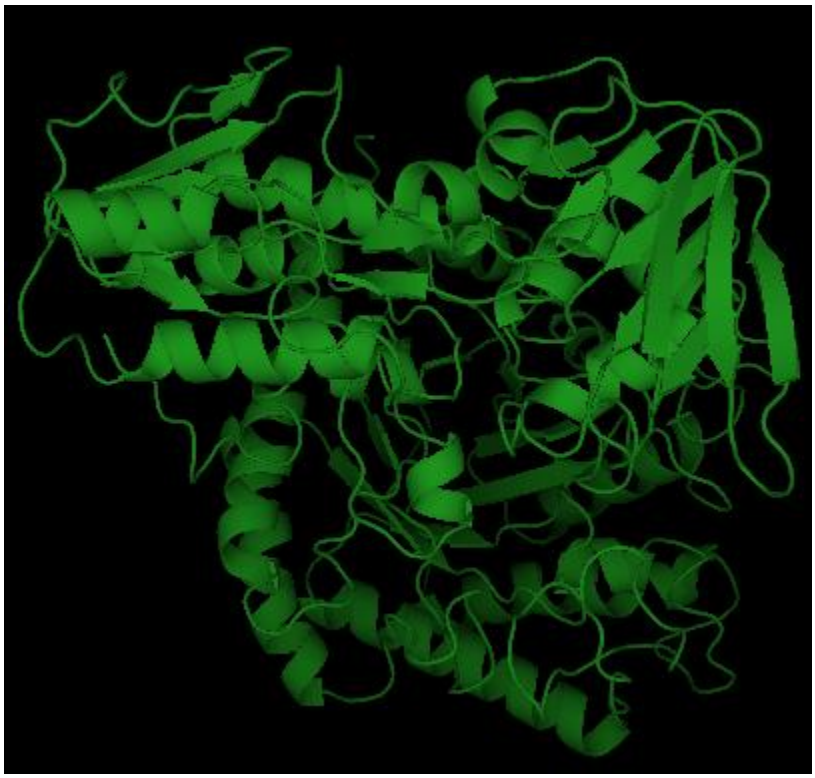


Fig. 12 Most stable conformation of MtFadH protein

4) Docking and Analysis

Docking was performed on simulated MtFadH protein by using ZINC Database.

But we were not able to find any competitive inhitors in this Datbase and so not a single inhibitors had good binding scores.

C) Polyketide synthase (Pks13)

1) Sequence Analysis

Sequence of this protein was retrieved from NCBI as accession id NP_218317 and did the sequence alignment by using Protein BLAST against Protein database and finds the best homolog (2HG4) of this protein. Homolog was selected on the basis of its query coverage and its identity.

Figure 13 displays the sequence alignment of Pks13 gene with its homolog (2HG4) from different species. It is seen that Pks13 shares high degree of similarity with its homolog as query coverage is almost 77% and also identity was 46% which is good.

```
336 bits(861)  1e-96  Compositional matrix adjust.  193/422(46%)  254/422(60%)  13/422(3%)

Query  119  IAIVGLSTRFPGEMNTPEQTWQALLEGRDGITDLPDGR-WSEFLEEPRLAARVAGARTRG  177
            IAIVG + RFPG++++PE  W+ +  G D I + P  R W     EP   AR+      G
Sbjct  41   IAIVGXACRFPGDVDSPESFWEFVSGGGDAIAEAPADRGW-----EPDPDARL------G  89

Query  178  GYLKDIKGFDSEFFAVAKTEADNIDPQQRMALELTWEALEHARIPASSLRGQAVGVYIGS  237
            G L     FD+ FF ++  EA   DPQQR+ LE++WEALE A     SLRG A GV+ G
Sbjct  90   GXLAAAGDFDAGFFGISPREALAXDPQQRIXLEISWEALERAGHDPVSLRGSATGVFTGV  149

Query  238  STNDYSFLAVSDPTVAHPYAITGTSSSIIANRVSYFYDFHGPSVTIDTACSSSLVAIHQG  297
             T DY       P     Y  TGT+SS+ + RV+Y     GP+ T+DTACSS L A+H
Sbjct  150  GTVDYGPRPDEAPDEVLGYVGTGTASSVASGRVAYCLGLEGPAXTVDTACSSGLTALHLA  209

Query  298  VQALRNGEADVVVAGGVNALITPMVTLGFDEIGAVLAPDGRIKSFSADADGYTRSEGGGM  357
             ++LR  E  + +AGGV    +P     F   G  LA DGR K FS  ADG+  +EG G+
Sbjct  210  XESLRRDECGLALAGGVTVXSSPGAFTEFRSQGG-LAADGRCKPFSKAADGFGLAEGAGV  268

Query  358  LVLKRVDDARRDGDAILAVIAGSAVNHDGRSNGLIAPNQDAQADVLRRAYKDAGIDPRTV  417
            LVL+R+  ARR+G  +LAV+ GSAVN DG SNGL AP+  AQ  V+RRA ++AG+    V
Sbjct  269  LVLQRLSAARREGRPVLAVLRGSAVNQDGASNGLTAPSGPAQQRVIRRALENAGVRAGDV  328

Query  418  DYIEAHGTGTILGDPIEAEALGRVVGRGRPADRPALLGAVKTNVGHLESAAGAASMAKVV  477
            DY+EAHGTGT LGDPIE  AL    G  R  D P  +G+VK+N+GH ++AAG A + K V
Sbjct  329  DYVEAHGTGTRLGDPIEVHALLSTYGAERDPDDPLWIGSVKSNIGHTQAAAGVAGVXKAV  388

Query  478  LALQHDKLPPSINFAGPSPYIDFDAMRLKMITTPTDWPRYGGYALAGVSSFGFGGANAHV  537
            LAL+H + P +++F  PSP I++D   + +++    WP       AGVSSFG  G NAHV
Sbjct  389  LALRHGEXPRTLHFDEPSPQIEWDLGAVSVVSQARSWPAGERPRRAGVSSFGISGTNAHV  448

Query  538  VV  539
            +V
Sbjct  449  IV  450
```

Fig. 13 Sequence Alignment of Pks13 with 2HG4

2) Homology Modeling and Validation

Homology models were obtained for MtPks13 using the template 2HG4. The best model was selected (Figure 14) on the basis of its **RMSD (root mean square deviation) value, verify-3D score, errat score** and **RC plot**.

Strsucture was also validated using RC plot, verify-3D score and Errat score (Figure 15) in which scores were not good.

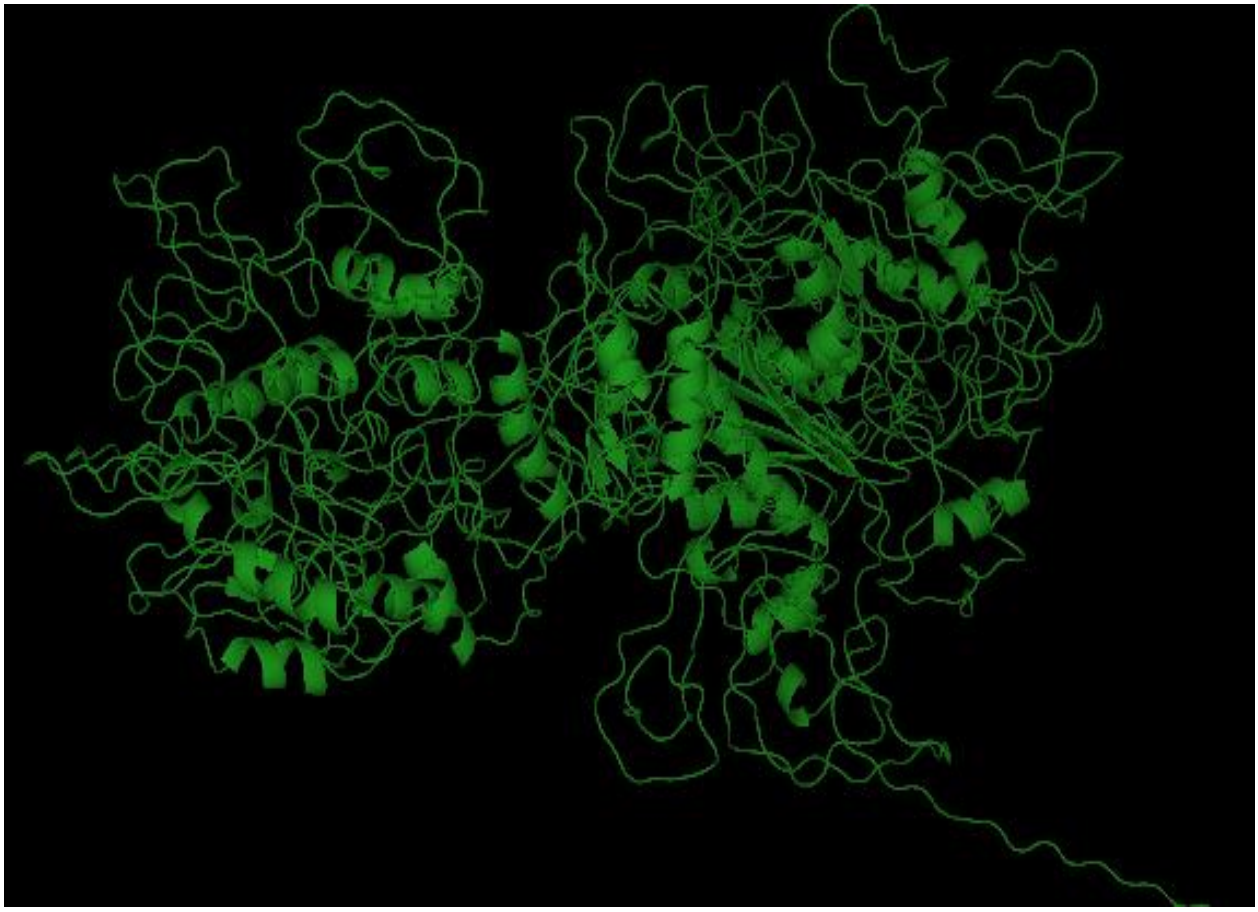


Fig. 14 Modeled MtPks13 protein structure with 2HG4 (template)

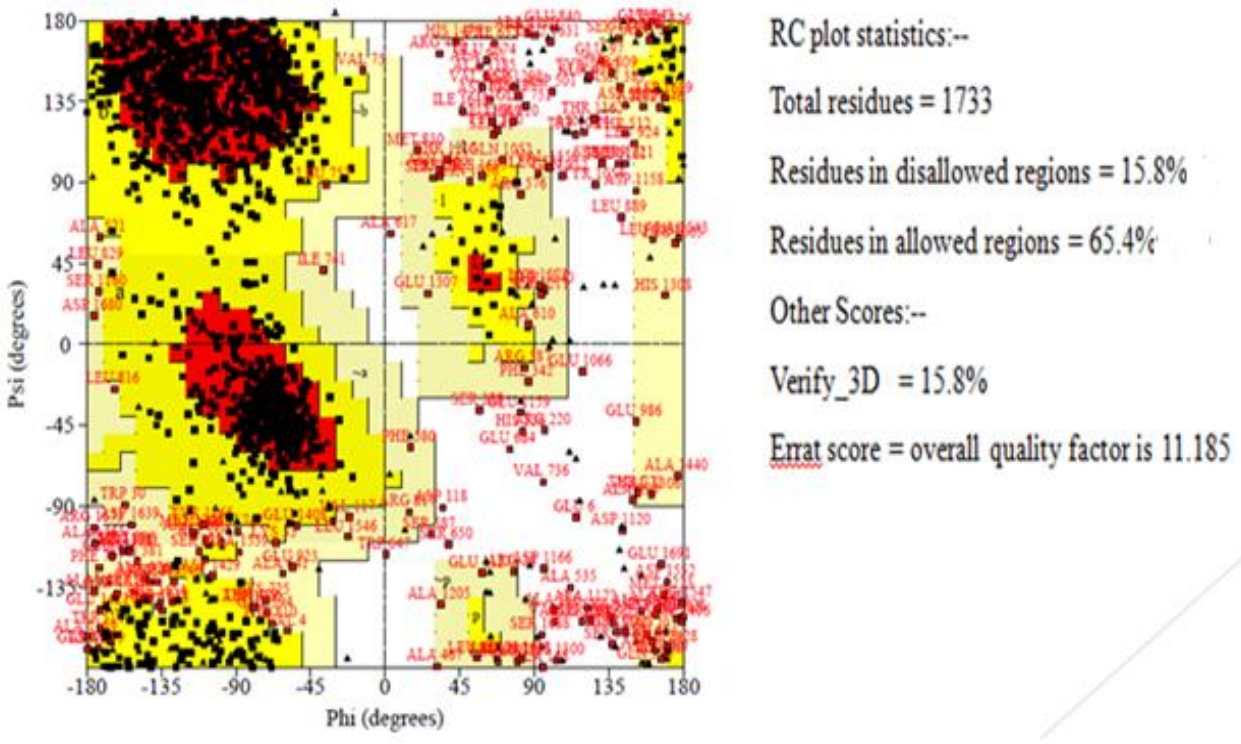


Fig. 15 Validation scores of modeled MtPks13 protein

3) Docking

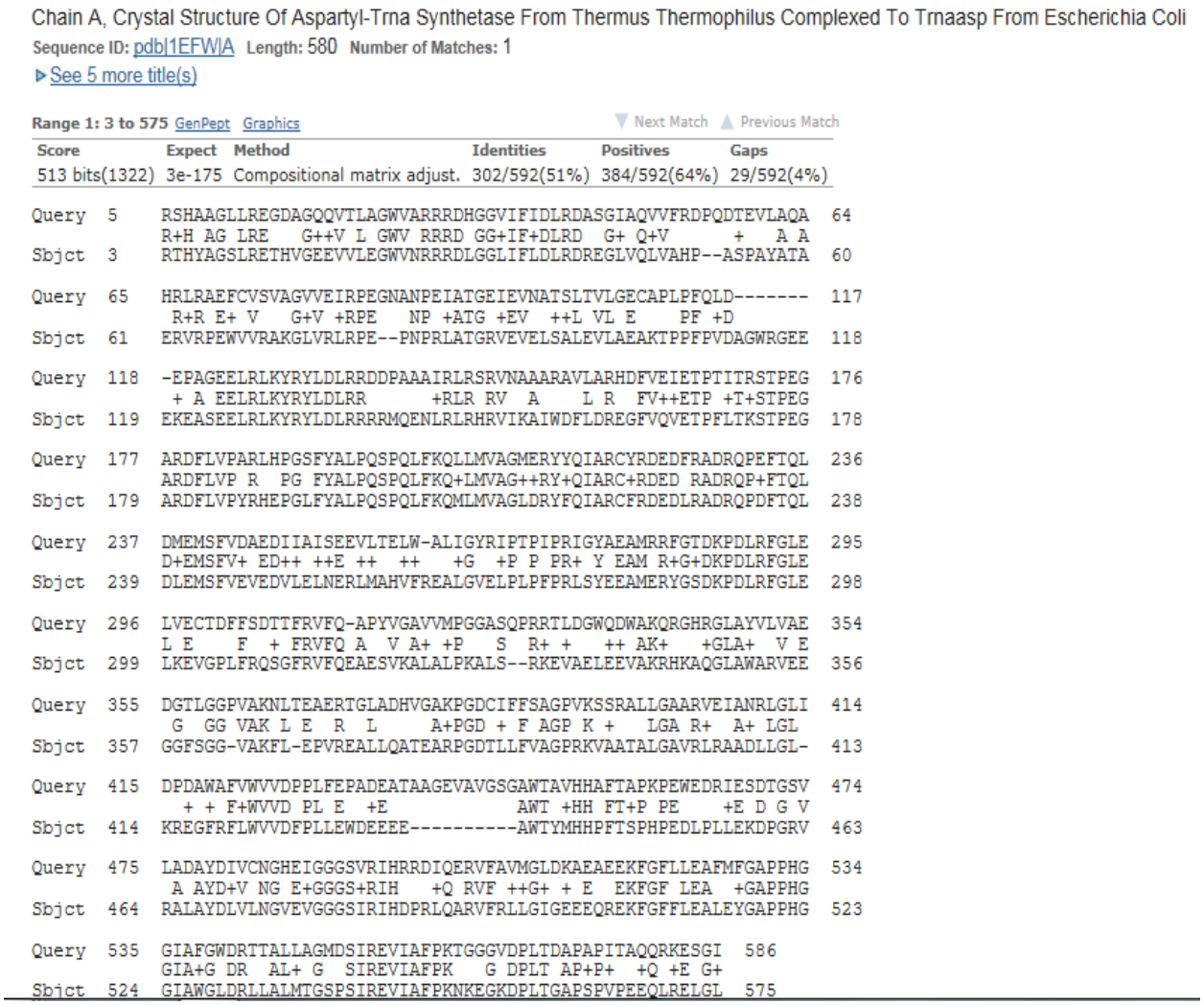
As validation score of MtPks13 protein were not good so one can not perform docking exercise on this protein as almost 15%-20% residues lie in unfavoured reagion in RC plot. So we further continue our project on other Drug Target.

D) Aspartyl synthetase (Asps)

1) Sequence Analysis

Sequence of this protein was retrieved from NCBI as accession id NP_218317 and did the sequence alignment by using Protein BLAST against Protein database and finds the best homolog (1EFW) of this protein. Homolog was selected on the basis of its query coverage and its identity.

Figure 16 displays the sequence alignment of Asps gene with its homolog (1EFW) from different species. It is seen that 1EFW shares high degree of similarity with its homolog as query coverage is almost 82% and also identity was 51% which is good.



Chain A, Crystal Structure Of Aspartyl-Trna Synthetase From Thermus Thermophilus Complexed To Trnaasp From Escherichia Coli
Sequence ID: pdb|1EFW|A Length: 580 Number of Matches: 1
See 5 more title(s)

Range 1: 3 to 575 GenPept Graphics Next Match Previous Match

Score	Expect	Method	Identities	Positives	Gaps
513 bits(1322)	3e-175	Compositional matrix adjust.	302/592(51%)	384/592(64%)	29/592(4%)

```
Query  5    RSHAAGLLREGDAGQQVTLAGWVARRRDHGGVIFIDLRDASGIAQVVFRDPQDTEVLAQA  64
            R+H AG LRE   G++V L GWV RRRD GG+IF+DLRD  G+ Q+V      +   A A
Sbjct  3    RTHYAGSLRETHVGEEVVLEGWVNRRRDLGGLIFLDLRDREGLVQLVAHP--ASPAYATA  60

Query  65   HRLRAEFCVSVAGVVEIRPEGNANPEIATGEIEVNATSLTVLGECAPLPFQLD-------  117
             R+R E+ V   G+V +RPE   NP +ATG +EV  ++L VL E    PF +D
Sbjct  61   ERVRPEWVVRAKGLVRLRPE--PNPRLATGRVEVELSALEVLAEAKTPPFPVDAGWRGEE  118

Query  118  -EPAGEELRLKYRYLDLRRDDPAAAIRLRSRVNAAARAVLARHDFVEIETPTITRSTPEG  176
             + A EELRLKYRYLDLRR      +RLR RV  A    L R  FV++ETP +T+STPEG
Sbjct  119  EKEASEELRLKYRYLDLRRRRMQENLRLRHRVIKAIWDFLDREGFVQVETPFLTKSTPEG  178

Query  177  ARDFLVPARLHPGSFYALPQSPQLFKQLLMVAGMERYYQIARCYRDEDFRADRQPEFTQL  236
            ARDFLVP R  PG FYALPQSPQLFKQ+LMVAG++RY+QIARC+RDED RADRQP+FTQL
Sbjct  179  ARDFLVPYRHEPGLFYALPQSPQLFKQMLMVAGLDRYFQIARCFRDEDLRADRQPDFTQL  238

Query  237  DMEMSFVDAEDIIAISEEVLTELW-ALIGYRIPTPIPRIGYAEAMRRFGTDKPDLRFGLE  295
            D+EMSFV+ ED++ ++E ++  ++   +G  +P P PR+ Y EAM R+G+DKPDLRFGLE
Sbjct  239  DLEMSFVEVEDVLELNERLMAHVFREALGVELPLPFPRLSYEEAMERYGSDKPDLRFGLE  298

Query  296  LVECTDFFSDTTFRVFQ-APYVGAVVMPGGASQPRRTLDGWQDWAKQRGHRGLAYVLVAE  354
            L E    F  + FRVFQ A  V A+ +P   S  R+ +   ++ AK+   +GLA+  V E
Sbjct  299  LKEVGPLFRQSGFRVFQEAESVKALALPKALS--RKEVAELEEVAKRHKAQGLAWARVEE  356

Query  355  DGTLGGPVAKNLTEAERTGLADHVGAKPGDCIFFSAGPVKSSRALLGAARVEIANRLGLI  414
             G  GG VAK L E  R  L     A+PGD + F AGP K +   LGA R+  A+ LGL
Sbjct  357  GGFSGG-VAKFL-EPVREALLQATEARPGDTLLFVAGPRKVAATALGAVRLRAADLLGL-  413

Query  415  DPDAWAFVWVVDPPLFEPADEATAAGEVAVGSGAWTAVHHAFTAPKPEWEDRIESDTGSV  474
              + + F+WVVD PL E  +E            AWT +HH FT+P PE    +E D G V
Sbjct  414  KREGFRFLWVVDFPLLEWDEEEE----------AWTYMHHPFTSPHPEDLPLLEKDPGRV  463

Query  475  LADAYDIVCNGHEIGGGSVRIHRRDIQERVFAVMGLDKAEAEEKFGFLLEAFMFGAPPHG  534
             A AYD+V NG E+GGGS+RIH   +Q RVF ++G+ + E  EKFGF LEA  +GAPPHG
Sbjct  464  RALAYDLVLNGVEVGGGSIRIHDPRLQARVFRLLGIGEEEQREKFGFFLEALEYGAPPHG  523

Query  535  GIAFGWDRTTALLAGMDSIREVIAFPKTGGGVDPLTDAPAPITAQQRKESGI  586
            GIA+G DR  AL+ G  SIREVIAFPK   G DPLT AP+P+  +Q +E G+
Sbjct  524  GIAWGLDRLLALMTGSPSIREVIAFPKNKEGKDPLTGAPSPVPEEQLRELGL  575
```

Fig. 16 Sequence Alignment of MtAsps with 1EFW

2) Homology Modeling and Validation

Homology models were obtained for MtAsps using the template 1EFW. The best model was selected (Figure 17) on the basis of its **RMSD (root mean square deviation) value, verify-3D score, errat score** and **RC plot**.

Structure was also validated using RC plot, verify-3D score and Errat score (Figure 18) in which scores were not good.

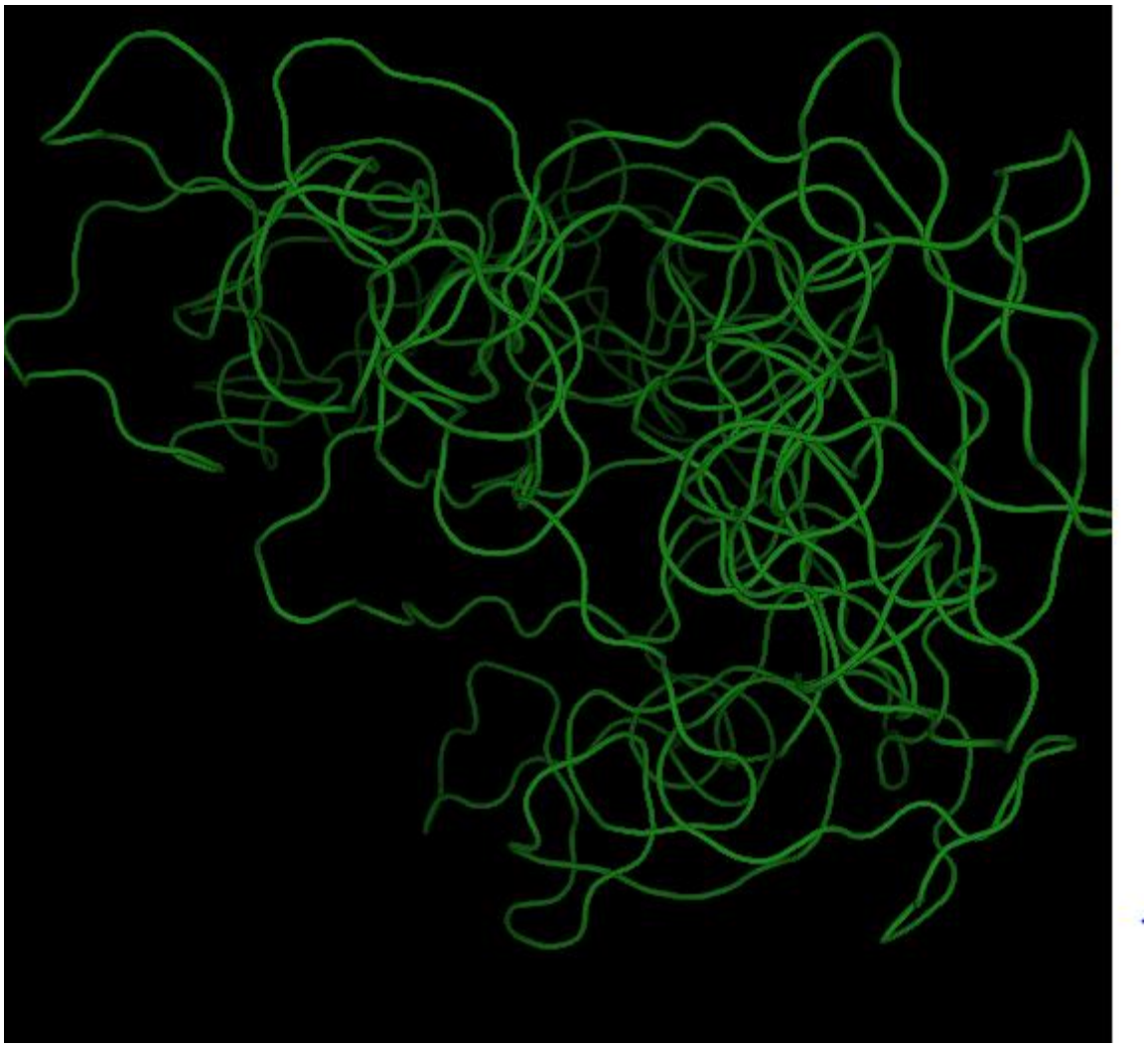


Fig. 17 Modeled MtAsps protein structure with 1EFW (template)

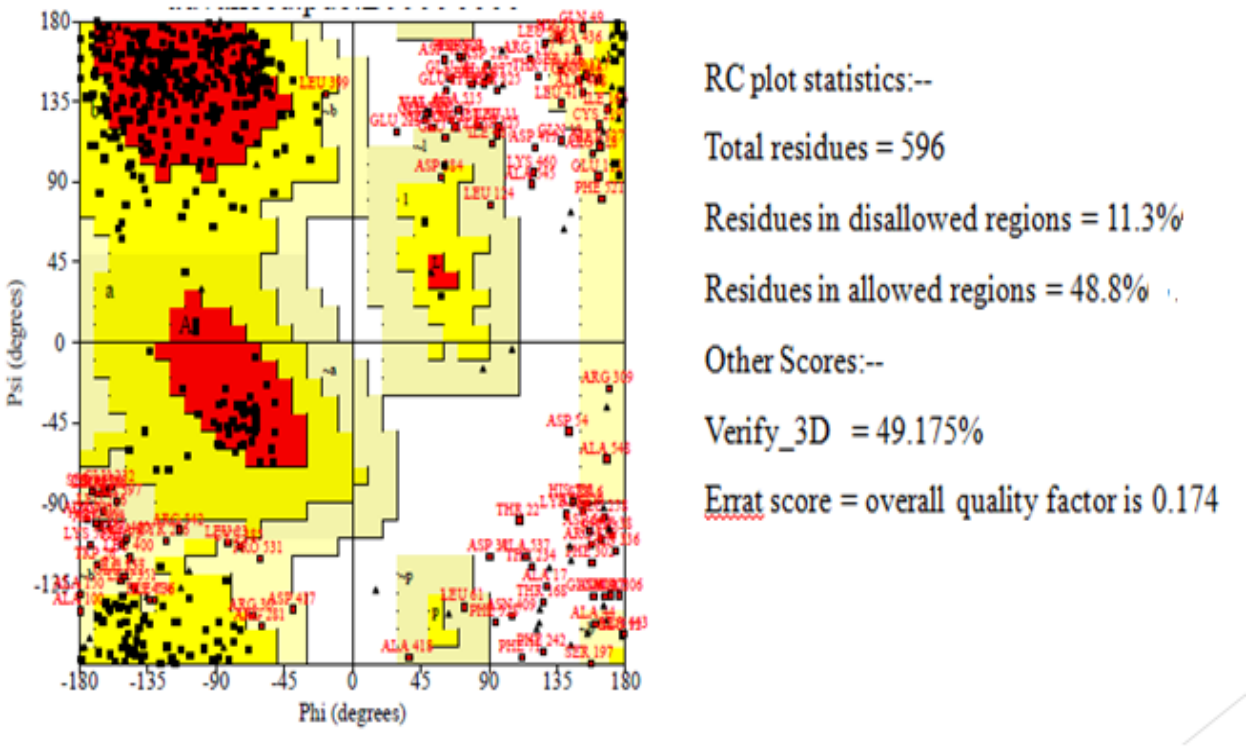


Fig. 18 Validation scores of modeled MtAsps protein

3) Docking

As again validation score of MtAsps protein were also not good so one cannot perform docking exercise on this protein also as almost 12% residues lie in unfavoured reagion in RC plot and other scores were also not good.

Conclusion

From four Drug Targets, it is clear that MtMurI can be a vital Drug Target in *Mycobacterium*. We could also identify some good potent inhibitors for this protein.

Also MtMurI is essential for bacterial growth and lacks a human homolog which makes it an attractive drug target. Analysis of the structure of MtMurI would shed light on its mechanism of action and lead to the identification of potent inhibitors.

These findings would expedite the pace of further experimental studies in this direction and pave the way to demystify a number of hitherto unresolved issues about the biology of this enzyme. Overall, this may provide an effective and robust strategy to counteract the tuberculosis pathogen.

Second MtFadH protein can also be a good drug target as its score was also good but one has to further analyse this protein and has to find some potent inhibitors for this protein either in some organisms or from any database.

References

[1] Zumla A, Nahid P, Cole ST: Advances in the development of new tuberculosis drugs and treatment regimens: [Internet: http://www.ncbi.nlm.nih.gov/pubmed/2362950]

[2] Sunil Sethi, Abhishek Mewara, Sunil Kumar Dhatwalia multidrug resistance in *Mycobacterium tuberculosis*: [Internet:http://www.biomedcentral.com/1471-2334/13/137]

[3] Statistics [Internet: http://www.who.int/mediacentre/factsheets/fs104/en/]

[4] Statistics [Internet: http://www.tbfacts.org/tb-statistics-india.html]

[5] Drug Resistance [Internet: http://en.wikipedia.org/wiki/Drug_resistance]

[6] Rabia Johnson,ElizabethM. Streicher, Gail E:Drug Resistance in *Mycobacterium tuberculosis*: [Internet: http://www.horizonpress.com/cimb/abstracts/v8/08.html]

[7] Ioerger TR1, O'Malley T, Liao R, Guinn KM :Identification of new drug targets: [Internet: http://www.ncbi.nlm.nih.gov/pubmed/24086479]

[8] Kimura C1, Mizugaki M, Yamanaka H, Fujino M, Morishima T: 2,4-Dienoyl-CoAreductases: significance: [Internet:http://www.ncbi.nlm.nih.gov/pubmed/15344554]

[9] Sengupta S, Ghosh S, Nagaraja V**:** Moonlighting function of glutamate racemase from *Mycobacterium tuberculosis*: racemization and DNA gyrase inhibition are two independent activities of the enzyme. Microbiology 2008, 154:2796-2803

[10] De Jonge BLM, Kutschke A, Uria-Nickelsen M, Kamp HD, Mills SD: Pyrazolopyrimidinediones Are Selective Agents for Helicobacter pylori That Suppress Growth through Inhibition of Glutamate Racemase (MurI). *Antimicrobial Agents and Chemotherapy* 2009, 53**:**3331-3336.

[11] E Kriege : [Internet: www.cmbi.ru.nl/edu/bioinf4/articles/homologymodeling.pdf]

[12] Energy Minimization: [Internet:http://www.charmmtutorial.org/index.php/Minimization]

[13]Equilibration [Internet: http://www.gromacs.org/Documentation/Terminology/Equilibration]

[14] [Internet:http://www4.ncsu.edu/~franzen/public_html /CH795N/lecture/IV/IV.html]

[15] [Internethttp://en.wikipedia.org/wiki/Tuberculosis#Signs_and_symptoms]